Sorbonne université
Master Mathématiques et applications

Cours d'analyse
pour l'Agrégation Externe de Mathématiques
2022-2023

Jimmy Lamboley

(et un chapitre de probabilités par Quentin Berger)

Version du 5 février 2023

Table des matières

Introduction

Contenu du polycopié

Ce polycopié contient un cours d'analyse destiné à des étudiants préparant l'Agrégation Externe de Mathématiques ; il se base sur le programme 2022[1] recopié en page 11, et en calquera d'ailleurs notablement le plan. À ce titre, ce n'est pas un cours strictement "linéaire", au sens où certaines remarques ou commentaires n'auront de sens que si le lecteur connaît déjà quelques notions qui ne seront abordées que dans des chapitres ultérieurs[2]. C'est à mon avis un aspect fondamental du concours de l'agrégation, d'ailleurs, d'obtenir une vision transverse du programme, à l'opposé de ce qui est souvent demandé à l'université, où l'on peut avoir l'illusion que chaque cours est plus ou moins indépendant des autres. C'est donc un cours à utiliser, pour la grande majorité, en "seconde lecture". Il a vocation à traiter la quasi-intégralité du programme d'analyse de l'agrégation et à être enseigné en moins de 6 mois, alors qu'il représente peu ou prou l'intégralité du programme d'une Licence de Mathématiques, avec une excursion modeste mais réelle vers quelques notions habituellement abordées en M1. Il y aura donc nécessairement des raccourcis et des ellipses, et ce cours ne saurait se substituer à différents ouvrages de Licence qui peuvent recouvrir chacun des chapitres abordés ici. Je glisserai donc de nombreuses références afin de permettre aux étudiants qui n'ont jamais vu (ou oublié !) telle ou telle notion d'aller creuser par eux-mêmes ; évidemment, ce ne seront que des propositions de références, l'étudiant préparant l'agrégation comprendra assez vite qu'un aspect central de son travail est de se constituer une bibliographie qui lui est propre et qu'il sait utiliser efficacement. Néanmoins, un objectif restera de fournir en un seul document une représentation fidèle et aussi exhaustive que possible du programme d'analyse de l'agrégation, incluant de multiples énoncés rarement évoqués en Licence, mais qui font désormais partie du folkore de l'agrégation[3].

Ce cours prépare autant à l'épreuve écrite qu'à l'épreuve orale d'Analyse & Probabilités. Il contient l'indispensable de ce qu'il faut connaître pour aller aux écrits, avec de nombreux exercices qui peuvent être vus comme un entraînement aux questions de l'écrit[4], et on fera très souvent référence à la liste des leçons d'oral pour mettre en lumière un énoncé ou une remarque. On notifiera également quand une démonstration ou un exemple peut constituer un bon exemple de développement. Encore une fois, on précisera autant que possible des références publiées pour que l'étudiant puisse aller chercher le livre de son choix le jour de l'épreuve orale[5]. Par contre, ce cours ne prépare pas en tant que tel à l'épreuve de Modélisation, quelle que soit l'option choisie. Il peut néanmoins permettre aux étudiants ayant choisi l'option B (Calcul Scientifique) de voir les bases du cours d'équations différentielles, ainsi que quelques éléments d'analyse numérique, même si on abordera surtout la partie "tronc commun" de ce sujet.

---

1. Ce programme a très peu évolué ces dernières années : on note une légère évolution dans le chapitre sur les distributions vers 2019, voir la note 31.

2. Le lecteur en question peut bien sûr sauter ces remarques en première lecture, mais nous ne saurions trop lui conseiller d'y revenir en seconde lecture pour profiter un peu de l'intérêt de cet ouvrage.

3. Ce folklore se développe désormais grâce (ou à cause, c'est selon) aux multiples pages web qui ont fleuri ces dernières années et où des étudiants fraîchement agrégés reproduisent (voire publient) et ainsi perpétuent de multiples ressources (notamment de plans et de développements), qui se transmettent ainsi d'année en année chez les candidats au concours. Ce folklore existait auparavant, mais était nettement plus localisé dans une "prépa agreg" donnée.

4. Certains exercices seront d'ailleurs tirés des annales du concours.

5. Rappelons qu'il n'est pas autorisé lors des épreuves orales d'amener un polycopié (à l'exception du rapport de jury, que je conseille à tout étudiant de ramener pour ses oraux).

Niveau du cours

Une des vocations avouées de ce cours, est de pouvoir s'adapter à un large public : très imparfaitement, on pourrait distinguer deux catégories de candidats, à savoir ceux qui visent d'être agrégés, et ceux qui savent déjà plus ou moins qu'ils le seront, mais souhaitent obtenir un (très) bon classement. Mon avis est que deux candidats qui appartiennent chacun à l'une de ces catégories devraient avoir des façons de travailler en partie similaires, et en partie très distinctes : le premier devrait passer la majeure partie de son temps à se concentrer sur les bases du programme, à combler les lacunes qui l'ont suivi durant les années passées, sur tel ou tel aspect du programme. C'est un travail déjà vertigineux et ambitieux, et mon conseil serait de ne surtout pas essayer d'aller travailler des sujets trop élaborés ou à la limite du programme [6] tant qu'on n'a pas acquis une certaine maîtrise. Le second, lui, devrait faire exactement la même chose (combler ses lacunes, car il en a certainement malgré ses ambitions) mais en bien moins de temps, pour ensuite aller creuser certaines directions du programme pour alimenter ses connaissances de résultats plus profonds et plus originaux.

Afin de guider un peu le lecteur dans ses priorités, les paragraphes et énoncés de ce document seront proposés en 3 catégories :

1. Sans étoile : le programme à proprement parler (tel que je l'interprète), et qui représentera une grosse majorité du document.
2. Une étoile * : des ouvertures possibles, qui nécessitent un investissement non négligeable mais pas démesuré, et qui peuvent mettre en valeur le programme une fois qu'on commence à mieux le maîtriser. Je dirais que ce n'est pas la priorité, que le candidat doit d'abord vérifier qu'il maîtrise les différents aspects d'un chapitre avant de se lancer dans ces considérations. Et si le temps le permet, le candidat pourra choisir la ou les directions "étoilées" qui lui paraissent les plus accessibles et en accord avec sa maîtrise du sujet.
3. Deux étoiles ** : des sujets difficiles, qui peuvent être culturels mais qui permettent d'avoir une bonne prise de recul, ou alors qui vont toucher les limites du programme et nécessitent d'avoir une maîtrise totale de tous les aspects du chapitre, voire de plusieurs chapitres. Ces sujets sont à réserver aux candidats très ambitieux, et peuvent être "sautés" absolument sans regret par les candidats qui n'ont pas une ambition particulière sur leur classement au concours.

Conseils de travail

J'en profite pour annoncer ce qui représente, selon moi, l'aspect le plus important qui doit permettre au candidat de tirer au mieux parti de sa préparation, et ainsi maximiser ses chances de réussite au concours : l'**auto-évaluation**. L'étudiant doit, pendant sa préparation et pendant les épreuves, avoir conscience de ce qu'il connaît bien, et de ce qu'il ne maîtrise pas encore. En ce sens, je ne saurais trop conseiller aux étudiants, dans leur préparation et dans leur gestion des épreuves, à tendre vers plus

6. Ça peut paraître une évidence, et c'est pourtant un écueil que je trouve très fréquent. Je me souviens parfaitement, lorsque je préparais l'agrégation, que la plupart des élèves de ma promo avaient travaillé et appris la démonstration des théorèmes de Hahn-Banach, qui figure dans les premières pages du livre de H. Brézis [Bré05], notamment dans l'idée de les proposer en développement : ces théorèmes étaient (et sont toujours) hors programme, ce qui n'est pas nécessairement incongru pour une idée de développement, bien que déjà a priori ambitieux. Mais ils présentent surtout deux difficultés à mon avis "éliminatoires" pour la quasi-totalité des candidats ; dans la version générale de [Bré05], ils reposent sur le Lemme de Zorn, variante de l'axiome du choix, ce qui oblige à des considérations largement en dehors des idées du programme, et touchent à des mathématiques à mon avis bien plus subtiles et difficiles que ce qu'est attendu d'un candidat à l'agrégation. Le second défaut est, à mon avis, que les utilisations et conséquences desdits théorèmes (avec en premier lieu l'étude des topologies faibles), sont également largement trop difficiles pour beaucoup de candidats (ce qui explique entre autres leur absence du programme), et ne constituent pas du tout le standard de ce qu'on doit savoir pour devenir agrégé. Qu'on ne se méprenne pas, un candidat peut très bien parler de ces théorèmes, c'est même évoqué dans le rapport de jury, et je proposerai même ces théorèmes dans ce document : mais d'une part je donnerai des variantes moins générales, plus simples à démontrer, et qui suffisent pour bon nombre d'applications, et je préciserai lesdites applications (se cantonner par exemple au cas des espaces de Hilbert est nettement plus raisonnable, et beaucoup plus dans l'idée du programme, même si dans ce cas les théorèmes de Hahn-Banach (dans leur version géométrique) deviennent trop simples à démontrer pour constituer un développement...). Et d'autre part, j'indiquerai aussi clairement que possible que cela ne peut être abordé que par les meilleurs candidats, ceux qui visent un excellent classement.

d'humilité[7] : à l'écrit par exemple, il faut absolument chiader au maximum les premières questions (souvent élémentaires mais, comme le rappelle chaque année le rapport de jury, très discriminantes[8]), et passer une partie très substantielle de son temps (voire tout son temps !) à plancher sur le premier quart ou tiers du sujet. Pour la majorité des candidats, le reste du sujet, plus difficile, sera inaccessible, et le temps passé à travailler sur ces parties ne rapportera pas beaucoup, voire pas du tout de points. Les épreuves écrites durant pas moins de 6 heures, il est plus raisonnable de tenter de récupérer l'intégralité des points destinés aux questions qui nous sont accessibles, en prenant le temps qu'il faut pour ne pas tomber dans les pièges classiques de raisonnement ou de rédaction qui seront à juste titre[9] pénalisés[10]. Pour les épreuves orales, la remarque sur la nécessité de maîtriser ce dont on parle vaut d'autant plus, puisqu'il est laissé au candidat une liberté non négligeable sur le degré de profondeur ou de technicité qu'il choisit de développer[11].

Il me semble important de comprendre qu'il ne faut pas essayer de tout savoir, au risque évident de ne rien maîtriser. Je représente souvent l'ensemble des notions à connaître à l'agrégation comme une pyramide ; une notion plus en hauteur dans cette pyramide ne saurait être bien comprise et maîtrisée si les fondations de ladite pyramide sont parsemées de trous ou sont friables[12]. Il faut donc considérer cette année de travail comme un moyen de *bien* construire sa propre pyramide, en cherchant à la faire grandir en faisant attention que chaque étage soit bien solide. Le temps de préparation étant limité, il va sans dire que cette méthode ne permettra pas à tous d'avoir une pyramide complètement[13] construite au moment des épreuves, mais ça n'est pas grave, au sens qu'on vous demande plus d'avoir des bases solides, que des connaissances très élevées[14]. Il est donc raisonnable de laisser de côté (pour plus tard ou pour toujours) certaines excursions qui seront proposées par ce document : les étoiles sont une façon d'aider le candidat à faire des choix, mais rappelons que cela reste subjectif et imparfait. Quant aux étudiants qui ont déjà des bases très solides, et qui sentent qu'ils ont plus ou moins déjà ce qu'il faut pour devenir agrégé, ils ne doivent évidemment pas pour autant s'empêcher de faire monter leur "pyramide" aussi haut qu'ils le souhaitent[15].

Pour représenter mon propos par une dernière métaphore, j'aime comparer les épreuves orales à un concours de saut en hauteur. Dans la première moitié de l'épreuve (défense de plan et développement), vous avez un complet contrôle sur ce que vous proposez au jury. Si vous proposez une hauteur bien

---

7. Il s'agit d'une remarque générale ; il arrive que quelques étudiants aient l'écueil inverse, ce qui est également dommage, mais plus rare.

8. Au sens où ces questions sont abordées par tous les candidats, mais avec un succès très varié ; dans le cadre d'un concours, c'est évidemment ce qui permet le mieux au jury de classer les candidats.

9. Rappelons quand même, si cela était nécessaire, qu'il s'agit d'un concours pour devenir enseignant.

10. Un bon exemple consiste à justifier, ou au moins dire si cela est évident, qu'un nombre est non nul quand on s'apprête à diviser par ce dernier.

11. On voit ici un aspect très négatif de la diffusion du folkore évoqué en note 3, car la plupart des sources qu'on peut trouver sur le net sont, à mon avis, d'un niveau bien trop élevé pour de nombreux candidats (et pour cause, il faut que l'étudiant produisant ces sources ait suffisamment de temps et de confiance en soi pour considérer que sa production vaut la peine d'être lue). Sans blâmer aucunement les personnes qui prennent le temps de produire ces sources, dont je me sers d'ailleurs largement pour préparer ce cours, je pense qu'elles peuvent donner une illusion de l'attente du concours ; certains étudiants feraient mieux de se cantonner à des plans et développements plus modestes qu'ils ont choisis et qu'ils maîtrisent, plutôt que d'essayer de recaser à tout va un joli énoncé, dont la démonstration ou le contexte les dépassera, et qui ne sera pas en accord avec le reste des connaissances du candidat, ce dont le jury n'aura aucun mal à se rendre compte rapidement. Notons que cette remarque vaut également pour la plupart des livres que j'ai trouvés et qui sont dédiés à l'agrégation externe. Je les trouve très souvent trop difficiles, et c'est entre autre ce qui a motivé l'écriture de ces notes, que j'espère plus accessibles sans manquer de jolies applications du programme.

12. L'image d'une pyramide en chateau de cartes est assez bonne car on comprend bien qu'il ne faut pas laisser de trou pour monter dans les étages, mais il ne faudrait pas croire que je pense la connaissance des étudiants si fragile :).

13. On entend ici, contenant tout le programme ; mais il va sans dire qu'en réalité cette pyramide n'a pas de "plafond" ou de "fin" si ce n'est celle où l'étudiant s'arrête de la construire.

14. Tout ceci est largement détaillé dans l'introduction des rapports de jury de ces dernières années.

15. À condition de le faire de la bonne manière : prendre en prérequis d'une leçon ou d'un développement un ensemble d'énoncés totalement hors programme n'est pas souvent une bonne stratégie, et peut laisser le jury peu convaincu. Si un bon étudiant souhaite inclure un énoncé de niveau élevé dans sa hotte d'agrégatif, je lui conseille de travailler afin de décomposer toutes les étapes nécessaires pour arriver à cet énoncé avec les seules briques (ou cartes, selon la métaphore choisie) fournies par le programme officiel. S'il constate que le cheminement est particulièrement vertigineux, il est probablement de bon ton de laisser tomber cet énoncé pour aller vers des choses moins ambitieuses, qui seront certainement mieux valorisées pour les épreuves orales. Je tenterai de donner quelques exemples de telles manœuvres dans ce cours.

trop ambitieuse pour vous, le jury vous verra alors simplement sauter bien en-dessous de *votre* barre, et n'aura pas vraiment une idée claire de votre niveau. Si par contre vous proposez une barre que vous savez que vous pouvez sauter avec quasi-certitude, il saura au moins que vous avez ces capacités, et dans la suite de l'épreuve, il pourra vous proposer de sauter un peu plus haut, et identifiera au mieux votre hauteur maximale. Donc pendant cette année de préparation, vous ne devez pas penser qu'il y a une "hauteur standard" attendue par le jury, qui serait le standard attendu pour être agrégé ; vous devez plutôt, suivant les sujets abordés, identifier votre "hauteur", et chercher à l'améliorer doucement mais sûrement [16].

Le lecteur trouvera au Chapitre 0 quelques éléments supplémentaires pour étayer les propos précédents.

Du général au particulier, ou le contraire ?

Un autre parti pris par ce cours, et qui n'est pas souvent pris, à mon avis, dans la littérature, et de *ne pas* aller du général au particulier, mais au contraire, d'aller aussi souvent que possible du particulier au général. Après une grosse quinzaine d'années d'enseignement, je me rends compte que les étudiants ont très souvent l'impression qu'une idée centrale des mathématiques est d'élaborer des concepts les plus généraux possibles avec des propriétés et des théorèmes les plus abstraits possibles, qui pourront s'appliquer dans les situations les plus variées. Dit autrement, l'étudiant peut penser inutile de comprendre les cas particuliers *avant* de se lancer dans la version la plus générale d'une notion ou d'une propriété, puisque de toute façon cette dernière contiendra bien toute l'information nécessaire. Vous l'aurez compris, je ne suis pas du tout d'accord, et je pense que le programme et l'idée du concours, est d'aller également à l'encontre de cette idée. Il n'est pas anodin que les titres de leçons se terminent très souvent par "Exemples" ou "Exemples et applications"[17]. Tout énoncé dont un étudiant ne saurait pas exhiber un exemple pertinent sera, lors d'une épreuve orale, très mal perçu par le jury. À titre d'exemple, si vous énoncez un résultat pour tout espace de Banach, mais que vous ne savez pas produire d'exemple plus riche que $\mathbb{R}^n$ comme espace de Banach, le jury doutera sérieusement de votre compréhension de cet énoncé. Au contraire, si vous arrivez à expliquer et développer un exemple pertinent, quand bien même vous n'abordez pas les extensions que cela peut cacher, cela peut s'avérer très valorisé [18]. Pour cette raison, vous verrez que contrairement à ce qui est souvent fait à ce niveau, ce cours ne débute pas par un cours de topologie métrique, mais bien par le cas particulier des réels. C'est aussi ce qui est fait dans le programme officiel, et il me semble indispensable que les étudiants prennent dès le début de l'année la mesure du degré de généralité dans lequel il convient de se placer (à l'écrit, mais encore plus à l'oral où la liberté du candidat est grande quant au contexte qu'il choisit de développer [19]). Et bien sûr, au-delà des contraintes spécifiques du concours, je pense que c'est dans cet ordre de l'apprentissage (du particulier au général) qu'on a les meilleures chances de comprendre et de s'approprier les outils mis en jeu. Cela m'obligera à certaines circonvolutions au cours de ce document, mais j'espère que cela ne perturbera pas trop la lecture ; rappelons que le document n'a pas vocation à être parfaitement linéaire.

Que faut-il savoir ?

J'en profite maintenant pour attaquer la question qui m'est le plus souvent posée au cours de l'année, et qui se résume, à des variantes près à "Doit-on savoir *tout* démontrer ?" (notamment dans ce qu'on

16. Il n'y a qu'une raison valable que je voie, à chercher à identifier la hauteur "standard" : c'est durant la phase où on doit décider si on passe le concours ou non. J'ai été amené à rencontrer pas mal de personnes en reconversion, qui veulent savoir avec relative précision, dès le début de l'année, s'ils ont ou non une chance au concours. C'est une démarche compréhensible et naturelle. Mais une fois qu'une décision a été prise (disons en octobre précédent le concours), il ne me paraît pas très productif de revenir sur la question pendant la suite de la préparation. Il vaut mieux consacrer son temps et son énergie à travailler et progresser.

17. Le rapport du jury étaye largement ce point.

18. Il ne faut pas pousser à l'extrême cette remarque ; d'une part j'ai plutôt en tête le cas des épreuves orales (si vous ne savez traiter qu'un cas particulier d'une question à l'écrit, ce n'est pas forcément une mauvaise chose, à condition de s'en rendre compte et de le dire, mais ça doit rester un évènement rare, qui ne sera valorisé que si la question est particulièrement difficile), et bien sûr cela ne doit pas vous exempter de connaître l'étendue du programme.

19. Même si les titres de leçons (et le rapport du jury qui les accompagne) limitent volontairement les débordements.

choisit de mettre dans une leçon). La réponse la plus courte est assez simple, c'est non. Mais il me faut étayer un peu plus pour guider le lecteur dans son apprentissage. Lorsqu'il s'agit des théorèmes centraux du programme, je pense que la priorité dans leur apprentissage n'est pas d'en maîtriser la démonstration, parfois subtile [20] [21], mais bien de savoir se demander pourquoi cet énoncé est important, à quel problématique il répond, à quels situations et exemples il peut s'appliquer, et quels sont les contre-exemples qu'on peut fournir quand on enlève l'une ou l'autre des hypothèses, etc. Ce sont les premières étapes de leur apprentissage, à mon avis. À ce titre, à quelques exceptions près qu'on pourra discuter dans ce document, tout énoncé pour lesquels on ne sait pas répondre à ces questions, devrait probablement être retiré de sa hotte d'agrégatif (pour toujours, ou jusqu'à ce qu'on trouve des réponses convaincantes). Quant aux démonstrations assez fines, certainement il peut être instructif et bénéfique de les travailler, mais dans de nombreux cas, on se rendra compte qu'elles nous "dépassent" un peu, au sens où probablement on n'aurait pas trouvé l'idée nous-même de la démonstration. Et donc je pense qu'il n'est pas du tout problématique, dans certains cas [22], de ne pas savoir démontrer un théorème central du programme, car la priorité restera de le connaître et de le comprendre, ce qui est somme toute assez différent. *Par contre*, un cours est parsemé de "propositions" classiques [23], dont les démonstrations doivent être vues comme des exercices importants et qui pourront guider le lecteur à savoir s'il a maîtrisé ou non les outils mis en jeu [24]. À ce titre, s'il vous plaît, passez un temps autrement plus important à essayer de trouver ces démonstrations par vous-même, qu'à les lire (et donc à avoir l'illusion de les comprendre). **Une démonstration comprise est une démonstration qu'on a faite soi-même** [25].

Pour finir sur ce point, je pense que les attentes du concours sont assez différentes suivant les sujets abordés. Sur les notions que l'on voit habituellement au niveau L1-L2, le jury aura une forte exigence de maîtrise, de recul, et de manipulation des outils. Il veut à la fois que vous sachiez démontrer les résultats de base (il peut brutalement [26] en demander une démonstration à l'écrit ou à l'oral), et que vous ayez une bonne compréhension du cheminement entre les outils (notamment dans l'organisation de votre plan pour l'épreuve orale). Sur les sujets qu'on voit plutôt en L3-M1 (l'intégration de Lebesgue, les fonctions holomorphes, l'analyse fonctionnelle, la théorie des distributions...), les attentes me paraissent plus modestes : ces sujets sont en effet bien plus élaborés, et on attend avant tout que vous sachiez les utiliser, ce qui ne nécessite pas le même recul. L'exemple typique d'une telle situation est l'utilisation du théorème de convergence dominée de Lebesgue [27]. Son énoncé même requiert un

20. Je ne vois pas de meilleur exemple que de citer mon propre oral d'agrégatif, sur la leçon qui s'appelait "Espaces $L^p, 1 \leq p \leq \infty$" (ce titre a été significativement modifié, mais l'exemple garde son sens) : le jury m'a demandé de démontrer que ces espaces étaient complets. C'était évidemment un énoncé de mon plan, et la question était plus que légitime. Mais je ne savais pas le démontrer de but en blanc (c'est pourtant un développement très souvent proposé ; mais je ne l'aimais pas (et ne l'avais pas proposé) parce que justement, je ne considérais pas suffisamment comprendre sa démonstration ; aujourd'hui encore, c'est le seul exemple de preuve de complétude que je connais où l'on passe "forcément" par la caractérisation par les séries "absolument convergentes" et je ne considère pas avoir vraiment compris le pourquoi de cette astuce). Le jury n'a pas été choqué de devoir me guider dans ma réponse, et même si ce dernier a probablement apprécié ma réactivité dans ma réponse, il m'a largement guidé, ce qui ne m'a pas empêché d'obtenir une excellente note.

21. Il y a des exceptions, bien sûr : la preuve du théorème de point fixe de Picard est un exemple de simplicité, et son énoncé complet (faisant référence aux suites récurrentes construites à partir de la fonction considérée) suggère significativement la preuve qui doit suivre et qui ne devrait pas poser problème à un candidat bien préparé au concours de l'agrégation externe.

22. Évidemment, je ne peux pas me lancer dans une liste détaillée des énoncés dont les preuves sont "indispensables" et celles qui peuvent être laissées de côté, d'autant que ça reste subjectif. Le candidat doit rester guidé par sa propre auto-évaluation.

23. Volontairement, je réserverai le mot "Théorème" à un nombre limité d'énoncé, afin de signifier leur importance, et possiblement leur difficulté. Le mot "Proposition" sera utilisé beaucoup plus fréquemment, et dans la quasi-totalité des cas, les preuves seront courtes et accessibles. Il convient de savoir par cœur certaines de ces propositions (mais probablement pas toutes !) car elles sont très fréquemment utilisées, mais si l'étudiant est pris d'un doute sur l'énoncé ou les hypothèses de l'une d'elle (ce qui est naturel, à mon avis, alors que ce sera plus rédhibitoire dans le cas d'un théorème), le candidat devrait pouvoir reproduire la démonstration en un temps limité, s'il suit les conseils donnés ici.

24. Pour revenir à mon oral sur les espaces L^p, le jury m'a demandé de prouver que si $f, g \in L^1(\mathbb{R})$, alors $f * g \in L^1(\mathbb{R})$ avec l'inégalité de norme qui l'accompagne (énoncé qui figurait également dans mon plan) et là par contre, au vu de l'importance et du caractère très classique en théorie de Lebesgue de la méthode employée, je pense qu'il m'aurait tenu rigueur de ne pas savoir répondre du tac au tac.

25. Sans l'avoir apprise par cœur évidemment.

26. J'entends par là, sans question intermédiaire ou sans indication.

27. L'autre exemple flagrant est la théorie des distributions, pour laquelle le jury étaye spécifiquement qu'il n'attend

certain nombre d'éléments de théorie de la mesure, et sa démonstration également. Mais la priorité réside la plupart du temps en son application, et on contourne souvent les subtilités de mesurabilité en sachant simplement qu'une fonction continue (par morceaux) est borélienne. Pour citer un autre exemple, prenons le chapitre d'analyse complexe : autant je pense que les étudiants doivent avoir une connaissance quasi-irréprochable de l'étude des séries entières, qui repose sur plusieurs outils de niveau L1-L2 (critère de convergence de série, notions de convergence des séries de fonctions...), autant il ne sera le plus souvent pas exigé d'avoir un fort recul sur la théorie des fonctions holomorphes. Si on regarde les premières moitiés des épreuves écrites des dernières années, on verra que la plupart du temps les questions sur l'holomorphie reposent seulement sur une utilisation directe d'énoncés du cours [28], sans chercher les subtilités derrière ces énoncés. En conclusion, tout étudiant qui passe un temps très substantiel de sa préparation à comprendre et maîtriser la théorie des fonctions holomorphes, mais qui par ailleurs s'avère fragile sur les séries entières car il les a laissées de côté en pensant que c'était moins important, a toutes les chances de se faire pénaliser durant les épreuves, écrites comme orales. Quant à l'étudiant qui, du fait qu'il prend beaucoup de temps pour réviser la théorie des séries entières, n'a pas le temps de maîtriser la théorie des fonctions holomorphes, il sera plus facilement pardonné de ce manque, en tout cas cela ne l'empêchera pas de réussir le concours.

Conclusion et remerciements

Il va sans dire, ce document sera parsemé d'erreurs (qu'on espère pas trop grosses), et est amené à être amélioré, par mon expérience dans son enseignement, mais surtout par la vigilance et l'enthousiasme des étudiants qui le travailleront. Alors je vous en conjure, plutôt que de pester dans votre coin, n'hésitez pas à m'écrire un mail [29] en me décrivant vos doléances. Si vous considérez que j'ai fait des oublis non négligeables, n'hésitez pas non plus à me suggérer des ajouts. J'en profite donc, en vrac, pour remercier Jean-Baptiste, Redouane, Paul, Guillaume, Sacha, Gilles, Sven, Edouard, Luca, Augustin, Adrien, Laurent, Eli, Maxime, et tous ceux que j'ai malheureusement oublié de mentionner.

J'insiste sur le fait qu'un étudiant agrégatif se doit de connaître (et s'il veut maximiser ses chances et son classement, de respecter) les "règles" du concours ; ces dernières sont précisément décrites dans le rapport du jury que je conseille de lire assidument.

Sur ce, bonne lecture, bon apprentissage, et bon courage !

Paris, le 26 août 2019, mis à jour le 23 août 2022
Jimmy Lamboley

aucune subtilité de construction des topologies adaptées aux espaces fonctionnels considérés, mais qu'il attend surtout que les candidats sachent donner des exemples intéressants de distributions.

28. Par exemple, la démonstration qu'une fonction définie par une série ou une intégrale à paramètre, est une fonction holomorphe, ou encore l'application du théorème de Cauchy pour calculer une intégrale.

29. jimmy.lamboley@imj-prg.fr

Programme d'analyse et probabilité de l'agrégation externe 2022

On reprend la description du programme de 2022 à partir du 6ième onglet, les 5 premiers décrivant le programme d'algèbre.

6 Analyse à une variable réelle

6.1 Nombres réels

Le corps $\mathbb{R}$ des nombres réels. Topologie de $\mathbb{R}$. Sous-groupes additifs de $\mathbb{R}$. Suites de nombres réels : convergence, valeur d'adhérence. Suites récurrentes. Limites inférieure et supérieure. Suites de Cauchy. Complétude de $\mathbb{R}$. Théorème de Bolzano-Weierstrass. Parties compactes de $\mathbb{R}$. Parties connexes de $\mathbb{R}$.

6.2 Séries numériques

Convergence des séries à termes réels. Séries géométriques, séries de Riemann. Séries à termes positifs. Sommation des relations de comparaison. Comparaison d'une série et d'une intégrale. Estimations des restes. Convergence absolue. Produits de séries. Séries alternées.

6.3 Fonctions définies sur une partie de $\mathbb{R}$ et à valeurs réelles

(a) Continuité

Limites, continuité. Théorème des valeurs intermédiaires, image d'un segment. Étude de la continuité des fonctions monotones. Continuité d'une fonction réciproque.

(b) Dérivabilité

Dérivée en un point, fonctions dérivables. Dérivée d'une fonction composée. Dérivée d'une fonction réciproque. Théorèmes de Rolle et des accroissements finis. Etude des variations d'une fonction. Dérivées d'ordre supérieur. Applications de classe C^k, de classe C^k par morceaux. Formule de Leibniz. Formule de Taylor-Young, formule de Taylor avec reste intégral, formule de Taylor-Lagrange. Calcul de développements limités et de développements asymptotiques.

6.4 Fonctions usuelles

Fonctions polynômes, fonctions rationnelles. Logarithmes. Exponentielles. Fonctions puissances. Fonctions circulaires et hyperboliques. Fonctions circulaires et hyperboliques réciproques.

6.5 Intégration

(a) Intégrale sur un segment des fonctions continues par morceaux

Calcul de primitives. Sommes de Riemann. Primitives d'une fonction continue. Méthodes usuelles de calcul d'intégrales. Changement de variable. Intégration par parties.

(b) Intégrales généralisées

Intégrales absolument convergentes. Intégration des relations de comparaison. Intégrales semi-convergentes.

6.6 Suites et séries de fonctions

Convergence simple, convergence uniforme. Continuité et dérivabilité de la limite. Cas des séries de fonctions ; convergence normale.
Théorèmes d'approximation de Weierstrass polynomial et de Weierstrass trigonométrique.

6.7 Convexité

Fonctions convexes d'une variable réelle. Continuité et dérivabilité des fonctions convexes. Caractérisations de la convexité. Inégalités de convexité.

7 Analyse à une variable complexe

7.1 Séries entières

(a) Rayon de convergence. Propriétés de la somme d'une série entiére sur son disque de convergence : continuité, dérivabilité par rapport à la variable complexe, primitives.

(b) Exponentielle complexe ; propriétés. Extension des fonctions circulaires au domaine complexe. Développement en série entiére des fonctions usuelles.

7.2 Fonctions d'une variable complexe

(a) Fonctions holomorphes. Conditions de Cauchy-Riemann. Intégrale d'une fonction continue le long d'un chemin C^1 par morceaux. Primitives d'une fonction holomorphe. Déterminations du logarithme. Théorème d'holomorphie sous le signe intégrale.

(b) Indice d'un chemin fermé C^1 par morceaux par rapport à un point.

(c) Formules de Cauchy. Analyticité d'une fonction holomorphe. Principe des zéros isolés. Principe du prolongement analytique. Principe du maximum.

(d) Singularités isolées. Séries de Laurent. Fonctions méromorphes. Théorème des résidus.

(e) Suites et séries de fonctions holomorphes. Stabilité de l'holomorphie par convergence uniforme.

8 Topologie

8.1 Topologie et espaces métriques

(a) Topologie d'un espace métrique. Topologie induite. Produit fini d'espaces métriques.

(b) Suites. Valeurs d'adhérence. Limites. Applications continues. Homéomorphismes.

(c) Compacité. Équivalence des définitions en termes de valeurs d'adhérence (Bolzano-Weierstrass) ou de recouvrements ouverts (Borel-Lebesgue). Connexité. Composantes connexes. Connexité par arcs.

(d) Applications lipschitziennes, applications uniformément continues. Théorème de Heine.

(e) Espaces métriques complets. Théorème du point fixe pour les applications contractantes.

8.2 Espaces vectoriels normés sur $\mathbb{R}$ ou $\mathbb{C}$.

(a) Topologie d'un espace vectoriel normé. Normes équivalentes. Cas des espaces de dimension finie. Normes $\|\cdot\|_p$ sur $\mathbb{R}^n$ et $\mathbb{C}^n$. Espaces de Banach. Séries absolument convergentes dans un espace de Banach.

(b) Applications linéaires continues, norme d'une application linéaire continue.

(c) Norme de la convergence uniforme. Espace des fonctions continues bornées sur un espace métrique, à valeurs dans un espace de Banach.

(d) Étude de la compacité de parties d'un espace vectoriel normé : Théorème de Riesz, Théorème d'Ascoli.

8.3 Espaces de Hilbert

(a) Projection sur un convexe fermé. Projection orthogonale sur un sous-espace vectoriel fermé.

(b) Dual d'un espace de Hilbert, Théorème de représentation de Riesz. Cas des espaces ℓ^2 et L^2. Bases hilbertiennes (dans le cas séparable). Exemples de bases de polynômes trigonométriques et de polynômes orthogonaux. Théorème de Lax-Milgram.

(c) Espace $H_0^1(]0,1[)$ et application au problème de Dirichlet en dimension un.

9 Calcul différentiel

9.1 Fonctions différentiables

(a) Applications différentiables sur un ouvert de $\mathbb{R}^n$. Différentielle (application linéaire tangente). Dérivée selon un vecteur.

(b) Dérivées partielles. Matrice jacobienne, vecteur gradient, matrice hessienne. Composition d'applications différentiables. Théorème des accroissements finis. Applications de classe C^1.

(c) Applications de classe C^k. Dérivées partielles d'ordre k. Interversion de l'ordre des dérivations. Formule de Taylor-Young, formule de Taylor avec reste intégral.

(d) Étude locale des applications à valeurs dans $\mathbb{R}$. Développements limités. Recherche des extrema locaux, caractérisation de la convexité des fonctions de classe C^1 et C^2 définies sur un ouvert convexe [30] $\mathbb{R}^n$.

(e) Difféomorphismes. Théorème d'inversion locale. Théorème des fonctions implicites.

9.2 Équations différentielles

(a) Équations différentielles de la forme $X' = f(t, X)$ sur $\mathscr{I} \times \Omega$ avec $\mathscr{I}$ intervalle ouvert de $\mathbb{R}$ et Ω ouvert de $\mathbb{R}^n$. Théorème de Cauchy-Lipschitz. Solutions maximales. Lemme de Grönwall. Théorème de sortie de tout compact (théorème « des bouts »).

(b) Cas des équations différentielles autonomes. Portrait de phase, comportement qualitatif. Stabilité des points d'équilibre (théorème de linéarisation).

(c) Systémes différentiels linéaires. Méthode de variation des constantes (fomule de Duhamel). Cas des coefficients constants. Application à la résolution d'équations différentielles linéaires d'ordre supérieur à 1.

9.3 Géométrie différentielle

(a) Sous-variétés de $\mathbb{R}^n$. Définitions équivalentes : graphe local, paramétrisation locale, équation locale. Espace tangent. Gradient. Cas des surfaces de $\mathbb{R}^3$, position par rapport au plan tangent.

(b) Construction de courbes planes définies par une représentation paramétrique. Etude métrique des courbes : abscisse curviligne, longueur d'un arc C^1.

(c) Extrema liés, multiplicateurs de Lagrange.

10 Calcul intégral

10.1 Notions de théorie de la mesure

Définition des espaces mesurables, tribu produit, cas particulier des tribus boréliennes. Définition d'une mesure positive, cas particuliers de la mesure de comptage, de la mesure de Lebesgue (construction admise) et des mesures de probabilité. Définition d'une mesure produit (construction admise).

30. Je recopie la faute (l'oubli du mot "de") que j'ai trouvée dans le programme officiel

Définition des fonctions mesurables, approximation par des fonctions étagées.

10.2 Intégration

(a) Intégrale des fonctions mesurables positives, Théorème de convergence monotone. Lemme de Fatou. Fonctions intégrables, théorème de convergence dominée.

(b) Fonctions intégrables à valeurs dans un espace vectoriel de dimension finie. Continuité, dérivabilité des intégrales à paramètres.

(c) Espaces L^p, où $1 \leq p \leq \infty$. Complétude. Inégalité de Hölder

(d) Théorème de Fubini. Changement de variables dans une intégrale multiple. Cas des coordonnées polaires, cas des coordonnées sphériques.

(e) Convolution. Régularisation et approximation par convolution.

10.3 Analyse de Fourier

(a) Séries de Fourier des fonctions localement intégrables périodiques d'une variable réelle. Lemme de Riemann-Lebesgue. Produit de convolution de fonctions périodiques. Théorèmes de Dirichlet, de Fejer et de Parseval.

(b) Transformation de Fourier sur les espaces $L^1(\mathbb{R}^d)$ et $L^2(\mathbb{R}^d)$. Théorème de Plancherel.

11 Probabilités

11.1 Définition d'un espace probabilisé

Evénements, tribus, mesure de probabilité. Indépendance d'événements et de tribus. Loi du 0-1, lemmes de Borel-Cantelli. Probabilités conditionnelles. Formule des probabilités totales.

11.2 Variables aléatoires, loi d'une variable aléatoire

(a) Loi discrète, loi absolument continue. Fonction de répartition et densité. Loi conjointe de variables aléatoires, indépendance de variables aléatoires. Espérance et variance d'une variable aléatoire à valeurs réelles, théorème de transfert. Moments. Exemples de lois : loi de Bernoulli, binomiale, géométrique, de Poisson, uniforme, exponentielle, de Gauss.

(b) Fonction caractéristique. Fonction génératrice d'une variable aléatoire à valeurs dans $\mathbb{N}$. Application aux sommes de variables aléatoires indépendantes.

11.3 Convergences de suites de variables aléatoires

(a) Convergence en probabilité, dans L^p, presque sûrement, en loi. Inégalité de Markov, inégalité de Bienaymé-Tchebychev, Théorème de Lévy.

(b) Loi faible et loi forte des grands nombres. Théorème central limite.

12 Calcul au sens des distributions et applications

[31]

En dimension $d > 1$, on considére comme admise la formule « d'intégration par parties »

$$\int_\Omega \partial_{x_j} f(x)dx = \int_{\partial\Omega} f(x)\nu_j(x)d\sigma(x)$$

sur un domainé $\Omega \subset \mathbb{R}^d$, de frontière $\partial\Omega$ « suffisamment régulière », avec $d\sigma$ la mesure de Lebesgue sur $\partial\Omega$ et $\nu(x)$ le vecteur unitaire extérieur en $x \in \partial\Omega$. Il est en revanche attendu une certaine familiarité

31. Il s'agit du principal changement des dernières années : auparavant seules les distributions tempérées étaient au programme, désormais les distributions "tout court" le sont aussi.

dans la manipulation de telles formules, par exemple dans le cas où Ω est une boule de $\mathbb{R}^d$.

12.1 Distributions sur $\mathbb{R}^d$

(a) Espace vectoriel des fonctions C^∞ à support compact. Stabilité par dérivation. Stabilité par multiplication par une fonction C^∞. Partitions de l'unité. Construction d'approximations de fonctions et densité dans les espaces de fonctions usuels.

(b) Distributions. Exemples de distributions : fonctions localement intégrables, masse de Dirac et ses dérivées, valeur principale de Cauchy. Principes du calcul par dualité-transposition. Dérivation ; multiplication par une fonction C^∞. Formule des sauts. Suites convergentes de distributions : définition, exemples. Notion de support d'une distribution ; cas des distributions à support ponctuel.

12.2 Espaces $\mathscr{S}(\mathbb{R}^d)$ et $\mathscr{S}'(\mathbb{R}^d)$

(a) Espace de Schwartz $\mathscr{S}(\mathbb{R}^d)$ des fonctions à décroissance rapide ainsi que toutes leurs dérivées. Fonctions gaussiennes et leurs dérivées. Stabilité par dérivation. Stabilité par multiplication par une fonction C^∞ à croissance lente. Transformation de Fourier sur $\mathscr{S}(\mathbb{R}^d)$. Convolution de deux fonctions de $\mathscr{S}(\mathbb{R}^d)$. Multiplication par une fonction C^∞ à croissance lente.

(b) Espace $\mathscr{S}'(\mathbb{R}^d)$ des distributions tempérées : formes linéaires T sur $\mathscr{S}(\mathbb{R}^d)$ telles qu'il existe $C > 0$ et $k \in \mathbb{N}$ vérifiant $|\langle T, \varphi\rangle| \leq C \sup\{|x^\alpha \partial^\beta \phi(x)|, x \in \mathbb{R}^d, |\alpha| \leq k, |\beta| \leq k\}$ pour tout $\phi \in \mathscr{S}(\mathbb{R}^d)$. Exemples de distributions tempérées : fonctions L^2 et représentation de Riesz, fonctions L^p, cas des fonctions périodiques, peigne de Dirac. Dérivation des distributions tempérées ; multiplication par une fonction C^∞ à croissance lente. Convolution d'une distribution tempérée avec une fonction de $\mathscr{S}(\mathbb{R}^d)$.

(c) Transformation de Fourier dans $\mathscr{S}'(\mathbb{R}^d)$. Formule d'inversion. Transformation de Fourier et dérivation, Transformée de Fourier d'un produit de convolution.

12.3 Applications

Calcul de dérivées et de transformée de Fourier de distributions. Formule de Poisson (dimension un).

Utilisation de la convolution et de la transformée de Fourier-Laplace pour la résolution d'équations différentielles linéaires en dimension un. Notion de solution élémentaire d'opérateurs différentiels à coefficients constants (cas du laplacien).

Notion de solution faible d'équations aux dérivées partielles linéaires : application, par exemple, à la résolution des équations de Laplace, de la chaleur, des ondes.

13 Méthodes numériques

13.1 Résolution de systèmes d'équations linéaires

Notion de conditionnement. Théorème de Gershgörin-Hadamard. Pivot de Gauss, décomposition LU. Méthodes itératives (par exemple méthode de Jacobi, méthode de Gauss-Seidel) ; analyse de convergence : normes subordonnées, rayon spectral.

Décomposition en valeurs singulières.

Exemple de la matrice de discrétisation par différences finies du laplacien en dimension un.

13.2 Méthodes itératives de résolution approchée d'équations réelles et vectorielles

Cas des systémes linéaires : méthodes itératives. Recherche d'éléments propres : méthode de la puissance.

Optimisation de fonctions convexes en dimension finie, méthode du gradient à pas constant, moindres carrés.

Problèmes non linéaires réels et vectoriels : méthode de dichotomie, méthode de Picard, méthode de Newton, vitesse de convergence et estimation de l'erreur.

13.3 Intégration numérique

Méthode des rectangles, estimation de l'erreur. Méthode de Monte-Carlo : vitesse de convergence, application au calcul d'intégrales multiples.

13.4 Approximation de fonctions numériques

Interpolation de Lagrange : polynôme de Lagrange d'une fonction en $(n+1)$ points, estimation de l'erreur.

13.5 Équations différentielles ordinaires

Aspects numériques du problème de Cauchy : méthode d'Euler explicite, consistance, stabilité, convergence, ordre.

13.6 Transformée de Fourier

Transformée de Fourier discrète sur un groupe abélien fini. Transformée de Fourier rapide.

Liste des leçons d'analyse et probabilités 2022

201 Espaces de fonctions. Exemples et applications.
203 Utilisation de la notion de compacité.
204 Connexité. Exemples et applications.
205 Espaces complets. Exemples et applications.
207 Prolongement de fonctions. Exemples et applications.
208 Espaces vectoriels normés, applications linéaires continues. Exemples.
209 Approximation d'une fonction par des fonctions régulières. Exemples et applications.
213 Espaces de Hilbert. Bases hilbertiennes. Exemples et applications.
214 Théorème d'inversion locale, Théorème des fonctions implicites. Exemples et applications en analyse et en géométrie.
215 Applications différentiables définies sur un ouvert de $\mathbb{R}^n$. Exemples et applications.
219 Extremums : existence, caractérisation, recherche. Exemples et applications.
220 Équations différentielles ordinaires. Exemples de résolution et d'étude des solutions en dimension 1 et 2.
221 Équations différentielles linéaires. Systèmes d'équations différentielles linéaires. Exemples et applications.
222 Exemples d'études d'équations différentielles linéaires et d'équations aux dérivées partielles linéaires.
223 Suites numériques. Convergence, valeurs d'adhérence. Exemples et applications.
226 Suites vectorielles et réelles définies par une relation de récurrence $u_{n+1} = f(u_n)$. Exemples. Applications à la résolution approchée d'équations.
228 Continuité et dérivabilité des fonctions réelles d'une variable réelle. Exemples et applications.
229 Fonctions monotones. Fonctions convexes. Exemples et applications.
230 Séries de nombres réels ou complexes. Comportement des restes ou des sommes partielles des séries numériques. Exemples.
234 Fonctions et espaces de fonctions Lebesgue-intégrables.
235 Problèmes d'interversion de limites et d'intégrales.
236 Illustrer par des exemples quelques méthodes de calcul d'intégrales de fonctions d'une ou plusieurs variables.
239 Fonctions définies par une intégrale dépendant d'un paramètre. Exemples et applications.
241 Suites et séries de fonctions. Exemples et contre-exemples.
243 Convergence des séries entières, propriétés de la somme. Exemples et applications.
245 Fonctions d'une variable complexe. Exemples et applications.
246 Séries de Fourier. Exemples et applications.
250 Transformation de Fourier. Applications.
253 Utilisation de la notion de convexité en analyse.
261 Loi d'une variable aléatoire : caractérisations, exemples, applications.
262 Convergences d'une suite de variables aléatoires. Théorèmes limite. Exemples et applications.
264 Variables aléatoires discrètes. Exemples et applications.
265 Exemples d'études et d'applications de fonctions usuelles et spéciales.
266 Illustration de la notion d'indépendance en probabilités.
267 Exemples d'utilisation de courbes en dimension 2 ou supérieure.

Pour les docteurs, la liste réduite en analyse et probabilités est constituée des leçons : **203, 208, 213, 214, 215, 219, 220, 222, 226, 228, 234, 235, 236, 241, 245, 246, 250, 262, 265, 266**.

Petit historique (très partiel) : Chaque année, il peut y avoir disparition et/ou apparition d'une ou deux leçons. Certains titres sont également modifiés. Regardons les changements les plus récents : [32]

- Quelques titres ayant disparu ces dernières années (et qui pourront réapparaître un jour) :
  202 Exemples de parties denses et applications (disparition pour la session 2020)
  218 Applications des formules de Taylor, (disparition pour la session 2019)
  224 Exemples de développements asymptotiques de suites et de fonctions (disparition pour la session 2020)
  260 Espérance, variance et moments d'une variable aléatoire (disparition pour la session 2020)
  263 Variables aléatoires à densité. Exemples et applications
- Les leçons **266, 267** sont apparues pour la session 2020, **265** en 2019.
- Quelques modifications de titre :
  Dans **209**, il y avait "polynômes et polynômes trigonométriques" à la place de "fonctions régulières"
  222 était précédemment "Exemples d'équations aux dérivées partielles linéaires"
  Dans **228**, il y a eu l'ajout de la mention "dérivation faible" pour les sessions 2020 et 2021 [33], puis son retrait pour la session 2022,
  234 était précédemment "Espaces L^p, $1 \leq p \leq +\infty$",
  245 était précédemment "Fonctions holomorphes sur un ouvert de $\mathbb{C}$. Exemples et applications"
- La leçon **233** "Analyse numérique matricielle : résolution approchée de systèmes linéaires, recherche de vecteurs propres, exemples" a été modifiée mais surtout transférée dans la liste des leçons d'Algèbre et Géométrie, et est devenue **149** "Valeurs propres, vecteurs propres. Calculs exacts ou approchés d'éléments propres. Applications."

Quelques titres plus anciens tirés de la liste 2010 :
206 Théorèmes de point fixe. Exemples et applications.
216 Étude métrique des courbes. Exemples.
217 Sous-variétés de $\mathbb{R}^n$. Exemples.
232 Méthodes d'approximation des solutions d'une équation F(X)=0. Exemples.
238 Méthodes de calcul approché d'intégrales et de solutions d'équations différentielles.
242 Utilisation en probabilités du produit de convolution et de la transformation de Fourier ou de Laplace.
250 Suite de variables aléatoires de Bernoulli indépendantes.
251 Indépendance d'événements et de variables aléatoires. Exemples.
252 Loi binomiale. Loi de Poisson. Applications.
254 Espaces de Schwartz $\mathscr{S}'(\mathbb{R}^d)$ et distributions tempérées.
255 Dérivation au sens des distributions. Exemples et applications.
256 Transformation de Fourier dans $\mathscr{S}(\mathbb{R}^d)$ et $\mathscr{S}'(\mathbb{R}^d)$
et d'autres tirés de la liste de 2005 :
208 Utilisation de la continuité uniforme en analyse
209 Utilisation de la dénombrabilité en analyse et probabilités
211 Utilisation de la dimension finie en analyse
212 Méthodes hilbertiennes en dimension finie et infinie
217 Etude locale de surfaces. Exemples.
237 Méthodes de calcul de valeurs approchées d'intégrales.
241 Exemples d'utilisation de fonctions définies par une série.
244 Fonctions holomorphes et méromorphes sur un ouvert de $\mathbb{C}$.

32. Notons aussi que la liste des leçons pour la session 2023 devrait être diffusée en même temps que le rapport du jury de la session 2022, soit autour de novembre 2023.
33. Donc en réalité seulement pour une session d'oraux puisqu'il n'y a pas eu d'oraux en 2020.

Descriptif des cours

— **Cours 1 à 5 : Chapitre I**

1. Existence admise de l'ensemble des réels, notion de borne supérieure, densité de $\mathbb{Q}$ dans $\mathbb{R}$, utilisation de $\mathbb{R} \cup \{\pm\infty\}$, sous-groupes de $\mathbb{R}$. Théorèmes de la limite monotone et de Bolzano-Weierstrass.
2. Séries réelles, séries de Riemann et théorèmes de comparaison entre séries à termes positifs. Résultat fondamental [convergence absolue $\Rightarrow$ convergence], avec deux preuves importantes. Quelques éléments sur les séries semi-convergentes. Théorème sur le produit de Cauchy et théorème de Fubini pour les séries doubles.
3. Fonctions réelles d'une variable réelle : définition de limite, de continuité, de dérivabilité. Théorème des valeurs intermédiaires, théorème des bornes, théorème de Rolle, égalité des accroissements finis. Les résultats sur les opérations sont abordés, on détaille surtout le cas de l'inverse. On introduit aussi l'uniforme continuité, et on montre le théorème de Heine.
4. Fonctions de classe C^n, formules de Taylor. On évoque rapidement l'étude des fonctions classiques. Etude des suites et séries de fonctions. On montre [convergence normale $\Rightarrow$ convergence uniforme] pour les séries de fonctions, et on étudie le comportement de l'intégration et la dérivation pour la convergence uniforme.
5. Construction de l'intégrale de Riemann des fonctions continues par morceaux. On évoque l'espace des fonctions réglées. On montre que le théorème fondamental de l'analyse. On évoque les intégrales à paramètre (on renvoit au Chapitre IV pour les résultats plus performants) et l'intégration numérique. On conclut par quelques notions sur les intégrales impropres.

— **Cours 6 à 10 : Chapitre II**

1. Définition et manipulation des nombres complexes, différences entre $\mathbb{R}$ et $\mathbb{C}$. Quelques éléments de topologie dans $\mathbb{C}$, nous repoussons au Chapitre III les démonstrations. Etude des séries entières (rayon de convergence, $\mathbb{C}$-dérivabilité dans le disque ouvert de convergence).
2. Etude de la fonction exponentielle définie par une série entière, et les fonctions qui s'en déduisent (fonctions trigonométriques, Logarithme principal). Etude des fonctions développables en série entière/analytique, différence entre fonction C^∞ et analytique dans le cas réel. Définition de l'holomorphie, exemples.
3. Enoncé et preuve des équations de Cauchy-Riemann. Notion d'intégrale sur un chemin, indice. Caractérisisation de l'existence d'une primitive, énoncé du théorème et de la formule de Cauchy dans le cas d'un ouvert étoilé ; on montre le second à partir du premier, mais on admet pour le moment le théorème de Cauchy. On en déduit le caractère analytique des fonctions holomorphes.
4. Conséquences de la formule de Cauchy, et théorème d'holomorphie des intégrales à paramètre. Etude les singularités d'une fonction holomorphe en un point, développement en série de Laurent et théorème des résidus.
5. Cours facultatif où on aborde des questions plus avancées sur les fonctions holomorphes : démonstration complète du théorème de Cauchy local, théorie de l'homotopie et théorème de Cauchy global, notion d'ensemble simplement connexe. Preuve de l'existence d'un développement en série de Laurent pour une fonction holomorphe sur un anneau et on conclut par quelques directions de travail.

— **Cours 11 à 15 : Chapitre III**

1. Définitions relatives aux espaces métriques, que ce soit sur les ensembles, les suites ou les fonctions.
2. Notion de complétude dans les espaces métriques, avec ses applications fondamentales, comme le théorème de prolongement des applications uniformément continues, et le théorème de point fixe de Picard. Notion de compacité, avec équivalence entre la définition topologique et la définition séquentielle. Généralisation du théorème des bornes et du théorème de Heine vus au Chapitre I.
3. Notion de connexité et de connexité par arcs. Etude des espaces vectoriels normés en reprenant toutes les notions du paragraphe sur les espaces métriques dans ce cas particulier.
4. On revient sur la compacité dans le cas des e.v.n.. On clarifie le cas des espaces de dimension finie, et on prouve le théorème de Riesz de non-compacité de la boule unité dans un e.v.n. de dimension infinie. On étudie ensuite l'espace des applications linéaires continues entre deux e.v.n., avec le cas particulier de la dimension finie et ses applications à l'analyse numérique matricielle.
5. Etude de l'espace des fonctions continues sur un segment de $\mathbb{R}$, avec le théorème de Weierstrass et le théorème d'Ascoli. On discute les possibles extensions à des espaces de départ et d'arrivée plus généraux. On conclut le chapitre par des suggestions d'approfondissement.

— **Cours 16 à 20 : Chapitre IV**

1. Quelques éléments sur la dénombrabilité, notion de tribu, exemple de tribu borélienne, définition d'une mesure (positive).
2. Propriétés classiques des mesures, définition de la mesure de Lebesgue (existence, unicité, et caractérisations admises, quelques éléments de référence sont donnés en fin de chapitre). On évoque quelques ensembles pathologiques et on étudie la notion de fonctions mesurables et de fonctions étagées.
3. Lemme d'approximation des fonctions mesurables par les fonctions étagées, construction rapide de l'intégrale de Lebesgue. Théorème de Beppo Levi, inégalité de Markov et ses conséquences, comparaison avec l'intégrale de Riemann et de Lebesgue sur un segment. On conclut avec l'interversion somme et intégrale dans le cas de fonctions mesurables positives, puis le lemme de Fatou.
4. Théorème de convergence dominée, interversion série-intégrale, théorèmes sur les intégrales à paramètre. Notion de tribu produit et de mesure produit. On passe vite sur les détails techniques, et on admet le théorème de construction d'une mesure produit. On en déduit les théorèmes de Fubini-Tonelli et Fubini. On conclut le cours par la notion de mesure image et le théorème de transfert.
5. On énonce sans démonstration le théorème de changement de variable. Étude des espaces L^p pour $p \in [1, +\infty]$, de la convolution, et théorèmes de densité (de C_c^0 est C_c^∞), en donnant des éléments de démonstration, avec la notion d'identité approchée et de régularisation par convolution.

— **Cours 21 à 25 : Chapitre V** par M. Martineau

— **Cours 26 à 28 : Chapitre VI** par M. Dupuy

— **Cours 29 à 32 : Chapitre VII**

1. Définition d'une fonction différentiable, comparaison avec la notion de dérivée directionnelle, exemples avec différentes stratégies de calcul. On aborde la notion de dérivée partielle et de calcul en coordonnées, et on évoque rapidement la notion de différentielle partielle en prévision du théorème des fonctions implicites.
2. Théorème des accroissements finis et ses principaux corollaires. Etudes des fonctions de classe C^1, et introduction aux différentielles d'ordre 2, théorème de Schwarz (sans démonstration). Différentiabilité de l'inverse d'une fonction.

3. Théorème d'inversion locale, théorème de fonctions implicites. On poursuit avec la notion de sous-variété de $\mathbb{R}^n$. On détaille peu les démonstrations, mais on explique les conditions en faisant le lien avec l'algèbre linéaire, et on détaille des exemples.
4. Caractérisations de la convexité d'une fonction avec le calcul différentiel, conditions d'optimalité (nécessaire, suffisante, d'ordre 1, d'ordre 2), sans et avec contrainte d'égalité (théorème des multiplicateurs de Lagrange).

— **Cours 33 à 37 : Chapitre VIII**

1. Etude des Séries de Fourier, lien entre régularité d'une fonction et comportement asymptotique des coefficients de Fourier. Théorème de Féjer, Théorème de Dirichlet, théorème de convergence uniforme de la série de Fourier.
2. Cadre $L^2(\mathbb{T})$ des séries de Fourier et théorème de Parseval. Étude de la transformation de Fourier dans le cadre de l'espace $L^1(\mathbb{R})$: théorème d'inversion de Fourier, et liens entre les propriétés de f et de $\widehat{f}$.
3. Transformation de Fourier dans les cadres $L^2(\mathbb{R})$ et $\mathscr{S}(\mathbb{R})$. On évoque sans trop de détails le cas des mesures finies et son importance en théorie des probabilités.
4. Définition d'une distribution, injection $L^1_{loc}(I) \subset \mathcal{D}'(I)$. Dérivation au sens des distributions, cas des fonctions C^1 par morceaux.
5. Distributions tempérées, transformée de Fourier des distributions tempérées.

— **Cours 35 à 38 : Chapitre IX**

1. Espaces préhilbertiens, orthogonalité, espaces de Hilbert, théorème de projection et théorème de Riesz sur le dual d'un espace de Hilbert.
2. Théorème de Lax-Milgram, bases hilbertiennes (existence et théorème de Parseval), exemples des polynômes orthogonaux, échantillonage de Shannon.
3. Espaces $H^1(]a,b[)$ (définitions équivalentes, injection dans C^0, produit) et H^1_0. Résolution du problème de Dirichlet sur un segment via le théorème de Riesz et le théorème de Lax-Milgram.
4. [*]Cours facultatif sur le théorème de Baire et ses conséquences principales : continuité sur un ensemble dense des limites simples de fonctions continues, théorème de Banach-Steinhaus, théorème de l'application ouverte, théorème du graphe fermé.

— **Cours 39 à 42 : Chapitre X**

1. Définitions de base, théorème de Cauchy Lipschitz local, propriétés d'unicité pour le problème de Cauchy, lemme de Gronwall.
2. Notions de solution globale, maximale, alternative d'explosion, condition d'existence de solution globale. Méthodes de résolution explicite des EDO.
3. Cas des équations différentielles linéaires, cas à coefficients constants. Etude des équations autonomes, théorème de linéarisation.

Chapitre 0

Conseils de travail

1 Préparer les épreuves de l'agrégation

On donne ici quelques remarques sur la nature des épreuves de l'agrégation externe de mathématiques, et quelques conseils sur la façon de les préparer. On s'appuiera sur des exemples liées aux épreuves d'Analyse et Probabilités, mais tous les conseils valent également pour les épreuves de Mathématiques Générales et d'Algèbre et Géométrie.

Certains conseils sont certainement *discutables et personnels*, c'est bien sûr à vous de décider la façon dont vous organisez votre travail. Néanmoins on étaye ces conseils par des extraits du rapport de jury de la session 2017 (je vous laisse vous reporter aux versions plus récentes du rapport).

1.1 Remarques générales

> *“Le jury insiste sur le fait qu'il convient, de manière générale, d'***assurer des bases solides et de faire preuve de la maîtrise de ces bases***, que ce soit*
>
> – *par une rédaction efficace et convaincante à l'écrit,*
> – *par des leçons calibrées et soignées*
> – *ou une présentation à l'épreuve de modélisation*
>
> *qui inspirent toute confiance quant à la solidité de ces bases.” (première page du rapport 2017)*

C'est peut-être le point qui me semble le plus important pour commencer : ce qui est attendu à l'agrégation n'est pas de connaître ou de manipuler des mathématiques très élaborées (du niveau Master ou supérieur), mais de témoigner d'une très grande solidité sur les “bases”, c'est-à-dire sur le programme de niveau L1-L2-L3. Que ce soit à l'écrit ou à l'oral, il faut savoir être irréprochable sur des notions de niveau L1 voire L2, et être performant sur des notions plus difficiles de L3, savoir comment les choses s'articulent.

A titre d'exemple, la première question de l'épreuve d'analyse 2017 consiste à démontrer que l'application

$$A \in \mathcal{M}_d(\mathbb{R}) \mapsto \sup_{x \in \mathbb{R}^d \setminus \{0\}} \frac{\|Ax\|_2}{\|x\|_2}, \tag{1.1}$$

est bien définie de $\mathcal{M}_d(\mathbb{R})$ dans $\mathbb{R}_+$ et que c'est une norme sur $\mathcal{M}_d(\mathbb{R})$.

Le rapport note sur cette question :

> *“Cette question a été particulièrement discriminante. La notion de borne supérieure est mal maîtrisée par un trop grand nombre de candidats. Pour montrer que l'application est bien définie, il faut montrer que le supremum de la quantité scalaire définie en (1.1) est fini. Ce point a été mal compris, bien que faisant partie d'un des fondamentaux de la topologie en dimension finie. Dans le même esprit, la définition précise d'une norme doit être connue.”*

Le problème relevé ici est de deux natures : en premier lieu, une partie des notions mises en œuvre ici relève du programme de L1-premier semestre, à savoir la définition d'une borne supérieure (et comment les manipuler) et la définition même d'une application (on doit vérifier que l'espace d'arrivée est le bon). Le second point est que cette question fait partie du programme de L2, donc elle a été traitée explicitement en cours, probablement plusieurs fois dans votre cursus universitaire, avec différents degrés de généralité et de complexité (la norme ici mentionnée est une norme dite subordonnée, et la construction est valable dans $\mathcal{L}(E, F)$ l'ensemble des applications linéaires continues entre deux e.v.n., ce qui se voit au niveau L3).

Autrement dit, l'évaluation de l'agrégation n'est pas la même que celle de l'université : vous avez pu apprendre des notions difficiles et profondes, notamment au niveau M1, et on vous a évalué sur la base de votre intuition sur ces notions, et de votre capacité à les mettre en œuvre dans des situations complexes. Cette année, vous serez en priorité évalué sur votre capacité à enseigner *"sur le créneau bac-3/bac+3"*, et donc on veut vous voir capable de prendre du recul et d'être très précis dans la manipulation de notions moins élaborées que ce que vous avez pu apprendre jusqu'à présent.

Mon premier conseil est donc de retravailler vos cours de base (depuis le L1) si vous constatez des manques à ce niveau-là :

– quand vous ferez un écrit blanc (ou quand vous travaillerez sa correction), mettez un point d'honneur dès le début de l'année à obtenir une rédaction parfaite pour les premières questions du sujet. À court terme, ce n'est pas forcément la stratégie optimale pour avoir la meilleure note (c'est loin d'être la moins bonne également !), mais à plus long terme (d'ici les épreuves écrites), ça me semble le meilleur moyen de progresser, et de la meilleure manière, c'est-à-dire en consolidant les notions 'dans l'ordre'. Notez que le sujet 2017 évoqué précédemment à pour vocation de manipuler des objets élaborés (à savoir l'exponentielle dans des e.v.n. de dimension infinie), mais que les deux premières parties du sujet (les deux autres parties étant abordées par très peu de candidats) se "cantonnent" à la dimension finie, et donc essentiellement au programme de L2.

– une autre façon de mettre en œuvre la révision de vos bases et de les tester est aussi la préparation de l'oral, qui doit être commencée dès le début de l'année ! Car les leçons couvrent tout le programme du L1 au L3 (avec de rares excursions vers le M1), parfois de façon transverse, et que c'est l'occasion de voir les notions de façon différente de ce dont vous avez l'habitude. Les leçons d'oral sont à mon avis un très bon moyen de tester votre recul et de déceler vos lacunes. Commencez donc au plus tôt à travailler cette épreuve.

Pour enfoncer le clou, concluons par un autre extrait du rapport :

> *"[...] le concours est exigeant et nécessite une préparation spécifique, un entraînement pour être capable de convaincre de sa solidité sur les bases du programme et non de sa dextérité sur des notions très pointues"* (page 11, en commentaire du taux de réussite assez faible des docteurs).

1.2 Déroulement de l'épreuve d'oral (AP ou MG) : la "leçon"

Le jour de l'épreuve, vous tirez un couplage (deux sujets) parmi une liste qui est connue à l'avance, et qui contient 36 leçons en AP et 35 en MG (voir les rapports pour les listes de leçons de l'agrégation Docteur, qui est une sous-partie de l'union des deux listes AP et MG).

> *"Quelques rares candidats prennent des libertés quant au libellé de la leçon et s'en écartent notablement : les titres des leçons définissent un champ précis qu'il faut traiter. Le hors-sujet est lourdement sanctionné"*

Vous avez 3h de préparation, avec un accès à la bibliothèque de l'agreg, ainsi que vos propres ouvrages (pas de polycopiés). L'épreuve dure 55 minutes, et se déroule en 3 temps que nous listons ci-dessous.

1.2.1 Le Plan et sa présentation

Pendant la préparation, vous devez rédiger un document manuscrit de maximum 3 pages (A4 avec des marges de 1cm) qui sera photocopié et distribué au jury (donc n'utilisez pas de couleur).

"Il est recommandé de soigner la présentation du plan écrit, de mettre des titres, d'encadrer les formules, etc. pour qu'il soit le plus lisible possible. En particulier il est vain de vouloir écrire petit dans l'espoir de placer plus de contenu ; on perd en clarté et le jury n'est pas disposé à utiliser une loupe. Les plans peuvent être complétés par une quatrième page consacrée aux figures (et exclusivement à celles-ci). Sur ce document écrit, le candidat doit clairement faire apparaître les résultats qu'il estime significatifs, représentatifs de la leçon et qu'il propose de développer devant le jury."

Le début de l'épreuve consiste en la défense du plan :

"Le candidat est convié à utiliser son temps de parole, 6 minutes maximum, pour **présenter, argumenter, mettre en valeur et faire une synthèse** *de sa leçon. [...]* **le plus important est que le plan soit bien structuré, maîtrisé par le candidat et qu'y figure une quantité substantielle d'exemples et de mathématique***. On attend du candidat qu'il explique pourquoi il adopte cette présentation. En fait, pour cette partie de l'épreuve, le candidat devrait s'imaginer dans la situation où il doit introduire à un auditoire, pendant 6 minutes, une leçon destinée ensuite à être développée sur plusieurs séances. Quel est l'intérêt du sujet ? Comment se positionne-t-il dans un paysage mathématique plus large ? Comment s'articulent les différentes sections qui composent la leçon ? Comment s'expliquent et se motivent les enchaînements ? C'est l'occasion de s'interroger sur les difficultés didactiques de la leçon, c'est-à-dire dans quel ordre et comment présenter les choses pour que le tout soit cohérent, compréhensible et pédagogiquement efficace. En quoi les exemples sélectionnés s'avèrent-ils pertinents ? Quelles figures illustrent particulièrement les notions en jeu ? Autrement dit, le jury attend une argumentation synthétique de la construction de la leçon.*

Il s'agit d'une épreuve orale. *Le document écrit transmis au jury se justifie pour servir de base pour la discussion et constitue un fil conducteur qui guide le jury pour mener la partie consacrée au dialogue et aux questions. Ce plan ne doit être ni une énumération d'énoncés, ni un cours ou un exposé complet avec développement des démonstrations. En revanche, il doit définir avec suffisamment de précision les notions mathématiques introduites, donner les énoncés complets (notamment les hypothèses) des résultats fondamentaux,* **citer des exemples et des applications**. **Le plan calibre l'ensemble de la leçon, il détermine l'orientation choisie par le candidat** *et est par conséquent très important. Les choix doivent être motivés lors de la présentation et éventuellement lors de la discussion avec le jury. Il est nécessaire d'être en mesure de démontrer les résultats importants*[a] *et de pouvoir les illustrer avec des exemples simples.[...]*
Il est contre-productif de chercher à remplir *à tout prix les 3 feuilles autorisées, surtout* **avec des éléments que le candidat ne maîtrise manifestement pas***.*
[...] Le jury attache une grande importance à la maîtrise du plan qui intervient de manière substantielle dans la grille de notation. Le plan ne peut être considéré comme maîtrisé si l'exposé ressemble à une récitation. Les résultats exposés doivent être compris ainsi que l'organisation d'ensemble de la leçon. Il est souhaitable que le candidat connaisse dans leurs grandes lignes les démonstrations des résultats figurant au programme du concours. Il est indispensable que le candidat ait une idée assez claire de la difficulté des démonstrations des résultats qu'il évoque dans son plan. **Le jury rappelle qu'il n'est pas nécessaire de s'aventurer au-delà des contours du programme pour prétendre à une très bonne note ! Plus généralement, il vaut mieux se cantonner à un niveau où l'on se sait solide plutôt que de chercher à placer des énoncés débordant du programme mais sans le recul, ni l'assurance technique nécessaires.** *Bien entendu, réussir cette partie de l'épreuve ne s'improvise pas et le discours doit avoir été réfléchi durant la préparation de l'épreuve, et plus largement* **tout au long du parcours qui conduit les candidats jusqu'au concours.***"*

a. Attention ; je ne pense pas qu'il faut prendre cette remarque au sens strict : il est d'ailleurs dit un peu plus loin "souhaitable que le candidat connaisse dans leurs grandes lignes les démonstrations des résultats", ce qui me parait plus juste. Voir aussi la note 20.

1.2.2 Le développement

Pendant la défense du plan, et dans le plan manuscrit lui-même, vous aurez indiqué deux “développements” au moins. Un développement correspond à un énoncé de votre plan (et donc en rapport avec son titre) que vous proposez de développer pendant 15 min au tableau et sans l'aide de vos notes (mais vous avez le droit de garder votre plan avec vous pendant toute l'épreuve, y compris pendant le développement). Vous aurez a priori défendu vos choix de développements pendant votre défense.

> *“Le jury veille à la cohérence du plan et des propositions de développements. Il est souhaitable que le candidat recherche une adéquation entre le niveau auquel il souhaite se placer et les développements proposés.* **Un candidat ne sera pas avantagé s'il présente un développement non maîtrisé, manifestement mal compris.** *Le niveau de l'agrégation ne peut toutefois se cantonner aux notions abordées dans une classe de Terminale ou une première année post-bac. Des choix de développement de ce niveau sont pénalisés.”*

Un développement est souvent la preuve d'un énoncé important du plan, mais ça peut tout à fait être la résolution d'un exercice que vous jugez intéressant car utilisant la plupart des outils de la leçon, ou ça peut être le développement en profondeur d'un exemple qui s'avère riche et pertinent, notamment parce qu'il illustre les notions de la leçon. En tout cas, il faut rédiger proprement vos démonstrations, et bien jauger la gestion de l'oral et de l'écrit.

> *“Lors du développement, le jury attend du candidat des explications sur la preuve et sur l'utilisation pertinente des notions développées durant l'exposé oral ; il peut être opportun, lors du développement, de se référer explicitement au plan présenté. Trop peu de candidats commencent leur développement par une rapide exposition des grandes idées ou des étapes qui le composent. Expliquer l'approche adoptée, et les difficultés, au début du développement est une démarche pédagogique appréciée. Il est inutile de se précipiter ou de parler trop vite ;* **on veillera au contraire à préciser ses notations, à soigner sa présentation, à placer sa voix et à regarder de temps en temps le jury.** *Le développement ressemble parfois à une succession plus ou moins convaincante de résultats intermédiaires ad hoc.* **La récitation mécanique d'un développement ne peut pas être convaincante ; le jury veille à ce que les futurs enseignants comprennent ce qu'ils exposent et sachent exposer ce qu'ils comprennent.**”

Il est possible qu'au cours de vos 15 min, vous perdiez vos moyens ou la mémoire sur un passage de la démonstration. Dans ce cas, et même si cela peut être pénalisé, vous pouvez demander au jury la permission de regarder vos notes. Cela pourra peut-être vous débloquer et vous éviter de rester 10 min sans rien avoir à dire. **Bien sûr, cela signifie que vous devez rédiger proprement vos développements sur vos notes pendant la préparation**. Ce sera l'occasion de les réviser, de savoir s'il y a des étapes du raisonnement qui vous échappent et nécessitent de la réflexion de votre part.

Vous pouvez aussi passer une étape qui ne vous revient plus, et y revenir à la fin, c'est d'ailleurs sans doute le mieux. Quoi qu'il arrive, il faut savoir communiquer avec le jury et ne pas rester à rien dire.

À mon avis, il faut donc prévoir au minimum 40min de vos 3h de préparation aux développements. On vous pardonnera plus facilement d'avoir un plan un peu court que de ne pas avoir deux développements (ou de ne pas les savoir).

1.2.3 Discussion et questions

Pendant les deux premières parties de l'épreuve (soit un peu plus de 20 min), vous étiez “maître” de votre discours, et sauf si cela s'était avéré nécessaire, le jury avait très peu parlé et ne vous avait pas interrompu. Ensuite, il engage la discussion avec vous.

> *"Le jury ne cherche en rien à déstabiliser le candidat pendant l'épreuve, mais au contraire cherche à valoriser les connaissances du candidat. Les premières questions, parfois très simples, visent souvent à préciser une notation, un point du plan ou du développement, pour que le jury s'assure de la compréhension par le candidat des notions qu'il vient d'exposer et non pour lui tendre un quelconque piège."*

Pendant le début de la discussion, le jury voudra principalement s'assurer que vous avez compris ce que vous avez écrit dans votre plan et/ou votre développement. Il peut pour cela vous demander de préciser un point du développement qu'il a trouvé vague, il peut pointer dans le plan un énoncé erroné et voir si vous arrivez à le corriger, ou enfin vous poser une question dont la réponse est une conséquence élémentaire des résultats de votre plan. S'il constate que vous ne maitrisez manifestement pas votre plan ou votre propre développement, vous serez significativement pénalisés.

La discussion se poursuit ensuite avec le cap fixé par ces premières questions : si vos connaissances s'avèrent cohérentes avec le niveau auquel vous avez placé votre leçon, il pourra poser des exercices de difficulté croissante, et/ou évoquer des parties du programme que vous avez choisi de ne pas aborder. Si par contre il se rend compte que vous en connaissez moins que ce que votre plan ou votre développement laissaient entendre, il posera des questions élémentaires pour tester vos connaissances des bases.

1.2.4 Ce que veut voir le jury (important !)

Comprenez que le jury veut certes s'assurer que vous avez des connaissances de bon niveau, mais il veut avant tout, et ce dans tous les aspects de l'épreuve (et pas seulement la dernière), vous voir réfléchir, vous voir organiser votre pensée, et la communiquer avec clarté. En ce sens, toute tentative de se remémorer un catalogue de données que vous auriez engrengé pendant l'année vous amènera inévitablement à craquer à un moment donné. Le jury ne sera pas dupe, et cherchera à voir d'une manière ou d'une autre si vous êtes capable de réfléchir sur les notions de la leçon, quitte à vous poser des questions très simples.

Pour insister sur ce point, je peux reporter plusieurs témoignages d'étudiants étonnés de leurs notes d'oral. Dans plusieurs cas, les étudiants avaient produit un plan qu'ils considèraient de bon niveau (car appris pendant l'année), et avaient réussi "sans encombre" leur développement, et pourtant ils obtenaient une note très basse, disons inférieure à 5/20. Ils n'arrivaient pas à saisir cette note alors qu'ils considéraient avoir réussi une bonne partie de l'épreuve. Selon moi, dans ces situations, le jury a pu effectivement trouver que plan et développement étaient corrects (même s'il est bien possible qu'il ait décelé des failles ou des erreurs qu'il n'a pas indiquées au candidat parce qu'il ne souhaitait pas creuser ces points ou déstabiliser le candidat), néanmoins la phase de discussion l'aura convaincu que vous ne maitrisiez pas le contenu de votre plan, ou que vous n'aviez pas *vraiment* compris votre développement. Le fait qu'un plan soit bon ne représente rien, puisque l'étudiant peut l'avoir appris par coeur, ou l'avoir recopié d'un livre de plan, puisqu'il en existe. Un développement peut également être appris par coeur, même si c'est plus délicat. Si le jury ne s'en rend pas compte pendant votre défense ou pendant votre développement (ce qui est quand même peu probable), il le constatera rapidement pendant la phase de discussion, et il réévaluera largement à la baisse son impression de votre performance.

1.3 Conseils de préparation

En préambule, **mon premier conseil** est que vous devez savoir vous auto-évaluer, mesurer vos connaissances et vos éventuelles lacunes avec lucidité et honnêteté. Et ce, pendant les épreuves et aussi pendant votre préparation, tout au long de l'année !

Par exemple, il me semble contre-productif d'essayer d'apprendre un développement qui ne correspond pas à votre niveau (en tout cas au moment où vous le travaillez ; il est tout à fait possible qu'après quelques mois et de meilleures connaissances vous puissiez finir par travailler avec efficacité un énoncé qui vous passait au-dessus de la tête au début de l'année). Il vaut mieux consolider d'abord vos bases (quitte à travailler des développements plus modestes). De même, commencez au plus vite à travailler les leçons qui vous semblent les plus accessibles. Vous verrez finalement qu'il y a pas mal de choses sur lesquelles vous n'étiez pas parfaitement au point.

Pendant votre préparation, essayez d'être le plus humble possible, sans tomber dans une modestie excessive. Il faut que vous ayez conscience des chapitres que vous devez encore travailler, et de ceux que vous maîtrisez vraiment, afin que vous puissiez vous reposer dessus pour aborder des notions plus difficiles. Ça n'a pas de sens d'étudier les espaces métriques (L2/L3) si vous n'êtes pas à l'aise sur l'analyse réelle (L1 ; c'est dans une certaine mesure un cas particulier, mais beaucoup plus facile à comprendre). Ça n'a pas de sens d'apprendre la démonstration du théorème de Weierstrass (sur la densité des polynômes dans les fonctions continues) si vous n'avez pas une certaine intuition de la continuité uniforme (ce qui implique que vous puissiez facilement citer des exemples et des contre-exemples)...

Je conclus par un dernier conseil avant de rentrer dans les détails : même si vous serez tentés de beaucoup travailler la forme des épreuves (en particulier pour les épreuves orales), la priorité doit rester, presque toujours [1], de travailler le fond : **il faut "faire des maths" avant tout**, car même si la forme (plan-développements) de l'épreuve orale sera un moyen pour le jury d'engager le dialogue avec vous, c'est bien votre capacité à faire des mathématiques qui sera évaluée en priorité.

Je peux d'ailleurs citer des exemples inverses, à savoir des étudiants particulièrement brillants mais très peu préparés aux épreuves. Même si ces étudiants n'ont certainement pas obtenu le classement *optimal* au vu de leur potentiel, il ont eu des notes correctes voire bonnes, car le jury a pu vérifier qu'ils connaissaient bien les notions mises-en-jeu. Ces étudiants ont généralement de bonnes notes à l'écrit ; quant à l'oral où la forme de l'épreuve requiert certainement un plus gros effort, même s'il est bien possible que leur plan ou leur développement contenaient de gros défauts du fait du manque de préparation en amont, le jury a eu tout le temps de la discussion pour constater que le candidat connaissait les notions au programme et savait s'en servir.

Au final, vous devez progresser en mathématiques durant cette année de préparation ! Et bien sûr constater et évaluer cette progression pour orienter votre travail tout au long de l'année.

1.3.1 Les écrits

Sur la forme, pas de secret : faites des écrits blancs, et faites les le plus sérieusement possible, en vous imaginant en situation : il est bien difficile de se concentrer pendant 6h d'affilée, et cela s'entraine.

Sur le fond, naturellement l'entrainement le plus efficace consiste à travailler les preuves classiques du cours (à la fois les résultats assez simples, mais aussi occasionnellement des preuves plus longues ou techniques : les épreuves écrites sont souvent un recueil de démonstrations plus ou moins sophistiquées où on vous guide pas-à-pas vers plusieurs résultats intéressants et qui dépassent largement le programme, tout en étant construits avec les briques du programme), et bien sûr des exercices : sur ce point, continuez votre auto-évaluation, et ne vous jetez pas tête baissée à faire des exercices sans savoir pourquoi vous les faites :

— Distinguez les exercices élémentaires que vous faites pour mettre en pratique une notion nouvelle ou que vous révisez, des exercices plus astucieux ou élaborés qui demandent de prendre un peu de recul par rapport à une ou des notions.

— Quand vous réussissez un exercice, demandez-vous si vous l'auriez réussi avec moins de questions intermédiaires (un exercice d'agreg peut tout-à-fait être un exercice de niveau L1 auquel on a retiré les questions intermédiaires qui guident l'étudiant), demandez-vous à quelle problématique il répond et si cette problématique pourrait trouver sa place dans une leçon d'oral...

— Comprenez pourquoi vous n'avez pas réussi à faire un exercice (est-ce parce que vous ne connaissiez pas le théorème qu'il fallait appliquer, ou parce que vous n'y avez pas pensé ? est-ce parce qu'il y avait une idée profonde que vous n'avez pas trouvée ? etc)

— Enfin, avez-vous simplement griffoné les idées de l'exercice, ou l'avez-vous rédiger proprement ? Il n'est bien sûr pas interdit de ne pas être constamment en mode "copie de concours", mais il faut

1. Pour le "presque" : on peut comprendre, à quelques jours des oraux, que vous passiez en phase "catastrophe", et que vous accélériez un peu votre apprentissage au détriment d'un peu de qualité, pour les sujets que vous n'avez pas eu le temps de travailler pendant l'année. Il est en effet naturel de pallier aux tirages catastrophes qui combinent deux leçons sur lesquelles vous êtes secs. Mais dans ce cas, ne vous faites pas d'illusion, et assumer d'être en situation catastrophe, et visez des plans et des développements qui visent des notes de 5/20, note qui n'est certes pas brillante, mais qui ne vous élimine pas complètement de l'admission. Une telle note devrait effectivement correspondre à des exigences limitées qui peuvent correspondre à ce que vous pouvez ingurgiter dans la précipitation. Mais à nouveau, si vous mettez la barre trop haut (rappelons qu'on parle de situation catastrophe, idéalement la majorité des couplages ne rentrent pas dans cette catégorie), vous risquez une note très basse.

vous assurer que vous connaissez les principes de rédaction, et que vous les maitrisez au point de transformer rapidement votre gribouillage en une démonstration propre et recevable par un correcteur.

Au final, privilégiez la qualité à la quantité. Notons que la liste précédente s'applique également aux écrits blancs que vous ferez : ne retenez pas que la note, essayez de revenir sur les erreurs que vous avez faites, et comprenez pourquoi vous les avez faites (et quelles révisions cela suggère).

1.3.2 Les plans

Comme cela a été relevé dans les extraits précédents, il est impensable [2] d'espérer apprendre par cœur 36 plans de leçons (en fait il suffit d'en apprendre... 35) et autant (ou le double) de développements. De toute façon, il existe maintenant des livres avec des plans et des développements tout-faits.

À mon avis, il peut être très utile pendant l'année de s'entrainer à faire quelques plans en environ 2 heures [3], mais **il n'est pas nécessaire de rédiger 36 plans. Je dirais même que je ne le conseille pas**, car fatalement vous allez en partie recopier les plans des livres, du net, ou de vos camarades, et avoir tendance à inclure des énoncés trop difficiles ; et surtout, **vous allez y consacrer un temps considérable, temps que vous ne passerez pas à faire des maths de façon plus productive**. **Je conseille plutôt**, pour chaque leçon, **d'établir une idée précise des énoncés indispensables** que vous n'aurez pas le droit d'oublier si vous choisissez cette leçon le jour J.

Vous exposerez les uns les autres pendant l'année un exemple de plan pour chacune des 36 leçons, et ce sera l'occasion pour vous de réfléchir aux énoncés indispensables d'une leçon et dont vous n'êtes peut-être pas encore familiers, ou auxquels vous n'auriez peut-être pas pensé. Ce sera aussi l'occasion de voir des interactions entre les différents chapitres du programme, de développer des exemples riches qui pourront intervenir dans plusieurs leçons. À vous de déceler le plus rapidement possible des lacunes dans certaines parties du programme, et de les combler.

Attention, les plans qui seront présentés (ou que vous trouverez sur le net) seront dans la plupart des cas très fournis et très ambitieux, notamment car ils n'auront pas été préparés en 2-3h. La réalité le jour de l'oral sera très différente, et c'est tout à fait normal. Ne vous découragez pas en tout cas si vous réalisez que vous ne connaissez pas tout ce qui se trouve dans un plan présenté par un camarade ; vous avez le temps de beaucoup progresser, et de toute façon ces plans ne sont pas le standard du concours.

Au fond, **la préparation de l'oral est très complémentaire à l'apprentissage de votre cours d'analyse (et d'algèbre !) et des révisions que vous ferez tout au long de l'année**. Dans un cas vous verrez les notions de façon plus linéaire, dans l'autre, de façon plus transverse. Ces deux façons de travailler doivent être parallèles et complémentaires, et se faire tout au long de l'année.

1.3.3 Les développements

C'est certainement la partie de l'épreuve qui demandera la **préparation la plus chronophage**. Effectivement, il est quasiment impensable d'espérer "improviser" deux développements le jour de l'épreuve. Il est très préférable, pour presque toutes les leçons, d'avoir à l'avance quelques idées (deux peuvent suffire, si elles sont vraiment bonnes) des développements que vous pourriez proposer. Cela peut d'ailleurs aider à guider un peu la construction de votre plan, même si cela peut s'avérer dangereux également ; vous ne devez pas faire de disgression trop notable afin de "faire rentrer vos développements" à tout prix. Parfois, on voit aussi un dernier paragraphe "Applications" qui contient justement les deux développements proposés et qui n'est pas vraiment lié au reste de la leçon. À défaut de pouvoir mieux faire c'est évidemment une possibilité, mais elle est souvent peu idéale.

Trouver un bon développement, même pendant l'année et sans le stress de l'épreuve, **n'est pas du tout une chose facile**. Il faut trouver un énoncé qui soit non trivial, mais pas trop ambitieux non plus. Vous devez pouvoir le comprendre au point de prendre du recul et d'avoir de l'intuition sur la logique de la démonstration. Vous devez être capable de la rédiger au tableau en 15 min sans se précipiter, et **sans l'apprendre par-cœur**. En ce sens, je conseille souvent d'être capable de le

2. Il serait plus juste de dire "chronophage et contre-productif"

3. C'est le temps que vous pouvez espérer consacrer à votre plan ; n'oublions pas qu'il faut choisir et réapprendre vos développements, ce qui peut rapidement grignoter vos 3 heures de préparation

faire en 12-13 min quand vous êtes au calme. Dans ce cas, vous pouvez éventuellement prévoir un supplément bonus que vous ne sortirez que s'il vous reste du temps à l'issue de votre démonstration. Comprendre un développement, ça signifie aussi savoir trouver des applications pertinentes, et/ou des contre-exemples lorsque l'une ou l'autre des hypothèses est enlevée. Essayez d'anticiper les premières questions du jury, qui seront souvent de cette nature.

Un point central également, est de savoir où se situe votre développement dans la littérature. Il faut trouver un livre où la preuve sera suffisamment détaillée pour que vous puissiez vous la remémorer, et avec un style de rédaction qui vous correspond. Ce n'est pas toujours facile. Il faudra de toute façon vous approprier la preuve et l'écrire à votre façon : je vous conseille de **faire un cahier de développements**, où vous mettrez au propre, une fois que vous la maitriserez parfaitement, votre version de la démonstration ; vous pourrez également y indiquer les exemples/contre-exemples qui se rapportent à l'énoncé du développement, ou les possibles généralisations. **Commencez ce cahier dès le début de l'année !** Idéalement il contiendra environ une trentaine de développements d'analyse et autant en algèbre à la fin de votre préparation ; **il est préférable que vous ayez une poignée de développements (disons 5 en analyse, 5 en algèbre, à adapter à votre situation) dont vous êtes satisfaits pour fin décembre.**

Si vraiment vous devez improviser un résultat le jour de l'épreuve, je vous conseille de proposer la démonstration d'un énoncé central du plan, mais de difficulté modeste. Éventuellement vous pourrez l'accompagner d'un exemple pertinent si la démonstration est trop courte. C'est une "forme" de développement souvent négligée mais qui peut s'avérer une très bonne idée, notamment parce que vous êtes sûrs de ne pas être hors sujet et qu'il sera bien référencé.

1.3.4 Le tirage et les impasses

Le jour J, vous tirerez deux leçons. Evidemment, la meilleure stratégie consiste à savoir "sortir le bon tirage", c'est-à-dire la leçon de votre choix. En ce sens, si vous arrivez à maîtriser cette méthode, on peut assurément dire que vous seriez capable de passer votre oral d'agreg d'ici quelques semaines et de vous en tirer brillamment.

Si comme nous autres humains vous ne maîtrisez pas (encore) cette technique, **vous allez devoir beaucoup travailler**. Une façon de voir votre travail de l'année me semble la suivante : dans l'état actuel de vos connaissances, vous êtes probablement en mesure de vous en tirer honnorablement sur 2 ou 3 leçons du programme, qui touchent aux notions qui vous sont le plus familières et que vous maîtrisez le mieux. Quant aux autres leçons, vous ne voyez pas trop quoi y mettre pour le moment. Voyez votre travail comme une façon d'inverser la tendance, afin d'arriver en juin et d'être capable de faire correctement un maximum de leçons. Evidemment si vous tombez sur vos leçons favorites, vous pourrez encore plus briller, mais je pense que le meilleur objectif consiste à arriver à l'épreuve sans "avoir peur" de votre tirage. Car idéalement **il n'y a pas de bon ou de mauvais tirage**, il y a juste une énorme quantité de travail afin de pouvoir traiter un maximum de tirages possibles.

Dans ce sens, chaque impasse sur une partie du programme est une prise de risque. Attention, je ne dis pas ici qu'il faut tout savoir de tous les chapitres, qu'il faut savoir tout démontrer. Il faut voir les connaissances comme une pyramide, dont vous devrez connaître parfaitement les premiers étages (et ce sur chacun des chapitres du programme). Si vous apprenez des énoncés élaborés (plus haut dans la pyramide) mais que certaines bases ne sont pas solides, il est à attendre que votre pyramide s'écroule. Voyez votre préparation comme la construction de cette pyramide ; plus celle-ci arrive haut, plus vous pourrez prétendre à un bon résultat. Mais prenez garde aux fondations de votre construction !

À titre d'exemple, les fonctions holomorphes sont souvent un chapitre qui est négligé pas les candidats. À mon avis, connaître un minimum sur les fonctions holomorphes (la définition, la formule de Cauchy (pas sa démonstration) et ses conséquences, et la manipulation des séries entières) ne requiert que quelques heures de travail, et est beaucoup plus accessible que certaines démonstrations difficiles que certains étudiants connaissent par cœur. Et ce sera un premier étage qui peut suffire aux possibles questions de l'écrit [4], et vous permettre au moins d'aborder plus sereinement certaines leçons.

4. Par exemple dans le sujet 2018, la 6ème question (numéro II.3) est de niveau "premier étage" sur les fonctions holomorphes (il faut montrer le caractère holomorphe d'une fonction définie par une intégrale à paramètre) ; la 20ème question du sujet (numéro III.2.b) est une question "niveau deux" où il faut appliquer le théorème de Cauchy sur un contour bien choisi. Il serait en particulier vraiment dommage de ne pas pouvoir traiter la question 6, très classique et

En dernier exemple, évoquons les probabilités ; c'était une impasse classique il y a une quinzaine d'année, car le formalisme nécessitait un effort auxquels les élèves ne souhaitaient pas se confronter. C'est désormais une impasse qui sera de moins en moins tolérée par le jury et qui est devenue beaucoup plus risquée. Il y a maintenant 4 leçons de probabilités, ce qui en fait l'un des chapitres les plus représentés de la liste. Certaines peuvent être traitées à un niveau élémentaire, d'autres recquièrent un peu plus de maturité ; mais dans tous les cas, et même si la construction formelle repose sur la théorie de la mesure et de l'intégration de Lebesgue, il n'est pas nécessaire (et le plus souvent il ne vous est pas demandé) d'en connaître les moindres détails. Comme pour les fonctions holomorphes, vous devez vous fixer comme premier objectif de connaître un strict minimum sur la théorie de Lebesgue (définition d'une fonction intégrable, théorèmes de convergence monotone et de convergence dominée). Quant au formalisme probabiliste, vous pouvez commencer par le cas des mesures discrètes qui peut se traiter sans faire référence à la théorie de Lebesgue (la leçon **264** peut d'ailleurs se faire dans ce cadre), et quand vous aurez les connaissances de base sur “Lebesgue”, vous pourrez utiliser ce formalisme dans un cadre plus général.

1.3.5 Les camarades

Si vous avez la chance de pouvoir préparer l'agrégation en étant entouré de gens qui ont le même objectif que vous, profitez-en ! N'hésitez pas à travailler ensemble, à confronter vos idées. Présentez vos développements à vos camarades pour voir s'ils sont convaincus par votre rédaction, discutez ensemble de la meilleure façon de traiter le développement, de l'erreur ou de l'imprécision qui se trouve dans tel livre mais pas dans tel autre, etc. Vous pouvez représenter les uns pour les autres une source très utile d'informations et d'entre-aide, n'hésitez pas à en profiter. Et au-delà d'être plus efficace, ce sera certainement aussi plus agréable.

1.3.6 Les sources

Dans votre travail, vous devrez prendre du recul par rapport à vos cours manuscrits ou polycopiés, et travailler avec de nombreux livres. Vous devez ainsi vous construire votre “bibliothèque personnelle”, afin qu'en fin d'année vous puissiez savoir quel livre présente tel chapitre (ou tel développement ou tel exemple) de la façon qui vous convient le plus. Par exemple, avec une leçon comme **228**, l'utilisation d'un livre classique comme le “Gourdon d'analyse” [Gou08b] peut s'avérer en partie catastrophique, car le cadre développé par le livre n'est pas du tout celui attendu dans la leçon (la continuité est traitée dans des espaces métriques ; la partie dérivabilité est par contre plus adaptée). Les livres seront un support très utile pour la préparation et la rédaction de vos plans le jour de l'oral, mais il faudra être familier avec ces livres pour bien savoir les utiliser.

On trouve aujourd'hui de nombreuses aides sur internet, avec des listes de développements et des plans. Ces sources peuvent être une chance, représentent un témoignage souvent fort utile des efforts faits par vos prédécesseurs ; mais ça peut se révéler à double tranchant si vous n'utilisez pas ces sources à la lumière des conseils donnés précédemment. Par exemple, la plupart des développements que je peux trouver sur ces sources sont souvent très ambitieux, et inadaptés si vous faites une leçon de niveau modeste. Soyez donc vigilants et critiques envers les sources que vous utilisez (ça vaut également pour les livres).

Enfin, le rapport du jury que j'ai cité très partiellement ici, doit représenter votre première source d'information. On y trouve quelques lignes de commentaires pour chaque leçon ; lisez les et tirez-en parti à la lumière des conseils précédents, c'est-à-dire en ayant conscience de vos connaissances et du niveau auquel vous pouvez sereinement aborder une leçon.

2 Principes de rédaction

A venir ; pour le moment on conseille ce petit manuel de bonne rédaction rédigé par C. Bertault

qui ne demande pas beaucoup de compréhension de l'holomorphie.

Chapitre I

Analyse à une variable réelle

Comme expliqué brièvement en introduction, nous avons choisi de débuter ce cours par un chapitre d'analyse réelle plutôt que par un cours de topologie (métrique), même si bien sûr, $\mathbb{R}$ muni de sa valeur absolue constitue l'exemple le plus simple d'espace vectoriel normé (voir le paragraphe 2 du Chapitre III). La première raison de ce choix est qu'il me semble important d'avoir déjà une grande aisance dans ce cas particulier, et que celle-ci devrait permettre une compréhension simple et naturelle des cas plus généraux qui seront abordés dans les chapitres suivants. La seconde est que dans le cadre de l'agrégation, c'est souvent dans ce cadre (réel) que les choses devront être placées : les leçons **223-224-228**[1] se placent exclusivement dans ce cadre, et **226-229-230-236-241-265** se placent majoritairement dans ce cas (avec une attente du jury à aborder également le cas des suites complexes ou à valeurs dans $\mathbb{R}^n$; il n'empêche que le cas réel constituera une base prépondérante de ces leçons).

Ce chapitre représente environ le contenu des 3 premiers semestres de cours d'analyse au niveau L1-L2 ; nous ne précisons pas de référence globale, n'importe quel livre de ce niveau fera l'affaire. On abordera, notamment en fin de chapitre, des propositions de sujets que l'étudiant agrégatif peut travailler à bon escient (et qui sont rarement abordés en licence, ou alors seulement survolés), et à partir desquels il pourra mettre à l'épreuve ses connaissances et sa maîtrise du chapitre, et aussi se construire un début de culture pour l'épreuve orale, des idées pour certaines des leçons citées précédemment, et des développements. Nous préciserons plus de références pour ces sujets, même si l'auteur précise encore une fois que l'étudiant peut et doit trouver sa référence favorite.

Certains sujets seront laissés de côté par manque de temps, à savoir principalement les méthodes de calcul sur les limites (croissances comparées...), les développements limités, et les primitives. Ces sujets ne sont évidemment pas à négliger, notamment pour les épreuves écrites. Mais ils ne représentent pas de difficulté théorique particulière, et je pense qu'il vaut mieux les apprendre en les manipulant concrètement sur des exercices. Le lecteur potentiellement fébril en calcul ou qui veut vérifier sa dextérité aura donc tout intérêt à ressortir ses planches d'exercices de niveau L1 sur ces sujets.

1 Le corps des nombres réels

Dans ce paragraphe assez informel, on utilise des notions introduites plus loin dans le chapitre, comme la notion de suite convergente. L'objectif est de donner quelques éléments au lecteur pour pouvoir se lancer dans la suite. Le plus important étant de connaître les deux propriétés fondamentales de $\mathbb{R}$, à savoir que c'est un ensemble complet, et qui satisfait la propriété de la borne supérieure.

1.1 $\mathbb{R}$, un ensemble intuitif et ses propriétés

Je ne sais pas si le lecteur s'en est rendu compte, mais il est probable qu'il n'a jamais vu une construction de l'ensemble des réels, ensemble qu'il manipule pourtant depuis le collège. Ce n'est effectivement plus dans l'air du temps, au niveau Licence (ou Master d'ailleurs), et même au niveau de l'agrégation, de s'"encombrer" de telles considérations.

1. Voir la liste page 17.

À ma connaissance [2], la construction du corps des réels était au programme de l'agrégation il y a quelques décennies, et faisait même l'objet d'une leçon. Les temps ont changé, et cette leçon n'a pas seulement disparu, elle n'est plus du tout dans l'esprit du concours. Dans ce paragraphe, on rappelle donc une vision intuitive des réels (on ira assez vite, on convient que chaque étudiant utilise cela depuis bien des années sans trop de soucis), et surtout ses propriétés fondamentales, qu'il faut bien sûr connaître. Au paragraphe 1.3 suivant, on donnera au lecteur curieux (et ambitieux) matière à creuser de possibles constructions, dont l'une au moins n'est pas dénuée d'intérêt dans le cadre de l'agrégation.

On admet donc qu'il existe un ensemble $\mathbb{R}$ contenant l'ensemble des rationnels $\mathbb{Q}$, et qui satisfait :

i) $\mathbb{R}$ est un corps [3],

ii) $\mathbb{R}$ est totalement ordonné ($\leq$ est une relation d'ordre telle que pour tout $(x, y) \in \mathbb{R}, x \leq y$ ou $y \leq x$),

iii) $\mathbb{R}$ est archimédien, c'est-à-dire que pour tout $(x, y) \in \mathbb{R}$ tels que $0 < x$ et $0 < y$, $\exists n \in \mathbb{N}$ tel que $y \leq nx$,

iv) $\mathbb{R}$ est **complet** : toute suite $(u_n)_{n\in\mathbb{N}}$ de Cauchy, c'est-à-dire telle que

$$\forall \varepsilon > 0,\ \exists n_0 \in \mathbb{N},\ \forall p \geq n_0, \forall q \geq n_0,\ \ |u_q - u_p| \leq \varepsilon,$$

est convergente dans $\mathbb{R}$ (voir Définition 2.1).

Les propriétés de $\mathbb{R}$ sont nombreuses, et la liste ne s'arrête pas là, mais il n'est pas anodin qu'on l'arrête ici. On peut en effet montrer qu'il y a une forme d'unicité de l'ensemble qui satisfait ces 4 propriétés ; mais comme on le disait, nous ne rentrerons pas dans les détails. La 4ième propriété est particulièrement importante, car c'est celle qui distingue $\mathbb{R}$ de $\mathbb{Q}$, qui de son côté satisfait les 3 premières.

On peut également noter une 5ième propriété tout-à-fait fondamentale :

v) $\mathbb{R}$ possède la **propriété de la borne supérieure** : si $A \subset \mathbb{R}$ est non vide et majoré, c'est-à-dire [4] s'il existe M tel que $\forall x \in A, x \leq M$ (on dit que M est un majorant de A), alors l'ensemble des majorants de A possède un plus petit élément qu'on notera $\sup(A)$ [5]. Autrement dit, il existe $\sup(A)$ qui est majorant de A et tel que pour tout M majorant de A, on a $M \geq \sup(A)$.

On peut voir v) comme un théorème, ou comme un axiome issu de la construction de $\mathbb{R}$, tout dépend de comment on le construit. Plus clairement :

— il existe une construction de $\mathbb{R}$ qui amène à un ensemble qui satisfait i-ii-iii-iv) et à partir de là, on prouve v).

— il existe une construction de $\mathbb{R}$ qui amène à un ensemble qui satisfait i-ii-v), et à partir de là, on prouve iii) et iv).

On donne quelques éléments sur ces démonstrations et sur ces constructions au paragraphe 1.3. Il va sans dire que cela implique une "propriété de la borne inférieure", que le lecteur pourra écrire.

Remarque 1.1. Ainsi, si A est une partie non vide et majorée de $\mathbb{R}$, le réel $M = \sup(A)$ est caractérisé par le fait que

— M majore A, c'est-à-dire

$$\forall x \in A, \quad x \leq M.$$

— M est le plus petit majorant de A, c'est-à-dire

$$\forall M' \in \mathbb{R}, \Big[\ \forall x \in A, x \leq M'\ \Big] \ \ \Rightarrow\ \ M \leq M'$$

2. Je n'ai pas retrouvé l'information sur internet, je me souviens juste d'un collègue "un peu" plus âgé regrettant, comme c'est souvent le cas, le bon vieux temps. Si un lecteur peut me confirmer ou infirmer (et me préciser des dates), qu'il n'hésite pas à me contacter.

3. Rappelons que l'usage le plus répandu est qu'un corps est par défaut commutatif ; on dira éventuellement "corps non commutatif" ou "corps gauche" quand on veut indiquer qu'on retire cette hypothèse.

4. Le "c'est-à-dire" ne se rapporte qu'au mot "majoré", pas à l'hypothèse que A est non vide.

5. Le fait de pouvoir le noter suppose une unicité : on invite le lecteur à prouver cette unicité.

On utilise très fréquemment la seconde propriété sous la forme de sa contraposée, à savoir :

$$\forall M' \in \mathbb{R}, \quad M' < M \quad \Rightarrow \left[\exists x \in A, x > M' \right],$$

c'est-à-dire que tout réel strictement inférieur à M n'est pas un majorant de A. Dans les démonstrations, pensez bien au fait que la borne supérieure est caractérisée par deux propriétés.

Notez qu'on peut montrer à partir de ces "axiomes"[6].

Théorème 1.2. *$\mathbb{Q}$ est dense dans $\mathbb{R}$.*

Démonstration. Dans le cas des sous-ensembles de $\mathbb{R}$, on peut définir le caractère dense de la façon suivante : $A \subset \mathbb{R}$ est dit dense si pour tout $x < y$ deux réels, $]x, y[\cap A$ est non vide.[7]

Soit donc deux réels $x < y$. On distingue 3 cas : si $x < 0$ et $y > 0$, on peut voir que $0 \in]x, y[\cap \mathbb{Q}$. Supposons désormais $0 \leq x < y$. On cherche un entier n_0 suffisamment grand pour que les nombres de la forme $\left(\frac{k}{n}\right)_{k\in\mathbb{N}}$ passent par l'intervalle $]x, y[$ au moins une fois, ce qui sera possible si $\frac{1}{n}$ est strictement inférieur à la longueur de cet intervalle. On utilise donc le caractère archimédien de $\mathbb{R}$ aux nombres $(1, \frac{y-x}{2})$: il existe $n_0 \in \mathbb{N}$ tel que $1 \leq n_0.\frac{y-x}{2}$. En particulier $n_0 \neq 0$, d'où $\frac{1}{n_0} < y - x$. Pour trouver k, on pose

$$I := \left\{ k \in \mathbb{N}, \ \frac{k}{n_0} > x \right\}.$$

Cet ensemble est non vide, encore une fois d'après le caractère archimédien[8], et comme il est inclus dans $\mathbb{N}$, il contient un plus petit élément[9], qu'on note k_0. On a donc bien $\frac{k_0}{n_0} > x$. Pour voir que $\frac{k_0}{n_0} < y$, on utilise que $k_0 - 1$ n'est pas dans I (car k_0 était le plus petit) et donc

$$\frac{k_0 - 1}{n_0} \leq x, \ \text{ce qui donne } \frac{k_0}{n} \leq x + \frac{1}{n_0} < x + y - x = y,$$

et montre bien que $]x, y[\cap \mathbb{Q}$ est non vide.

Il reste à traiter le 3ième cas où $x < y \leq 0$. On applique le cas précédent au couple $0 \leq -y < -x$, et donc l'intervalle $]-y, -x[\cap \mathbb{Q}$ est non vide, et comme $\mathbb{Q}$ est stable par multiplication par (-1), on a le résultat[10]. $\square$

Que faut-il retenir de ce paragraphe ? D'une part, que l'on a le droit d'utiliser notre intuition, voir les réels comme des "nombres à virgules" ou encore comme des limites de rationnels, sans risquer d'être en dehors des clous du programme. D'autre part, que les propriétés iv) et v) (absolument fondamentales), peuvent chacunes être considérées tant comme des théorèmes que comme des axiomes issus des constructions de $\mathbb{R}$, ces constructions n'étant pas à connaître.

1.2 Préliminaires topologiques

Comme on l'a dit en introduction, on a choisi de faire figurer ce chapitre d'analyse réelle avant le chapitre de Topologie III. On utilisera quand même quelques éléments de vocabulaire usuel. On se donne $D \subset \mathbb{R}$:

— un élément x de $\mathbb{R}$ est dit adhérent à D s'il existe $(x_n)_{n\in\mathbb{N}}$ suite d'éléments de D qui converge vers x, et on notera $x \in \overline{D}$[11]. On appellera $\overline{D}$ l'adhérence de D. En considérant des suites constantes, on constate que $D \subset \overline{D}$. Enfin, on dira que D est fermé si $\overline{D} = D$.

6. En fait, on n'utilisera que le caractère archimédien.

7. Logiquement, vous devriez avoir en mémoire au moins une autre définition de la densité, par exemple dans le cas des espaces métriques, voir aussi le paragraphe 1.2 et l'exercice 1.5. Prenez le temps de comprendre pourquoi la définition utilisée ici avec les intervalles ne se généralise pas bien à d'autres types d'espaces.

8. On peut appliquer le caractère archimédien au couple $(x, \frac{1}{n_0+1})$ qui donne l'existence de k tel que $x \leq \frac{k}{n_0+1}$. En particulier $x < \frac{k}{n_0}$.

9. Qu'on pourrait calculer avec la fonction partie entière, voir les remarques de l'exercice 1.4.

10. Le lecteur peu convaincu par cette phrase peut écrire qu'il existe $r \in]-y, -x[\cap \mathbb{Q}$ et cherchera à constater que dans ce cas $(-r) \in]x, y[\cap \mathbb{Q}$.

11. Une définition équivalente : x est adhérent à D si et seulement si pour tout $\varepsilon > 0$, $]x - \varepsilon, x + \varepsilon[\cap D$ est non vide, voir le Chapitre III

— on dira qu'un point $x \in D$ est intérieur à D s'il existe $\alpha > 0$ tel que $]x-\alpha, x+\alpha[\subset D$. On dira que D est ouvert si tous ses points sont intérieurs à D.

— on appellera voisinage de x un ensemble qui contient $]x-\alpha, x+\alpha[$ pour un certain $\alpha > 0$. Par exemple, un ensemble est donc ouvert s'il est voisinage de tous ses points.

— on l'a déjà fait au paragraphe précédent, mais rappelons qu'on dira que D est dense dans $\mathbb{R}$ si pour tout $x < y$ deux réels, $]x, y[\cap D$ est non vide.

On utilisera également l'ensemble $\mathbb{R}\cup\{\pm\infty\}$ [12] [13]. Attention $\mathbb{R}\cup\{\pm\infty\}$ n'est plus un corps (on ne peut additionner $+\infty$ et $-\infty$, ni multiplier 0 et ∞ [14]). C'est par contre un espace qu'on peut classiquement munir d'une métrique, voir l'exemple 1.6 du Chapitre III. Il peut être très commode, notamment pour reformuler des propriétés et unifier des cas. Par exemple :

Propriété de la bornée supérieure reformulée : si A est un sous-ensemble non vide de $\mathbb{R}$, alors on peut poser [15] :

$$\sup(A) := \begin{cases} \text{le plus petit des majorants de } A \text{ si } A \text{ est majoré} \\ +\infty \text{ sinon} \end{cases}$$

Cette notation peut être très commode, mais attention, il n'empêche que dans les démonstrations, on est parfois amené à distinguer deux cas. On a vite fait d'écrire des bêtises en travaillant avec $+\infty$ [16].

On a déjà utilisé à plusieurs reprises la notion d'intervalle. Rappelons quand même que par définition, un intervalle est un ensemble *convexe* de $\mathbb{R}$: I est un intervalle si et seulement si

$$\forall (x,y) \in I,\ \forall z \in \mathbb{R}, \quad x < z < y \Longrightarrow z \in I,$$

et qu'on peut montrer (voir l'exercice 1.6) que tout intervalle se trouve dans la liste suivante :

$$[a,b],\ [a,b[,\ ]a,b],\ ]a,b[,\ [a,+\infty[,\ ]a,+\infty[,\ ]-\infty,b],\ ]-\infty,b[,\ \emptyset,\ \mathbb{R}, \tag{1.1}$$

où $a \leq b$ sont des réels.

On pourra ajouter

$$[a,+\infty],\ \ [-\infty,b],\ \ \mathbb{R}\cup\{\pm\infty\},$$

si on accepte de travailler avec $\mathbb{R}\cup\{\pm\infty\}$, mais ce ne sont pas des intervalles de $\mathbb{R}$.

On conclut par un résultat très classique sur les sous-groupes de $\mathbb{R}$:

Proposition 1.3. *Soit G un sous-groupe de $(\mathbb{R},+)$. Alors*

— *soit il existe $a \in \mathbb{R}_+$ tel que $G = a\mathbb{Z}$,*

— *soit G est dense dans $\mathbb{R}$.*

Démonstration. Si $G = \{0\}$, on prend $a = 0$. Sinon, G contient un élément non nul $x \in G \setminus \{0\}$, qu'on peut supposer strictement positif quitte à l'échanger avec $-x \in G \setminus \{0\}$. L'ensemble $G \cap \mathbb{R}_+^*$ est non vide (contient x) et minoré (par 0), il possède donc une borne inférieure $a \geq 0$.

---

12. Où $\{\pm\infty\}$ vaut pour $\{+\infty, -\infty\}$.

13. Une notation très répandue pour cet ensemble est $\overline{\mathbb{R}} := \mathbb{R}\cup\{-\infty,+\infty\}$. Pourquoi pas, à condition de comprendre qu'il s'agit d'une notation, et qu'elle est dangereuse : il ne s'agit pas de l'adhérence de $\mathbb{R}$! Effectivement, si on travaille dans l'espace métrique usuel $\mathbb{R}$, alors son adhérence (qu'on note également $\overline{\mathbb{R}}$) sera simplement $\mathbb{R}$, puisqu'en tant qu'espace ambiant, c'est un fermé.

14. Dans le chapitre IV, on établiera une convention pour cette multiplication, qui étendra à tout $\mathbb{R}\cup\{\pm\infty\}$ la multiplication usuelle, mais il ne faut pas perdre de l'esprit que cette extension de l'opération de multiplication ne sera plus continue.

15. On définit même parfois, par convention, $\sup(\emptyset) = -\infty$, même si je déconseille d'user trop fréquemment de ces généralisations par toujours intuitives.

16. À titre d'exemple, imaginons qu'on arrive à démontrer $2\sup(A) \leq \sup(A)$. On ne pourra déduire $\sup(A) \leq 0$ que si on sait que A est borné : en effet $2 * (+\infty) \leq +\infty$ est une inégalité tout à fait valable. De même, je vous laisse voir pourquoi $\sup(A) + 2 \leq \sup(A)$ n'amène pas nécessairement à une contradiction.

— Supposons $a > 0$. Supposons par l'absurde que $a \notin G$. Comme $a < 2a$, $2a$ n'est pas un minorant de $G \cap \mathbb{R}_+^*$, donc il existe $b \in G$ tel que $a \leq b < 2a$, mais comme $a \notin G$, on a en fait $a < b < 2a$. On répète l'argument entre a et b, ce qui donne l'existence de $c \in G$ tel que $a < c < b$. Mais alors

$$0 < b - c < 2a - a = a, \qquad \text{et par structure de groupe,} \qquad b - c \in G$$

ce qui contredit le fait que a minore $G \cap \mathbb{R}_+^*$ [17]. Ainsi $a \in G$, et par suite $a\mathbb{Z} \subset G$. Pour montrer l'inclusion réciproque, on se donne $g \in G$, et on pose $n = E(\frac{g}{a})$. Alors avec la définition de la partie entière (voir Exercice 1.4), on obtient $na \leq g < (n+1)a$, i.e. $0 \leq g - na < a$. Mais par structure de groupe, $g - na \in G$, donc par définition de a, $na - g = 0$, i.e. $g \in a\mathbb{Z}$, ce qui conclut ce cas.

— Supposons $a = 0$ [18] : montrons que G est dense dans $\mathbb{R}$. Soit $(x, y) \in \mathbb{R}^2$ avec $x < y$. Ainsi $y - x \in \mathbb{R}_+^*$ et $\inf(G \cap \mathbb{R}_+^*) = 0$, donc il existe $g \in G$ tel que $0 < g < y - x$. On pose $n = E\left(\frac{x}{g}\right) + 1$. Alors par définition de la partie entière, $(n-1)g \leq x < ng$, d'où par suite

$$x < ng = (n-1)g + g \leq x + g < y, \quad \text{d'où } ng \in]x, y[\cap G.$$

On a donc montré que G intersecte tout intervalle de $\mathbb{R}$, c'est-à-dire que G est dense dans $\mathbb{R}$.

□

1.3 **Vers les constructions de $\mathbb{R}$

Une idée naturelle (voir aussi [God98, page 51]), au vu de l'intuition qu'on s'en fait, serait de définir un réel comme un nombre à virgule (on dira plutôt "écrit sous forme décimale"), c'est-à-dire comme une suite d'entiers entre 0 et 9 (si on choisit de travailler en base 10, ce qui ne doit pas avoir d'importance a posteriori), et d'un endroit où placer la "virgule". On pourrait ainsi voir l'ensemble des rationnels comme l'ensemble des réels dont l'écriture décimale devient périodique à partir d'un certain rang. Si tant est qu'on pouvait formaliser une telle construction (on peut), cela poserait un premier problème (pas trop grave) de non unicité de l'écriture décimale pour certains nombres [19], mais surtout un très sérieux problème pour définir addition et multiplication de deux réels, qui ne se prêtent pas bien à cette écriture "infinie". Même si cela peut être rendu rigoureux (d'après Wikipedia), on présentera une autre construction plus classique. Mais avant cela, nous montrons l'équivalence des axiomes iii-iv) et v) du paragraphe précédent.

1.3.1 *Equivalences des constructions

On donne ici des idées de preuves de l'"équivalence" des constructions évoquées au paragraphe précédent (pour lire ce paragraphe, il vaut mieux être déjà familier avec les notions de base sur la complétude qui se trouve au paragraphe 1.4 du Chapitre III) :

***Éléments de preuve de iii) et iv) en supposant v) :** À partir de v), on montre le théorème de la limite monotone (voir théorème 2.8), qui affirme qu'une suite décroissante et minorée est convergente dans $\mathbb{R}$. On applique ce résultat à la suite $\left(\frac{1}{n}\right)_{n \in \mathbb{N}^*}$, qui a donc une limite [20]. Par suite,

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1} \xrightarrow[n \to +\infty]{} 0$$

et donc si $0 < x < y$, la définition de cette convergence avec $\varepsilon = \frac{x}{y}$ montre qu'il existe $n_0 \in \mathbb{N}$ tel que

$$\frac{1}{n_0(n_0+1)} \leq \frac{x}{y}, \quad \text{d'où } y \leq n_0(n_0+1)x,$$

17. On a utilisé un peu plus tôt que des nombres supérieurs strictement à a n'étaient pas minorant, et ici que a est un minorant de $G \cap \mathbb{R}_+^*$, on a donc bien utilisé les deux caractéristiques de la borne inférieure, voir la Remarque 1.1.

18. Ici, on va seulement utiliser une des caractéristiques de la borne inférieure, à savoir que si $M \in \mathbb{R}_+^*$, alors M n'est pas un minorant de G. La raison est que le fait que $a = 0$ soit un minorant de $\mathbb{R}_+^* \cap G$ est ici "trivial", et ne va donc pas nous aider.

19. Rappelons si besoin que $0,999999\ldots = 1$

20. Je ne suis pas en mesure de conclure que cette limite est zéro, il faut donc utiliser l'astuce qui suit.

et cela implique le caractère archimédien de $\mathbb{R}$.

Pour montrer la complétude, considérons la suite $(u_n)_{n\in\mathbb{N}}$ de Cauchy. Elle est alors bornée, et on peut en extraire une sous-suite monotone (voir l'exercice 2.26). Cette sous-suite sera donc monotone et bornée, donc convergente dans $\mathbb{R}$, encore d'après le théorème 2.8 qui découle de l'axiome v). On conclut avec le lemme classique qu'une suite de Cauchy ayant une sous-suite convergente, est elle-même convergente, voir le paragraphe 1.4 du Chapitre III. □

****Éléments de preuve de v) en supposant iii-iv) :** on va utiliser une méthode de dichotomie, très similaire à la preuve du théorème 2.13. Soit A un sous-ensemble non vide de $\mathbb{R}$, qui est majoré par M. On sait qu'il existe $x \in A$. Dans le cas où x est un majorant de A, alors x est une borne supérieure pour A. Supposons donc que x ne majore pas A. On construit maintenant deux suites $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ de manière récurrente : on pose $a_0 = x$ et $b_0 = M$, et pour tout $n \in \mathbb{N}$, si $(a_0, b_0), \dots, (a_n, b_n)$ sont construits, on pose

$$\begin{cases} a_{n+1} = a_n \text{ et } b_{n+1} = \frac{a_n+b_n}{2} & \text{si } \frac{a_n+b_n}{2} \text{ est un majorant de } A \\ a_{n+1} = \frac{a_n+b_n}{2} \text{ et } b_{n+1} = b_n & \text{sinon.} \end{cases}$$

On vérifie que cette construction récurrente fournit deux suites bien définie et telles que

— la suite $(a_n)_{n\in\mathbb{N}}$ est croissante et aucun de ses termes n'est un majorant de A.

— la suite $(b_n)_{n\in\mathbb{N}}$ est décroissante et ses termes sont tous des majorants de A.

— pour tout $n \in \mathbb{N}$, $\;b_n - a_n = \frac{M-x}{2^n}$, donc tend vers 0 (par le caractère Archimédien de $\mathbb{R}$).

Ces suites sont dites adjacentes, on vérifie qu'elles sont donc de Cauchy : par convergence de $(b_n - a_n)_{n\in\mathbb{N}}$, étant donné $\varepsilon > 0$, il existe n_0 tel que $|b_n - a_n| \leq \varepsilon$, et donc

$$\forall q \geq p \geq n_0, \quad 0 \leq a_q - a_p \leq b_q - a_p \leq b_p - a_p \leq \varepsilon$$

et de même pour $(b_n)_{n\in\mathbb{N}}$. Par l'axiome de complétude, on en déduit que $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ sont convergentes dans $\mathbb{R}$, et comme $(b_n - a_n)_{n\in\mathbb{N}}$ converge vers 0, elles ont la même limite qu'on note ℓ. On va conclure en prouve que ℓ est la borne supérieure de A :

— pour tout $y \in A$, on a $\forall n \in \mathbb{N}, y \leq b_n$ car b_n est un majorant de A. À la limite (voir l'exercice 2.22), on obtient $\forall y \in A, \;\; y \leq \ell$ donc ℓ est bien un majorant de A.

— Soit M' est un majorant de A. Pour $n \in \mathbb{N}$, a_n ne majore pas A, donc $a_n < M'$ (en effet, à n fixe, il existe $y \in A$ tel que $a_n < y$ et comme M' est majorant, $y \leq M'$). En passant à la limite, on obtient $\ell \leq M'$, ce qui conclut la preuve.

□

1.3.2 **Construction de $\mathbb{R}$ comme complété de $\mathbb{Q}$

On donne quelques éléments sur la construction classique de $\mathbb{R}$ qui amène à l'existence d'un ensemble qui satisfait i)-ii)-iii)-iv). On trouvera plus de détails sur ce Lien, dans [LFA93, page 5] [21], et dans [RDO91]. L'idée de base est de considéré l'ensemble $\mathbb{Q}$ construit (comme corps de fraction de $\mathbb{Z}$, voir un cours d'algèbre), et de remplir les trous en incluant les "limites" de suites de rationnels. Le hic dans cette expression, est que pour savoir qu'une suite est convergente par sa définition, on doit a priori connaître sa limite, et donc on tourne en rond. D'où l'idée naturelle de ne pas utiliser les suites convergentes, mais les suites de Cauchy (de rationnels).

En fait, cette construction est générale, et elle consiste à construire le "complété" d'un espace métrique (X, d), à savoir le "plus petit" ensemble complet qui le "contient". On renvoit au paragraphe 1.4 du Chapitre III pour plus d'éléments sur ce point.

En tout cas, il n'est pas rare de dire que l'on définit $\mathbb{R}$ comme le complété de $\mathbb{Q}$. Attention quand même, une subtilité non négligeable se cache ici : dans la définition d'une suite de Cauchy, on débute par "$\forall\varepsilon > 0$", ce qui veut a priori dire "$\forall\varepsilon \in \mathbb{R}_+^*$". Cela ne saurait convenir si on est en train de construire

21. Référence non vérifiée.

$\mathbb{R}$. C'est une difficulté mineure, car on sait qu'il suffit de remplacer cette expression par "$\forall \varepsilon \in \mathbb{Q}_+^*$", mais ça vaut le coup de s'en rendre compte.

On peut aller plus loin et montrer que tout réel possède un unique développement décimal propre, c'est-à-dire que x s'écrit

$$x = \sum_{k=0}^{+\infty} \frac{x_k}{10^k}, \quad \text{avec } x_0 \in \mathbb{Z}, \forall k \in \mathbb{N}^*, \ \ x_k \in [\![0,9]\!], \quad \text{et} \quad \Big[\forall N \in \mathbb{N}, \exists n \geq N, x_n \neq 9\Big].$$

La dernière propriété entre crochet signifie que le développement n'est pas stationnaire à 9, ce qui renvoie à l'ajectif "propre" employé précédemment, et permet de récupérer l'unicité.

Enfin, pour voir une forme d'unicité du corps satisfaisant i-ii-iii-iv), on renvoit au théorème 6. de ce lien.

1.3.3 **Construction de $\mathbb{R}$ par la méthode des coupures de Dedekind

Voir [God98, page 51], ainsi que l'exercice 6. page 41 de ce lien. Cette construction prouve l'existence d'un corps qui satisfait i-ii-iii).

1.4 Exercices

Exercice 1.4 (Sur la propriété d'Archimède)**.** L'axiome iii), appelé propriété d'Archimède, ne doit pas paraître mystérieux, ni être invoqué à tout va. On le précise car c'est un aspect de la construction, mais une fois cette construction achevée, on l'utilise le plus souvent sans y faire référence. On va en voir ici une caractérisation particulièrement intuitive, et voir qu'il permet de montrer l'existence de la fonction partie entière, et c'est souvent par ce biais qu'on l'utilise sans le dire :

1. Montrer que la propriété iii) est équivalente à
$$\forall \varepsilon > 0, \ \ \exists n_0 \in \mathbb{N}^*, \forall n \geq n_0, \ \ \frac{1}{n} \leq \varepsilon.$$
Cela signifie en fait que la suite $\left(\frac{1}{n}\right)_{n\in\mathbb{N}^*}$ tend vers 0 (dans le paragraphe 1.3.1 on verra un résultat un peu plus astucieux, à savoir que c'est équivalent au fait que la suite $\left(\frac{1}{n}\right)_{n\in\mathbb{N}^*}$ soit convergente dans $\mathbb{R}$, sans préciser la limite).
2. Soit $x \in \mathbb{R}$. Montrer que l'ensemble
$$\{n \in \mathbb{Z}, \ \ n \leq x\}$$
est non vide et majoré. Il a donc un plus grand élément [22], qu'on notera $E(x)$ ou $\lfloor x \rfloor$, et qu'on appellera partie entière de x. Montrer que
$$\forall n \in \mathbb{Z}, \quad n = E(x) \quad \Leftrightarrow \quad n \leq x < n+1.$$
Cela fournit une définition équivalente de la partie entière, à savoir l'unique entier $n \in \mathbb{Z}$ tel que $n \leq x < n+1$.

Exercice 1.5 (Sur les définitions topologiques dans le cas de $\mathbb{R}$)**.** Soit $D \subset \mathbb{R}$. Montrer que D est dense dans $\mathbb{R}$ si et seulement si pour tout $x \in \mathbb{R}$, il existe une suite $(a_n)_{n\in\mathbb{N}}$ d'éléments de D qui converge vers x.

Exercice 1.6 (Sur les intervalles)**.** Soit I un intervalle non vide de $\mathbb{R}$. On pose $b = \sup(I)$ et $a = \inf(I)$ dans $\mathbb{R} \cup \{\pm\infty\}$.

22. Attention, il ne faut pas confondre, ici on n'utilise pas la propriété de la borne supérieure, on fait référence à une propriété de $\mathbb{Z}$, non de $\mathbb{R}$.

— Montrer que $]a, b[\subset I \subset [a, b]$, et en déduire que I est dans la liste (1.1).
— Identifier parmi (1.1) les intervalles qui sont fermés, ouverts, ni ouverts ni fermés, à la fois ouverts et fermés.

Exercice 1.7 (On suit principalement [RDO91]). On pose

$$C(\mathbb{Q}) := \Big\{(u_n)_{n\in\mathbb{N}},\ \ \forall\varepsilon \in \mathbb{Q}_+^*, \exists n_0 \in \mathbb{N},\ \ \forall p \geq n_0, \forall q \geq n_0,\ \ |u_q - u_p| \leq \varepsilon\Big\}.$$

1. Montrer que $C(\mathbb{Q})$ peut être muni d'une structure d'anneau.
2. On pose
$$I = \Big\{(u_n) \in \mathbb{Q}^{\mathbb{N}},\ \ \forall\varepsilon \in \mathbb{Q}_+^*, \exists n_0 \in \mathbb{N},\ \ \forall n \geq n_0, |u_n| \leq \varepsilon\Big\}.$$
Montrer que $I \subset C(\mathbb{Q})$, et que I est un idéal de $C(\mathbb{Q})$.
3. Miontrer que $C(\mathbb{Q})/I$ est un corps. On le note $\mathbb{R}$.
4. Définir une injection θ qui permet de voir $\mathbb{Q}$ comme un sous-corps de $\mathbb{R}$.
5. On pose $\varphi : C(\mathbb{Q}) \to C(\mathbb{Q})/I$ la surjection canonique. On définit
$$C(\mathbb{Q})_+ := I \cup \Big\{(x_n)_{n\in\mathbb{N}} \in C(\mathbb{Q}), \exists n_0 \in \mathbb{N},\ \ \forall n \geq n_0, u_n > 0\Big\},\quad \mathbb{R}_+ := \varphi(C(\mathbb{Q})_+).$$
Enfin, pour $(X, Y) \in \mathbb{R}$, on définit la relation d'ordre $\leq$ par
$$Y \geq X \ \ \Leftrightarrow\ \ Y - X \in \mathbb{R}_+.$$
Montrer que $(\mathbb{R}, \leq)$ est un corps totalement ordonné.
6. Pour $x \in \mathbb{R}$, on peut définir $|X| = \max(X, -X)$, et ainsi définir une notion de convergence et de suite de Cauchy pour une suite de réels.
   (a) Soit $(x_n)_{n\in\mathbb{N}} \in C(\mathbb{Q})$, $X = \varphi((x_n)_{n\in\mathbb{N}})$. Montrer que la suite $(\theta(x_n))_{n\in\mathbb{N}}$ converge vers X dans $\mathbb{R}$.
   (b) Montrer que si $X, Y \in \mathbb{R}^2$ et $X < Y$, alors l'intervalle $]X, Y[$ contient au moins un rationnel, c'est-à-dire un élément de la forme $\theta(x)$, $x \in \mathbb{Q}$.
   (c) Montrer que $\mathbb{R}$ est archimédien.
   (d) Montrer que $\mathbb{R}$ est complet.

---

2 Suites et séries numériques

Une suite réelle peut-être vue comme un “vecteur infini de réels”, qu'on note généralement $(u_n)_{n\in\mathbb{N}}$ avec la propriété que pour tout $n \in \mathbb{N}$, $u_n \in \mathbb{R}$. On peut aussi la voir comme une application $u : \mathbb{N} \to \mathbb{R}$, mais pour un $n \in \mathbb{N}$ donné, la notation $u(n)$ en place de u_n est rarement utilisée. Néanmoins, il peut être commode d'écrire u à la place de $(u_n)_{n\in\mathbb{N}}$ (ici n est une lettre muette), nous considèrerons ces deux objets comme rigoureusement égaux. La notation $\mathbb{R}^{\mathbb{N}}$ désignera donc l'ensemble des suites réelles : elle convient bien à l'intuition d'un vecteur infini, et correspond également à la notation Y^X comme désignant l'ensemble des fonctions dont l'espace de départ est X et l'espace d'arrivée est Y.

Bien sûr, une suite peut aussi être indexée sur un ensemble du type $[\![n_0, +\infty[\![:= [n_0, +\infty[\cap \mathbb{N}$ où $n_0 \in \mathbb{N}$ [23] est le premier indice à partir duquel la suite est définie. On ne fera pas l'effort de prendre ce degré de généralité qui alourdirait les notations. Le lecteur prendra la peine d'utiliser les définitions et les propositions au cadre dans lequel il les emploie.

23. On peut même autoriser $n_0 \in \mathbb{Z}$.

On en profite dès le début de ce cours pour conjurer les étudiants à faire la différence (intellectuellement et dans leurs écrits) entre u_n (alors que n a été préalablement fixé) et $(u_n)_{n\in\mathbb{N}}$; la première notation désigne le n-ième élement de la suite (ce qui nécessite qu'on ait introduit préalablement la variable n), la seconde désigne la suite dans son intégralité. L'auteur de ces notes s'efforcera de faire cet effort tout au long du document. [24] La notation (u_n) pour $(u_n)_{n\in\mathbb{N}}$ est également très répandue, et probablement acceptable, mais nous l'éviterons ici : sans pour autant la déconseiller, nous suggérons au lecteur d'éviter la paresse au maximum dans son usage des notations.

Concluons ce paragraphe introductif par les faits clairs suivants : on peut additionner, multiplier par un scalaire, et multiplier entre elles (termes à termes), des suites réelles (ces propriétés sur l'espace d'arrivée $\mathbb{R}$ se transposent facilement aux fonctions à valeurs dans $\mathbb{R}$) [25]. Aussi, on dira qu'une suite $(u_n)_{n\in\mathbb{N}}$ est majorée, minorée, bornée, quand son image $u(\mathbb{N}) = \{u_n, n \in \mathbb{N}\}$ est majorée, minorée, bornée, respectivement, voir aussi l'exercice 2.15. Enfin, $\mathbb{N}$ et $\mathbb{R}$ ayant un ordre, on peut parler de monotonie, croissance et décroissance d'une suite $(u_n)_{n\in\mathbb{N}}$, voir l'exercice 2.16.

2.1 Suites ayant une limite, suites convergentes

Nous reviendrons sur ce titre un peu surprenant dans la Remarque 2.6.

2.1.1 Définitions et premières propriétés

Commençons par donner la classique définition d'une suite convergente :

Définition 2.1. [Convergence dans $\mathbb{R}$]

— Soit $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, et $\ell \in \mathbb{R}$. On dit que la suite $(u_n)_{n\in\mathbb{N}}$ converge vers ℓ (ou "tend vers ℓ" ou "a pour limite ℓ") si

$$\forall \varepsilon > 0, \quad \exists n_0 \in \mathbb{N}, \quad \forall n \geq n_0, \quad |u_n - \ell| \leq \varepsilon.$$

On notera alors

$$\lim_{n\to+\infty} u_n = \ell, \qquad \text{ou encore} \qquad u_n \xrightarrow[n\to+\infty]{} \ell.$$

— On dira d'une suite réelle $(u_n)_{n\in\mathbb{N}}$ qu'elle est convergente (ou "convergente dans $\mathbb{R}$" pour plus de précision, voir aussi la Remarque 2.6) s'il existe $\ell \in \mathbb{R}$ tel que $(u_n)_{n\in\mathbb{N}}$ converge vers ℓ.

Remarque 2.2. Le fait de pouvoir noter $\lim\limits_{n\to+\infty} u_n$ pour la limite d'une suite convergente est dû au fait qu'une limite de suite est unique : on renvoit à l'exercice 2.17.

Le résultat suivant, très classique, donne une condition nécessaire de convergence :

Proposition 2.3. *Si $(u_n)_{n\in\mathbb{N}}$ est une suite réelle qui converge dans $\mathbb{R}$, alors $(u_n)_{n\in\mathbb{N}}$ est bornée.*

Démonstration. Soit $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ une suite convergente. Il existe $\ell \in \mathbb{R}$ tel que $u_n \xrightarrow[n\to+\infty]{} \ell$. Appliquons la définition de la convergence avec $\varepsilon = 42$ [26]. Il existe $n_0 \in \mathbb{N}$ tel que

$$\forall n \geq n_0, |u_n - \ell| \leq 42. \tag{2.1}$$

Posons alors $M = \max\{|u_0|, |u_1|, \ldots, |u_{n_0-1}|, 42 + |\ell|\}$. Soit $n \in \mathbb{N}$.

24. On réitèrera cet effort et cette requête pour la distinction entre f et $f(x)$ dans la section 3.

25. *On dit que l'espace $(\mathbb{R}^{\mathbb{N}}, +, \cdot, \times)$ est une algèbre (où on note $\cdot$ pour la multiplication entre un réel et une suite, et $\times$ pour la multiplication termes à termes entre deux suites, au sens où $(\mathbb{R}^{\mathbb{N}}, +, \cdot)$ est un espace vectoriel, et que $\times$ est une opération interne et bilinéaire.

26. Je m'efforcerai de choquer le lecteur avec ces choix farfelus (mais dont le choix n'a rien d'anodin pour un lecteur un peu connaisseur), afin qu'il comprenne qu'il **doit** faire un choix, et non dire "on prend $\varepsilon > 0$ quelconque". N'oublions pas qu'une propriété montrée pour tout $\varepsilon \in A$ n'implique pas a priori que la propriété est vraie pour au moins un $\varepsilon \in A$: il faut en effet préciser que A est non vide pour que cela devienne vrai. Ici, on préfère choisir un élément bien explicite de $A := \mathbb{R}_+^*$ (à savoir 42) que de le faire pour tout $\varepsilon \in A$ et de conclure le raisonnement en disant que A est non vide (ce qu'aucun étudiant ne pense à faire, et qui de toute façon passerait par le choix d'un exemple tel que... $42 \in A$!). Pour le lecteur choqué de cette remarque et qui rétorque qu'il est évident que $\mathbb{R}_+^*$ est non vide, je réponds : le problème n'est pas là, à l'écrit comme à l'oral, si on utilise une propriété, quand bien même évidente, il faut l'écrire/le dire. Parce que c'est ça qu'on appelle la rigueur, et que le jour où il ne sera pas évident (ou faux !) que l'ensemble A est non vide, il faudra bien se rendre compte qu'il y a quelque chose de subtil à montrer ; prenons donc les bons réflexes de rédaction tant que les raisonnements sont simples, c'est ce qui nous permettra d'aborder avec les bons automatismes les raisonnements les plus délicats.

— Premier cas : si $n < n_0$, alors par définition $|u_n| \leq M$.
— Deuxième cas : si $n \geq n_0$, on peut utiliser (2.1), et en déduire :

$$|u_n| \leq |u_n - \ell| + |\ell| \leq 42 + |\ell| \leq M.$$

Dans tous les cas, on a montré

$$\forall n \in \mathbb{N}, \quad |u_n| \leq M$$

et donc la suite $(u_n)_{n\in\mathbb{N}}$ est bien majorée. $\square$

Dans le cas de l'espace d'arrivée $\mathbb{R}$ (ça ne sera pas le cas dans $\mathbb{C}$ ou dans un espace métrique quelconque), on peut aussi avoir des suites "non convergentes (dans $\mathbb{R}$)", mais qui ont des limites :

Définition 2.4. [Limites infinies] Soit $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$. On dit que la suite $(u_n)_{n\in\mathbb{N}}$ tend vers $+\infty$ (ou "diverge vers $+\infty$" ou "a pour limite $+\infty$") si

$$\forall A \in \mathbb{R}, \quad \exists n_0 \in \mathbb{N}, \ \forall n \geq n_0, \ u_n \geq A. \tag{2.2}$$

On notera alors

$$\lim_{n\to+\infty} u_n = +\infty, \qquad \text{ou encore} \qquad u_n \xrightarrow[n\to+\infty]{} +\infty.$$

On laisse le lecteur deviner la définition de la divergence vers $-\infty$.

Remarque 2.5. Il ne faut pas oublier que l'objet $\lim\limits_{n\to+\infty} u_n$ n'est pas défini tant qu'on n'a pas montré qu'une telle limite existe. Or il est de très mauvais ton, à l'écrit comme à l'oral, de manipuler un objet dont on n'a pas justifié qu'il existe (d'autant plus quand cela fait partie de la question). Si on vous demande de montrer l'existence de la limite d'une suite $(u_n)_{n\in\mathbb{N}}$ et de la calculer, il sera très mal perçu que votre réponse commence par contenir l'expression $\lim\limits_{n\to+\infty} u_n$. Cette remarque vaut pour tout objet mathématique dont l'existence est sujette à condition (c'est le cas pour une somme infinie, une division, une intégrale, une dérivée, etc).

Remarque 2.6 (Sur l'emploi des mots convergence/divergence)**.** Stricto sensu, une suite réelle est dite convergente si elle a pour limite un réel. De même dans un espace métrique, on parlera de suite convergente s'il y a un objet *de l'espace métrique considéré* vers lequel la suite se rapproche, voir le paragraphe 1 du Chapitre III. Et donc on est invité à dire d'une suite réelle qu'elle est divergente quand elle n'est pas convergente. Néanmoins, dans le cas des réels, il n'est pas très commode de parler de suites divergentes, au sens où celles-ci se divisent en deux catégories, celles qui ont une limite non réelle (à savoir celles qui tendent vers $+\infty$ ou $-\infty$), et celles qui n'ont vraiment pas de limite (ni réelle, ni infinie).

Une élément qui accentuera ces ambiguités, est l'utilisation de l'objet $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ [27]. C'est en effet un espace qu'on peut classiquement munir d'une métrique, voir l'exemple 1.6 du Chapitre III ; dans ce cas, une suite réelle $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ peut également être vue comme un élément de $\overline{\mathbb{R}}^{\mathbb{N}}$, auquel cas si $u_n \xrightarrow[n\to+\infty]{} +\infty$, on devrait dire qu'elle est divergente si elle est vue dans $\mathbb{R}$, mais convergente si elle est vue dans $\overline{\mathbb{R}}$...

Pour ces raisons, je conseille de bannir l'expression "converge" ou "diverge", et d'utiliser autant que possible "converge dans $\mathbb{R}$", "diverge sans limite". Quant aux suites qui tendent vers $\pm\infty$, on pourra utiliser un peu ce qu'on veut (voir les 3 choix proposés dans la définition 2.4), vu qu'on précisera toujours la limite dans l'expression employée.

2.1.2 Opérations sur les limites et formes indéterminées

Commençons par les opérations qui ne posent pas de problème :

Proposition 2.7. *Soit $(u_n)_{n\in\mathbb{N}}, (v_n)_{n\in\mathbb{N}}$ deux suites réelles, qui tendent vers $\ell \in \mathbb{R} \cup \{\pm\infty\}$ et $\ell' \in \mathbb{R} \cup \{\pm\infty\}$ respectivement, et $\alpha \in \mathbb{R}$.*

27. Rappelons qu'on a décidé de ne pas utiliser la notation $\overline{\mathbb{R}}$, voir la note 13.

— *on suppose que* $\{\ell, \ell'\} \neq \{+\infty, -\infty\}$ [28]. *Alors* $(u_n + v_n)_{n\in\mathbb{N}}$ *a pour limite* $\ell + \ell'$.
— *on suppose que* $(|\ell|, \alpha) \neq (+\infty, 0)$. *Alors* $(\alpha u_n)_{n\in\mathbb{N}}$ *a pour limite* $\alpha\ell$.
— *on suppose que* $\{|\ell|, |\ell'|\} \neq \{+\infty, 0\}$. *Alors* $(u_n v_n)_{n\in\mathbb{N}}$ *a pour limite* $\ell.\ell'$.
— *on suppose* $\ell \neq 0$. *Alors la suite* $\left(\frac{1}{u_n}\right)$ *est bien définie à partir d'un certain rang, et a pour limite* $\frac{1}{\ell}$.
— *on suppose que* $\ell = 0$ *mais que* $\left(\frac{1}{u}\right)$ *est bien définie à partir d'un certain rang. Alors* $\left(\frac{1}{|u_n|}\right)$ *a pour limite* $+\infty$. *Si* (u_n) *est positive à partir d'un certain rang, alors* $\left(\frac{1}{u_n}\right)$ *a pour limite* $+\infty$, *et si* (u_n) *est négative à partir d'un certain rang,* $\left(\frac{1}{u_n}\right)$ *a pour limite* $-\infty$.

Démonstration. La preuve se décompose en de nombreux cas, chacun se traitant avec une manipulation élémentaire des définitions. On ne traite qu'un seul cas, et on invite le lecteur à en écrire d'autres, en suffisance pour que ce type de preuve devienne un automatisme.

Supposons que $(u_n)_{n\in\mathbb{N}}$ converge vers $\ell \in]0, +\infty[$, et $(v_n)_{n\in\mathbb{N}}$ tend vers $+\infty$. On veut montrer que $(u_n v_n)_{n\in\mathbb{N}}$ tend vers $+\infty$. Soit $A \in \mathbb{R}$, et supposons $A > 0$ [29].
En appliquant la définition de la convergence de $(u_n)_{n\in\mathbb{N}}$ pour le nombre $\frac{\ell}{2} > 0$ [30], il existe $n_0 \in \mathbb{N}$ tel que

$$\forall n \geq n_0, \ |u_n - \ell| \leq \frac{\ell}{2}, \quad \text{donc en particulier,} \quad u_n \geq \ell - \frac{\ell}{2} = \frac{\ell}{2}.$$

En appliquant la définition de la divergence de $(v_n)_{n\in\mathbb{N}}$ vers $+\infty$ avec $\frac{2A}{\ell} \in \mathbb{R}$ [31], il existe $n_1 \in \mathbb{N}$ tel que $\forall n \geq n_1, \ v_n \geq \frac{2A}{\ell}$. On pose $n_2 = \max\{n_0, n_1\}$. Alors pour tout $n \geq n_2$,

$$u_n v_n \geq \frac{2A}{\ell}.\frac{\ell}{2} = A$$

d'où le résultat. □

Tous les cas qui ne sont pas traités dans cet énoncé, et qu'on peut résumer par

$$+\infty - \infty, \ \ 0.\infty, \ \ \frac{1}{0} \tag{2.3}$$

(où dans le dernier cas, on n'a pas d'idée du "signe" de 0), sont appelés des formes indéterminées. On peut leur ajouter, en considérant le quotient et les puissances, qu'on n'a pas traités directement dans la proposition précédente [32]

$$\frac{\pm\infty}{\pm\infty}, \quad \frac{0}{0}, \quad 1^{\pm\infty}.$$

28. Attention à faire la différence entre $\{a, b\} = \{x, y\}$ qui implique que a vaut [x ou y], et b vice versa (les ensembles sont donc égaux indépendamment de l'ordre), et $(a, b) = (x, y)$ qui signifie $a = x$ et $b = y$.

29. Je trouve souvent, dans la littérature ou dans la tête des étudiants, la définition

$$\forall A > 0, \ \exists n_0 \in \mathbb{N}, \forall n \geq n_0, u_n \geq A$$

à la place de (2.2) pour la définition du fait que u tend vers $+\infty$. Ce n'est pas faux, mais cette restriction à $A > 0$ au lieu de $A \in \mathbb{R}$ n'est pas naturelle, pas plus que si on avait mis $\forall \varepsilon \in]0, 1]$ dans la définition de la convergence vers un réel. Il est vrai qu'il suffit de considérer ε assez petit, tout comme A assez grand, mais encore une fois je pense qu'il est bon de garder le cas général quand c'est posssible, et d'ajouter des hypothèses quand c'est nécessaire, comme c'est le cas ici, car on va multiplier deux inégalités.

30. Ce choix classique permet d'arriver à la propriété classique qu'une suite qui converge vers un réel positif reste "bien au-dessus de 0" à partir d'un certain rang. On remontre cette propriété rapidement ici, un dessin peut guider le choix du ε considéré ici dans l'utilisation de la limite

31. Si vous lisez ces lignes sans avoir essayé de les écrire vous-même (et peut-être même après avoir essayé), vous pouvez vous demander d'où on sort ce choix ; hé non, ce n'est pas du chapeau qu'on le sort... Il découle simplement du calcul final de cette preuve. Effectivement, une preuve ne s'écrit pas toujours linéairement ; sa trame peut sortir naturellement, comme c'est le cas ici, mais on peut être amené à laisser un trou judicieux ou à corriger un premier choix naïf une fois arrivé au cœur ou à la fin de la preuve. C'est une démarche tout à fait naturelle de la rédaction mathématique, essayez de vous y exercer.

32. Car le quotient de ramène à un inverse et un produit, et que la puissance se réécrit avec des exponentielles et logarithmes. D'ailleurs, en ce qui concerne l'étude d'une suite $(u_n^{v_n})_{n\in\mathbb{N}}$, je conseille de toujours se ramener à l'étude de $(\exp(v_n.\ln(u_n)))_{n\in\mathbb{N}}$.

2.1.3 Théorèmes d'existence de limite

Théorème 2.8. *[Théorème de la limite monotone] Soit $(u_n)_{n\in\mathbb{N}}$ une suite réelle.*

- *si $(u_n)_{n\in\mathbb{N}}$ est croissante et majorée, alors $(u_n)_{n\in\mathbb{N}}$ converge vers un réel.*
- *si $(u_n)_{n\in\mathbb{N}}$ est croissante et n'est pas majorée, alors $(u_n)_{n\in\mathbb{N}}$ tend vers $+\infty$.*

Ainsi, si $(u_n)_{n\in\mathbb{N}}$ est croissante, elle admet une limite dans $\mathbb{R}\cup\{+\infty\}$. De plus, cette limite est

$$\lim_{n\to+\infty} u_n = \sup_{n\in\mathbb{N}} u_n := \sup\left\{u_n, n\in\mathbb{N}\right\}.$$

Remarque 2.9. Attention, si la suite $(u_n)_{n\in\mathbb{N}}$ prend ses valeurs dans $\mathbb{R}\cup\{\pm\infty\}$, on doit adapter l'énoncé en : si $(u_n)_{n\in\mathbb{N}}$ est croissante, alors elle admet une limite dans $\mathbb{R}\cup\{\pm\infty\}$. En effet, on ne peut pas a priori dire que la limite n'est pas $-\infty$, car il ne faut pas oublier de considérer la suite constante $\forall n\in\mathbb{N}, u_n=-\infty$.

Démonstration du théorème 2.8 : Supposons $(u_n)_{n\in\mathbb{N}}$ majorée. L'ensemble $A=\{u_n, n\in\mathbb{N}\}$ est ainsi non vide et majoré, donc par la propriété de la borne supérieure (notée v) page 34), il possède une borne supérieure qu'on note $M=\sup(A)\in\mathbb{R}$. Montrons que $(u_n)_{n\in\mathbb{N}}$ converge vers M. Soit $\varepsilon>0$. Par définition $M-\varepsilon$ n'est pas un majorant de A, donc il existe $n_0\in\mathbb{N}$ tel que $u_{n_0}\geq M-\varepsilon$. Mais alors par monotonie,

$$\forall n\geq n_0,\ \ u_n\geq u_{n_0}\geq M-\varepsilon \quad \text{d'une part, et} \quad u_n\leq M\leq M+\varepsilon \text{ d'autre part.}$$

d'où en particulier $\forall n\geq n_0,\ \ |u_n-M|\leq\varepsilon$ et la convergence de u vers M.

Supposons désormais $(u_n)_{n\in\mathbb{N}}$ non majorée. Soit $A\in\mathbb{R}$: ce n'est pas un majorant de $(u_n)_{n\in\mathbb{N}}$, donc il existe $n_0\in\mathbb{N}$ tel que $u_{n_0}\geq A$. Par croissance de u, on obtient

$$\forall n\geq n_0,\ \ u_n\geq u_{n_0}\geq A$$

d'où le fait que u tend vers $+\infty$. □

Proposition 2.10 ("Théorème des encadrements ou des gendarmes[33]"). *Soit u,v,w trois suites réelles.*

- *On suppose que $u\leq v\leq w$, et que u et w ont une même limite réelle $\ell\in\mathbb{R}$. Alors la suite v est convergente, et converge vers ℓ.*
- *On suppose que $u\leq v$ et que u tend vers $+\infty$. Alors v tend vers $+\infty$.*[34]

Démonstration. Faisons le travail uniquement dans le premier cas de limites réelles, et laissons le lecteur assidu écrire le cas d'une limite infinie. Attention, on ne peut pas dire qu'on "passe à la limite dans les inégalités" car a priori on ne sait pas que la suite v a une limite. On propose deux preuves distinctes, la première, plus élémentaire, la seconde plus expéditive mais utilisant le concept de limite inf/sup[35].

Preuve 1 : Soit $\varepsilon>0$. D'après la convergence de u, il existe $n_0\in\mathbb{N}$ tel que $\forall n\geq n_0,\ |u_n-\ell|\leq\varepsilon$. D'après la convergence de w, il existe $n_1\in\mathbb{N}$ tel que $\forall n\geq n_1,\ |w_n-\ell|\leq\varepsilon$. Posons $n_2=\max\{n_0,n_1\}$. On peut réécrire les propriétés précédentes de la façon suivante :

$$\forall n\geq n_2,\ \ \ell-\varepsilon\leq u_n\leq\ell+\varepsilon, \quad \text{et} \quad \ell-\varepsilon\leq w_n\leq\ell+\varepsilon.$$

Avec l'hypothèse $u\leq v\leq w$, on obtient :

$$\forall n\geq n_2,\ \ \ell-\varepsilon\leq u_n\leq v_n\leq w_n\leq\ell+\varepsilon,$$

ce qui implique

$$\forall n\geq n_2, |v_n-\ell|\leq\varepsilon,$$

d'où le résultat.

33. Il vaut peut-être mieux éviter la version enfantine de la dénomination de ce résultat.
34. On laisse le lecteur trouver la "version $-\infty$" de l'énoncé.
35. Voir l'exercice 2.25.

Preuve 2 : Précisément, l'avantage des limites inf/sup est de toujours exister (dans $\mathbb{R} \cup \{\pm\infty\}$). On peut donc écrire en passant aux limites inf et aux limites sup :

$$\liminf_{n\to+\infty} u_n \leq \liminf_{n\to+\infty} v_n \leq \liminf_{n\to+\infty} w_n, \quad \text{et} \quad \limsup_{n\to+\infty} u_n \leq \limsup_{n\to+\infty} v_n \leq \limsup_{n\to+\infty} w_n.$$

Mais par hypothèse de convergence, on a

$$\liminf_{n\to+\infty} u_n = \limsup_{n\to+\infty} u_n \leq \liminf_{n\to+\infty} w_n = \limsup_{n\to+\infty} w_n = \ell.$$

Ainsi

$$\liminf_{n\to+\infty} v_n = \limsup_{n\to+\infty} v_n = \ell$$

et donc par propriété des limites inf/sup, v converge vers ℓ. □

On va désormais aborder un résultat central de ce chapitre : pour ce faire nous allons aborder la notion de suite extraite :

Définition 2.11. Soit $(u_n)_{n\in\mathbb{N}}$. On appelle suite extraite de $(u_n)_{n\in\mathbb{N}}$ une suite de la forme $(u_{\varphi(n)})_{n\in\mathbb{N}}$ où $\varphi : \mathbb{N} \to \mathbb{N}$ est une fonction strictement croissante, qu'on appellera extraction. Si $\ell \in \mathbb{R}$ est limite d'une suite extraite de $(u_n)_{n\in\mathbb{N}}$, on dira que ℓ est une valeur d'adhérence de $(u_n)_{n\in\mathbb{N}}$.[36]

Remarque 2.12. Une autre notation répandue consiste à considérer $n_0 < n_1 < \dots n_k < \dots$ des entiers, et de noter $(u_{n_k})_{k\in\mathbb{N}}$ la suite extraite. Mais cette notation a ses limites (par exemple si on prend des suites extraites de suites extraites).

Théorème 2.13 (Bolzano-Weierstrass)**.** *Soit $(u_n)_{n\in\mathbb{N}}$ une suite réelle bornée. Alors il existe une suite extraite de $(u_n)_{n\in\mathbb{N}}$ qui est convergente dans $\mathbb{R}$.*

Remarque 2.14. Une fois qu'on aura fait un peu de topologie, on pourra remplacer l'énoncé de ce théorème par le fait que les ensembles $[a, b]$ où a et b sont deux réels, sont des ensembles compacts.

Aussi, une autre façon de donner l'énoncé est de dire que toute suite bornée admet au moins une valeur d'adhérence.

Démonstration. Nous allons utiliser la méthode dite de dichotomie, via une construction par récurrence :

Initialisation : Comme u est bornée, il existe $(a, b) \in \mathbb{R}^2$ telle que $\forall n \in \mathbb{N}, u_n \in [a, b]$. Posons $a_0 = a, b_0 = b$, et $\varphi(0) = 0$.

Première étape :[37] Les ensembles

$$\left\{p \in \mathbb{N}, u_p \in \left[a_0, \frac{a_0 + b_0}{2}\right]\right\} \text{ et } \left\{p \in \mathbb{N}, u_p \in \left[\frac{a_0 + b_0}{2}, b_0\right]\right\}$$

ont pour union $\mathbb{N}$, au moins l'un d'eux est donc infini[38]. Si le premier est infini, on pose $a_1 = a_0$ et $b_1 = \frac{a_0+b_0}{2}$, sinon on pose $a_1 = \frac{a_0+b_0}{2}$ et $b_1 = b$.

Hérédité : on considère construits $(a_k, b_k)_{k\in[\![0,n]\!]}$ tels que

- $\forall k \in [\![0, n-1]\!], a_k \leq a_{k+1}, \quad b_{k+1} \leq b_k$
- $\forall k \in [\![0, n]\!], b_k - a_k = \frac{b-a}{2^k}$
- $\forall k \in [\![0, n]\!], [a_k, b_k]$ contient une infinité d'éléments de u.

On considère

$$\left\{p \in \mathbb{N}, u_p \in \left[a_n, \frac{a_n + b_n}{2}\right]\right\} \text{ et } \left\{p \in \mathbb{N}, u_p \in \left[\frac{a_n + b_n}{2}, b_n\right]\right\}$$

dont un au moins est infini. Si le premier est infini, on pose $a_{n+1} = a_n$ et $b_{n+1} = \frac{a_n+b_n}{2}$, sinon on pose $a_{n+1} = \frac{a_n+b_n}{2}$ et $b_{n+1} = b$. On a bien

$$a_n \leq a_{n+1}, \quad b_{n+1} \leq b_n, \quad b_{n+1} - a_{n+1} = \frac{b-a}{2^{n+1}}, \text{ et } [a_{n+1}, b_{n+1}] \text{ contient une infinité d'éléments de } u,$$

36. **Attention, dans des cadres topologiques plus complexes, ceci n'est pas la bonne définition de valeur d'adhérence.
37. Cette étape est superflue, elle est mise là pour aider le lecteur à saisir la construction qu'on est en train de faire.
38. Il est possible qu'ils soient tous les deux infinis.

d'où la construction au rang suivant.

On construit désormais l'application φ également par récurrence : $\varphi(0) := 0$ et

$$\forall n \in \mathbb{N},\ \varphi(n+1) := \min\{p > \varphi(n), u_p \in [a_{n+1}, b_{n+1}]\},$$

ce dernier minimum étant bien défini car l'ensemble considéré est un ensemble d'entiers naturels, non vide car infini. La fonction φ est bien une extraction car elle est à valeurs dans $\mathbb{N}$ et strictement croissante, et on a de plus

$$\forall n \in \mathbb{N},\quad a_n \leq u_{\varphi(n)} \leq b_n.$$

La suite $(a_n)_{n\in\mathbb{N}}$ est croissante est majorée par b, donc par le théorème 2.8, elle converge vers $\ell \in \mathbb{R}$. De même, la suite $(b_n)_{n\in\mathbb{N}}$ converge, et comme $(a_n - b_n)_{n\in\mathbb{N}}$ tend vers 0, $(b_n)_{n\in\mathbb{N}}$ converge également vers ℓ. On conclut par la propriété des encadrements 2.10 que $(u_{\varphi(n)})_{n\in\mathbb{N}}$ converge vers ℓ. □

2.1.4 Exercices

Exercice 2.15 (Suite bornée)**.** Etant donné une suite réelle $(u_n)_{n\in\mathbb{N}}$, montrer que les propriétés

— $\exists (a,b) \in \mathbb{R}^2,\ \ \forall n \in \mathbb{N}, a \leq u_n \leq b$,

— $\exists M \in \mathbb{R},\ \ \forall n \in \mathbb{N}, |u_n| \leq M$,

sont équivalentes. On dira que $(u_n)_{n\in\mathbb{N}}$ est bornée et on utilisera l'une ou l'autre des assertions de façon équivalente.[39]

Exercice 2.16 (Suite croissante)**.** Etant donné une suite réelle $(u_n)_{n\in\mathbb{N}}$, montrer que les propriétés

— $\forall (p,q) \in \mathbb{N}^2$ tel que $p \leq q, u_p \leq u_q$,

— $\forall n \in \mathbb{N}, u_n \leq u_{n+1}$,

sont équivalentes. On dira que $(u_n)_{n\in\mathbb{N}}$ est croissante et on utilisera l'une ou l'autre des assertions de façon équivalente.

Exercice 2.17.

1. Montrer que la définition 2.1 est équivalente à la suivante[40] :

$$\forall \varepsilon > 0,\quad \exists n_0 \in \mathbb{N},\ \ \forall n \geq n_0,\quad |u_n - \ell| < \varepsilon.$$

2. Prouver l'unicité de la limite d'une suite $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$.
3. Montrez avec la définition de la convergence que la suite $\left(\frac{1}{n^2}\right)_{n\in\mathbb{N}^*}$ converge vers 0.
4. Montrer avec la définition de la divergence vers $+\infty$ que $\lim\limits_{n\to+\infty} \sqrt{n+2} = +\infty$.
5. Montrer avec la définition de la convergence que la suite $((-1)^n)_{n\in\mathbb{N}}$ n'est pas convergente.

Exercice 2.18 (Caractérisation séquentielle de la borne supérieure[41])**.** Soit A une partie non vide de $\mathbb{R}$, et $M \in \mathbb{R} \cup \{+\infty\}$. Montrer que[42]

$$M = \sup(A) \quad \Leftrightarrow \quad \begin{cases} \forall x \in A, \quad x \leq M \\ \exists (x_n)_{n\in\mathbb{N}} \in A^{\mathbb{N}}, \quad x_n \xrightarrow[n\to+\infty]{} M \end{cases}$$

39. Attention, quand on abordera les suites à valeurs dans des espaces plus généraux, par exemple $\mathbb{C}$, seule la deuxième version sera possible.

40. Et donc on utilisera l'une ou l'autre ; il arrive que l'une soit plus commode que l'autre, et le lecteur est invité à s'en rendre compte quand cela arrive. J'en profite pour remarquer qu'il est de très mauvais ton, à l'écrit ou à l'oral, de ne pas faire strictement la différence entre les usages de $\leq$ et de $<$. Quand on a le choix, je pense qu'il est préférable d'utiliser $\leq$, qui se comporte mieux dans les calculs (par exemple, si on passe à la limite).

41. Ce résultat est à connaître.

42. On a donc une nouvelle caractérisation de la borne supérieure. Dans la preuve, distinguez bien les deux cas $M < +\infty$ et $M = +\infty$ quand cela est nécessaire. Notez aussi que la première condition est triviale si $M = +\infty$, c'est le cas où A n'est pas majoré.

Exercice 2.19 (Notations de Landau)**.** Soit $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ deux suites.

— On écrit $u = O(v)$ s'il existe $M \in \mathbb{R}$ et $n_0 \in \mathbb{N}$ tels que

$$\forall n \geq n_0, |u_n| \leq M|v_n|.$$

— On écrit $u = o(v)$ si

$$\forall \varepsilon > 0,\ \exists n_0 \in \mathbb{N},\ \forall n \geq n_0,\ |u_n| \leq \varepsilon |v_n|.$$

— On écrit $u \sim v$ si $u - v = o(v)$, c'est-à-dire si

$$\forall \varepsilon > 0,\ \exists n_0 \in \mathbb{N},\ \forall n \geq n_0,\ |u_n - v_n| \leq \varepsilon |v_n|.$$

1. Montrer que si $u = o(v)$ alors $u = O(v)$.
2. On suppose que v ne s'annule pas[43]. Montrer que[44]

$$u = O(v) \iff \left[\left(\frac{u_n}{v_n}\right)_{n\in\mathbb{N}} \text{ est bornée,}\right]$$

$$u = o(v) \iff \left[\frac{u_n}{v_n} \xrightarrow[n\to+\infty]{} 0\right], \qquad \text{et} \qquad u \sim v \iff \left[\frac{u_n}{v_n} \xrightarrow[n\to+\infty]{} 1\right].$$

Exercice 2.20 (Suites extraites d'une suite convergente)**.** Soit $(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, et φ une extraction.

1. Montrer que pour tout $n \in \mathbb{N}$, $\varphi(n) \geq n$.
2. Montrer que si $(u_n)_{n\in\mathbb{N}}$ tend vers $\ell \in \mathbb{R} \cup \{\pm\infty\}$, alors $(u_{\varphi(n)})_{n\in\mathbb{N}}$ tend aussi vers ℓ.

Exercice 2.21 (Formes indéterminées)**.** Exhiber des exemples de suites qui permettent d'observer que les formes décrites en équation (2.3) sont effectivement indéterminées (par exemple, trouver différents couples de suites (u, v) qui tendent vers $+\infty$ et $-\infty$ respectivement, mais telles que $(u + v)$ tende vers $+\infty$, ou converge vers un réel, ou tende vers $-\infty$, ou diverge sans limite).

Exercice 2.22 (Passage à la limite dans les inégalités (à ne pas confondre avec la proposition 2.10))**.** Soit $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ deux suites réelles qui tendent respectivement vers $\ell \in \mathbb{R} \cup \{\pm\infty\}$ et $\ell' \in \mathbb{R} \cup \{\pm\infty\}$, et telles que pour tout $n \in \mathbb{N}, u_n \leq v_n$. Montrer que $\ell \leq \ell'$.

Exercice 2.23 (Cas particulier du théorème des gendarmes)**.** Soit $(u_n)_{n\in\mathbb{N}}, (v_n)_{n\in\mathbb{N}}$ deux suites réelles, telles que

$$0 \leq |u| \leq v, \qquad \text{et} \qquad v_n \xrightarrow[n\to+\infty]{} 0.$$

Montrer que $(u_n)_{n\in\mathbb{N}}$ a une limite et converge vers 0.

Exercice 2.24 (Suites adjacentes)**.**

1. Soit $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ deux suites réelles telles que
   — $(u_n)_{n\in\mathbb{N}}$ est croissante, $(v_n)_{n\in\mathbb{N}}$ est décroissante,
   — la suite $(v_n - u_n)_{n\in\mathbb{N}}$ converge vers 0.

   Montrer qu'il existe un réel ℓ tel que $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ convergent vers ℓ, et que de plus on a

$$\forall n \in \mathbb{N}, u_n \leq \ell \leq v_n.$$

2. Appliquer le résultat précédent aux suites

$$\forall n \in \mathbb{N}^*,\ \ u_n := \sum_{k=0}^{n} \frac{1}{k!} \quad \text{et } v_n := \sum_{k=0}^{n} \frac{1}{k!} + \frac{1}{n.n!}$$

   et en déduire que le nombre $e := \lim\limits_{n\to\infty} u_n$ est irrationnel.

43. Pour les cas de o et $\sim$, il suffit de supposer que v ne s'annule pas à partir d'un certain rang.
44. Attention, de nombreux élèves oublient les définitions initiales et les remplacent par les caractérisations données ici. C'est évidemment commode à condition d'en avoir le droit. Personnellement, je préfère éviter l'utilisation des notations de Landau pour des suites susceptibles de s'annuler, comme ça le problème ne se pose pas, et je redouble de vigilance quand je manipule une suite qui s'annule souvent.

Exercice 2.25 (Limites inférieures/supérieures)**.** Une application standard du théorème 2.8, est de pouvoir définir les limites inférieures et supérieures d'une suite réelle. L'avantage étant que ces objets existent toujours, alors que la limite d'une suite n'est a priori pas définie. Cela peut être commode quand on veut passer à la limite avant de savoir qu'une suite est convergente, voir par exemple la seconde démonstration de la proposition 2.10. Etant donnée $(u_n)_{n\in\mathbb{N}}$ une suite réelle, on pose dans $\mathbb{R}\cup\{\pm\infty\}$,

$$\forall n\in\mathbb{N},\quad v_n := \inf_{p\in[\![n,+\infty[\![} u_p,\qquad w_n := \sup_{p\in[\![n,+\infty[\![} u_p.$$

1. Montrer que v et w ont une limite dans $\mathbb{R}\cup\{\pm\infty\}$, qu'on note respectivement ℓ_1 et ℓ_2. Justifiez également que
$$\ell_1 = \sup_{n\in\mathbb{N}}\left(\inf_{p\in[\![n,+\infty[\![} u_p\right),\qquad \ell_2 = \inf_{n\in\mathbb{N}}\left(\sup_{p\in[\![n,+\infty[\![} u_p\right).$$
2. Calculer ℓ_1, ℓ_2 dans le cas $\forall n\in\mathbb{N},\ \ u_n = (-1)^n + \frac{1}{n+1}$.
3. Montrer que
$$\text{la suite } (u_n)_{n\in\mathbb{N}} \text{ converge dans } \mathbb{R}\cup\{\pm\infty\} \iff \ell_1 = \ell_2,$$
et qu'alors la limite de $(u_n)_{n\in\mathbb{N}}$ est la valeur commune $\ell_1 = \ell_2$.
4. * Montrer plus généralement que ℓ_1 est la plus petite valeur d'adhérence de la suite $(u_n)_{n\in\mathbb{N}}$ *(c'est-à-dire que $(u_n)_{n\in\mathbb{N}}$ admet une suite extraite dont la limite est ℓ_1, et que si ℓ est la limite d'une suite extraite de (u_n), alors $\ell \geq \ell_1$.)* De même, ℓ_2 est la plus grande valeur d'adhérence de (u_n).

On note
$$\ell_1 = \varliminf_{n\to+\infty} u_n = \liminf_{n\to+\infty} u_n,\qquad \ell_2 = \varlimsup_{n\to+\infty} u_n = \limsup_{n\to+\infty} u_n.$$

Exercice 2.26. * Montrer que si $(u_n)_{n\in\mathbb{N}}$ est une suite bornée, alors on peut en extraire une suite monotone. [45]

2.2 Séries de nombres réels

Nous allons introduire la classe particulière de suites que sont les séries (voir néanmoins la remarque 2.29), pour laquelle on a des critères spécifiques pour étudier la convergence. Notons qu'on calculera peu de séries explicitement dans ce paragraphe, car il n'y a pas de méthode miracle pour ce faire de façon générale. Une série aussi simple que $\sum_{n\in\mathbb{N}^*}\frac{1}{n^3}$ reste encore un mystère pour les mathématiciens [46] ; on verra dans les exemples et dans la suite du cours de nombreux calculs de séries, mais qui sont tous le résultat de méthodes spécifiques.

2.2.1 Définition et premières propriétés

Définition 2.27. On appelle série de nombres réels (ou série numérique) une suite $(S_n)_{n\in\mathbb{N}}$ dont le terme général est de la forme

$$\forall n\in\mathbb{N},\quad S_n = \sum_{k=0}^{n} u_k$$

où $(u_k)_{k\in\mathbb{N}}$ est une suite réelle. On appelle $(u_k)_{k\in\mathbb{N}}$ le terme général de la série, et $(S_n)_{n\in\mathbb{N}}$ la suite des sommes partielles.

45. Ceci est une version un peu améliorée du théorème de Bolzano-Weierstrass : en effet la suite monotone que l'on extrait est convergente, par le théorème de la limite monotone.

46. Il s'agit de la constante d'Apéry, et s'écrit aussi $\zeta(3)$; wikipédia m'informe qu'il s'agit d'un nombre irrationnel (ce qui se montre en étudiant son développement en fractions continues) mais qu'on ne sait pas s'il est transcendant (ni que $\frac{\zeta(3)}{\pi^3}$ est transcendant, ce qui est une question plus naturelle du fait de la formule de $\zeta(2n)$ pour $n\in\mathbb{N}^*$).

Comme pour les suites, on peut bien sûr considérer des séries indicées à partir de $n_0 \in \mathbb{Z}$, mais la présentation sera restreinte au cas $n_0 = 0$, sauf dans certains exemples.

Remarque 2.28. On lit souvent dans la littérature "soit $\sum u_n$ une série", avant de se soucier de la convergence. J'éviterai ces raccourcis d'écriture, certes agréables pour la lecture, mais qui à mon avis prêtent à confusion. En accord avec le choix de non-paresse évoqué en introduction du paragraphe 2, je paierai donc le prix de dire "la série de terme général $(u_n)_{n\in\mathbb{N}}$".

Remarque 2.29. Cette définition peut paraître un peu inutile du fait suivant : il n'y a pas de différence entre suite réelle et série numérique, au sens où toute série numérique est une suite numérique (par définition), et réciproquement, toute suite numérique $v \in \mathbb{R}^{\mathbb{N}}$ peut-être vue comme une série numérique : en effet, en considérant le terme général

$$u_0 = v_0, \qquad \text{et} \qquad \forall n \in \mathbb{N}^*, \ \ u_n := v_n - v_{n-1},$$

on observe en effet facilement [47] que

$$\forall n \in \mathbb{N}, v_n = \sum_{k=0}^{n} u_k.$$

On appelle ce calcul une somme télescopique, et on s'en sert à l'occasion quand on veut appliquer un résultat sur les séries à des suites. Il n'empêche qu'il existe des méthodes spécifiques et riches à l'étude spécifique des séries, ce qui justifie cette définition et les paragraphes qui suivent.

Définition 2.30. — On dit que la série de terme général $(u_k)_{k\in\mathbb{N}}$ converge si la suite des sommes partielles $(S_n)_{n\in\mathbb{N}} = (\sum_{k=0}^{n} u_k)_{n\in\mathbb{N}}$ converge dans $\mathbb{R}$, et on note

$$\sum_{n\in\mathbb{N}} u_k \qquad \text{ou} \qquad \sum_{k=0}^{+\infty} u_k \qquad \text{ou} \qquad \sum_{k\geq 0} u_k$$

la valeur de la limite, à savoir $\lim\limits_{n\to+\infty} \sum\limits_{k=0}^{n} u_k$.

On peut dans ce cas définir

$$\forall n \in \mathbb{N}, \ \ R_n = \sum_{k=n+1}^{+\infty} u_k$$

la suite des restes de la série de terme général $(u_n)_{n\in\mathbb{N}}$.

— Si la suite $(S_n)_{n\in\mathbb{N}}$ tend vers $+\infty$ (idem pour $-\infty$), on peut également se permettre de définir $\sum_{k=0}^{+\infty} u_k = +\infty$.

— Si la suite $(S_n)_{n\in\mathbb{N}}$ diverge sans limite, ne pourra pas écrire $\sum_{k=0}^{+\infty} u_k$.

Voici la première condition nécessaire de convergence (dans $\mathbb{R}$) d'une série :

Proposition 2.31. *Si la série de terme général $(u_k)_{k\in\mathbb{N}}$ converge dans $\mathbb{R}$, alors ce terme général converge vers 0.*

Remarque 2.32. Attention, cette condition n'est pas suffisante !!! On verra au paragraphe 2.2.2 la série harmonique de terme général $\left(\frac{1}{n}\right)_{n\in\mathbb{N}^*}$ qui fournit un contre-exemple (et plus généralement les séries de Riemann de terme général $\left(\frac{1}{n^\alpha}\right)_{n\in\mathbb{N}^*}$ où $\alpha \in]0,1]$).

Quand le terme général u ne tend pas vers 0, on dit parfois que la série est grossièrement divergente.

Démonstration. Notons $\forall n \in \mathbb{N}, S_n = \sum_{k=0}^{n} u_k$ et ℓ limite réelle de $(S_n)_{n\in\mathbb{N}}$. On peut écrire

$$\forall n \in \mathbb{N}^*, \ \ u_n = S_n - S_{n-1}$$

et comme les deux suites $(S_n)_{n\in\mathbb{N}}$ et $(S_{n-1})_{n\geq 1}$ convergent vers la même limite $\ell \in \mathbb{R}$, par différence, (u_n) converge vers 0. □

47. Détaillez ce calcul, soit par calcul direct et manipulation de sommes, soit par récurrence.

Concluons par des remarques élémentaires : on peut additionner et multiplier par des scalaires des séries convergentes (dans $\mathbb{R}$), et ainsi manipuler $\sum_{n\in\mathbb{N}}$ comme on manipule des sommes finies. Attention quand même à bien s'assurer que les sommes mise en jeu sont convergentes. On pourra faire attention à ce pas écrire

$$\sum_{n\in\mathbb{N}}(1-1)=\sum_{n\in\mathbb{N}}1-\sum_{n\in\mathbb{N}}1, \tag{2.4}$$

où on était pourtant parti d'une série bel et bien convergente (car nulle !) [48].

On peut aussi additionner une série qui tend vers $+\infty$ avec une série convergente et conclure que la série somme tend aussi vers $+\infty$. Mais on ne souhaite pas encombrer le lecteur d'énoncés, on lui conseille de garder un peu de vigilance dans la manipulation du symbole $\sum_{n\in\mathbb{N}}$, et s'il est pris d'un doute, de revenir aux sommes partielles et de justifier un passage à la limite.

2.2.2 Séries à termes positifs

Comme le titre le suggère, on considère ici des séries de terme général $u \in \mathbb{R}_+^{\mathbb{N}}$. Ainsi la suite des sommes partielles $(S_n)_{n\in\mathbb{N}}$ est croissante, et donc par le théorème 2.8, $(S_n)_{n\in\mathbb{N}}$ a toujours une limite dans $\mathbb{R}_+\cup\{+\infty\}$. Si S est majorée, alors la série est convergente, et sinon elle diverge vers $+\infty$. Donc ici, pas de danger à manipuler $\sum_{k\in\mathbb{N}}u_k \in \mathbb{R}_+\cup\{+\infty\}$, à condition d'éviter les soustractions.

Evidemment, si le terme général est positif seulement à partir d'un certain rang, tous les éléments de ce paragraphe s'appliquent.

Comparaison série-intégrale : Commençons par un premier moyen d'étudier la convergence d'une série à terme positifs, qui utilise un objet qui n'apparaitra qu'en partie 5 mais que le lecteur a certainement déjà rencontré : la comparaison avec des intégrales. Il peut paraître surprenant de pouvoir décider de la convergence d'une somme discrète en utilisant un objet plus élaboré qu'est une intégrale. La raison est que dans la version continue (les intégrales), on a des moyens de calculs dans les cas où on a accès à des primitives ; en l'occurrence, les méthodes évoquées ici ne permettent pas de calculer les valeurs des sommes considérées (ce qui est un challenge autrement plus élaboré).

Proposition 2.33 (Comparaison série-intégrale)**.** *Soit $f : [0,+\infty[\to\mathbb{R}$ une fonction positive, décroissante, continue par morceaux* [49] [50]*. Alors la série de terme général $(f(k))_{k\in\mathbb{N}}$ converge dans $\mathbb{R}$ si et seulement si la fonction f est intégrable sur $\mathbb{R}_+$ (c'est-à-dire que l'intégrale $\int_0^{+\infty}f(t)dt$ est convergente).*

Remarque 2.34. On retient rarement parfaitement cet énoncé, on le redémontre souvent à chaque fois qu'on veut l'utiliser, ça prend souvent moins de temps et c'est souvent moins ambiguë. Au besoin, on peut même remplacer l'énoncé par un dessin bien choisi où on peut voir les sommes partielles (et la série) comme l'aire de rectangles qui figurent au-dessus ou en dessous du graphe de la fonction f.

Démonstration. Par décroissance de f, on a

$$\forall k\in\mathbb{N}, \forall t\in[k,k+1], \quad f(k)\geq f(t)\geq f(k+1)$$

48. L'objet de gauche de l'égalité ne pose pas de problème (et est nul), l'objet de droite n'a pas de sens car il revient à écrire $+\infty-\infty$.

49. Etant donné une fonction $f : I\to\mathbb{R}$ où I est intervalle, on dira que f est continue par morceaux si :

- premier cas, $I=[a,b]$ est un segment avec $(a,b)\in\mathbb{R}^2$; alors on demande qu'il y ait une subdivision finie $x_0=a<x_1<x_2<\ldots<x_p=b$ (où $p\in\mathbb{N}$) telle que pour tout $i\in[\![0,p-1]\!]$, la restriction $f_{|]x_i,x_{i+1}[}$ est continue est admet des limites réelles en x_i^+ et x_{i+1}^-.
- deuxième cas, si $I=[a,b[$ avec $b\in\mathbb{R}\cup\{+\infty\}$ ou si $I=]a,b]$ avec $a\in\mathbb{R}\cup\{-\infty\}$ ou si $I=]a,b[$ avec $(a,b)\in\overline{\mathbb{R}}$, on dira que f est continue par morceaux sur I si elle l'est sur tout segment compact inclus dans I (via la définition du cas précédent).

50. Cette hypothèse n'est pas vraiment fondamentale, elle permet simplement de parler d'intégrale en restant dans le cadre de l'intégrale de Riemann ; comme les applications de ce résultat seront explicites, il n'y aurait à mon avis pas beaucoup de sens à aller chercher des subtilités comme les fonctions Riemann-intégrables ou Lebesgue-intégrables.

et en intégrant les inégalités,

$$\forall k \in \mathbb{N}, \quad f(k) \geq \int_k^{k+1} f(t)dt \geq f(k+1).$$

Ainsi,

— Si on suppose que f est intégrable sur $\mathbb{R}_+$, alors

$$\forall n \in \mathbb{N}, S_n = \sum_{k=0}^{n} f(k) = \sum_{k=-1}^{n-1} f(k+1) \leq f(0) + \int_0^n f(t)dt \leq f(0) + \int_0^{+\infty} f(t)dt.$$

La suite des sommes partielles étant majorée et croissante (du fait de la positivité de f), on conclut qu'elle converge dans $\mathbb{R}$.

— Si maintenant on suppose S convergeante dans $\mathbb{R}$, alors on peut écrire

$$\forall x \in \mathbb{R}_+, \int_0^x f(t)dt \leq \int_0^{E(x)+1} f(t)dt = \sum_{k=0}^{E(x)} \int_k^{k+1} f(t)dt \leq \sum_{k=0}^{E(x)} f(k) \leq \sum_{k\in\mathbb{N}} f(k)$$

et donc l'intégrale est convergente (encore une fois, car f est positive).

□

Remarque 2.35. On peut plus précisément obtenir :

$$\forall n \in \mathbb{N}, \quad \int_0^{n+1} f(t)dt \leq \sum_{k=0}^{n} f(k) \leq f(0) + \int_0^n f(t)dt$$

ce qui peut fournir un équivalent des sommes partielles quand la série est divergente [51].

Dans le cas convergent, on peut écrire :

$$\forall n \in \mathbb{N}, \quad \int_{n+1}^{+\infty} f(t)dt \leq \sum_{k=n+1}^{+\infty} f(k) \leq \int_n^{+\infty} f(t)dt,$$

qui peut fournir un équivalent des restes.

Exemple 2.36. Ce critère s'applique à l'exemple le plus célèbre de séries classiques, à savoir les séries de Riemann. En effet, pour $\alpha \in \mathbb{R}_+$, on peut montrer que la fonction $\forall x > 0, \quad f(x) = \frac{1}{x^\alpha}$ est positive, décroissante et continue, et donc la série

$$\sum_{n\geq 1} \frac{1}{n^\alpha}$$

est convergente si et seulement si l'intégrale

$$\int_1^{+\infty} \frac{1}{t^\alpha} dt$$

est convergente, ce qui équivaut par calcul à $\alpha > 1$. En effet

$$\int_1^A \frac{1}{t^\alpha} dt = \begin{cases} \ln(A) \text{ si } \alpha = 1 \\ \frac{1}{1-\alpha}\left(\frac{1}{A^{\alpha-1}} - 1\right) \end{cases}$$

d'où le résultat en regardant $A \to +\infty$.

Etant donné que si $\alpha \leq 0$ [52], la série de terme général $\left(\frac{1}{n^\alpha}\right)_{n\in\mathbb{N}^*}$ diverge grossièrement, et on a finalement

$$\sum_{n\in\mathbb{N}^*} \frac{1}{n^\alpha} \text{ est convergente dans } \mathbb{R} \text{ si et seulement si } \alpha > 1.$$

51. Traitez le cas de la série harmonique de terme général $\left(\frac{1}{n}\right)_{n\in\mathbb{N}^*}$ par exemple.

52. Dans le cas $\alpha = 0$, les deux arguments fonctionnent, même s'il est un peu absurde d'appliquer la proposition 2.33 à ce cas.

Maintenant qu'on a en main quelques exemples de séries convergentes et divergentes (essentiellement on a étudié les cas des séries géométriques et des séries de Riemann), on peut étudier la convergence de séries plus élaborées en les comparant à ces séries classiques, via l'énoncé suivant :

Proposition 2.37 (Critères de comparaison pour les séries)**.** *Soit u, v deux suites réelles positives*[53]*. On note*

$$\forall n \in \mathbb{N},\ U_n = \sum_{k=0}^{n} u_k, \quad \textit{et} \quad V_n = \sum_{k=0}^{n} v_k.$$

Dans le cas où les séries sont convergentes, on définit

$$\forall n \in \mathbb{N},\ R_n = \sum_{k=n+1}^{+\infty} u_k, \quad \textit{et} \quad R'_n = \sum_{k=n+1}^{+\infty} v_k$$

les restes.

- *On suppose $u = O(v)$.*
  - *si la série de terme général v converge, alors la série de terme général u converge, et $R = O(R')$.*
  - *si la série de terme général u diverge, alors la série de terme général v diverge et $U = O(V)$.*
- *On suppose $u = o(v)$.*
  - *Si la série de terme général v converge, alors la série de terme général u converge et $R = o(R')$.*
  - *si la série de terme général u diverge, alors la série de terme général v diverge et $U = o(V)$.*
- *Si $u \sim v$, alors les séries de terme général u et de terme général v ont même nature :*
  - *soit elles sont convergentes toutes les deux, et $R \sim V$.*
  - *soit elles divergent toutes les deux vers $+\infty$, et $U \sim V$.*

Démonstration. Nous ne détaillons que le cas des suites équivalentes, les deux autres cas se traitant de façon similaire.

- Supposons $u \sim v$ et $\sum_{k\in\mathbb{N}} v_k$ convergente. Par définition de $u \sim v$ pour $\varepsilon = 42$, il existe $n_0 \in \mathbb{N}$ tel que

  $$\forall n \geq n_0,\ |u_n - v_n| \leq 42|v_n|, \quad \text{ce qui implique, avec la positivité de } v_n, \quad u_n \leq 43 v_n.$$

  Par suite[54]

  $$\forall n \geq n_0, \quad \sum_{k=0}^{n} u_k \leq \sum_{k=0}^{n_0-1} u_k + 43 \sum_{k=n_0}^{+\infty} v_k$$

  donc les sommes partielles associées à $(u_n)_{n\in\mathbb{N}}$ sont bornées, la série est donc convergente. Il reste à obtenir l'estimation des restes : soit $\varepsilon > 0$. Il existe $n_0 \in \mathbb{N}$ tel que

  $$\forall n \geq n_0,\ |u_n - v_n| \leq \varepsilon |v_n|, \quad \text{ce qui implique,} \quad (1-\varepsilon) v_n \leq u_n \leq (1+\varepsilon) v_n.$$

  Et donc

  $$\forall n \geq n_0, (1-\varepsilon) \sum_{k=n}^{+\infty} v_k \leq \sum_{k=n}^{+\infty} u_k \leq (1+\varepsilon) \sum_{k=n}^{+\infty} v_k,$$

  d'où le résultat.

53. Ne négligez pas cette hypothèse ! ! ! Par contre, vous pourrez toujours utiliser cette proposition pour $|u|$ et $|v|$ à la place de u et v si vous considérez de la convergence absolue.

54. On vient en fait de montrer que $u \sim v$ implique $u = O(v)$; on peut conclure en invoquant ce cas si on l'a déjà détaillé.

— Supposons $u \sim v$ et $\sum_{k\in\mathbb{N}} v_k$ divergente. On en déduit cette fois que $v = O(u)$, donc par contraposée du premier point, la série de terme général u est divergente. Soit $\varepsilon > 0$. Il existe $n_0 \in \mathbb{N}$ tel que

$$\forall n \geq n_0, \ \ (1-\frac{\varepsilon}{2})v_n \leq u_n \leq (1+\frac{\varepsilon}{2})v_n.$$

Et donc

$$\forall n \geq n_0, \sum_{k=0}^{n_0-1} u_k + (1-\frac{\varepsilon}{2}) \sum_{k=n_0}^{n} v_k \leq \sum_{k=0}^{n} u_k \leq \sum_{k=0}^{n_0-1} u_k + (1+\frac{\varepsilon}{2}) \sum_{k=n_0}^{n} v_k.$$

Pour gérer l'inégalité de droite, on utilise d'une part que $\sum_{k=n_0}^{n} v_k \leq \sum_{k=0}^{n} v_k$, et d'autre part que comme $\sum_{k=0}^{n} v_k$ tend vers $+\infty$, il existe $n_1 \in \mathbb{N}$ tel que

$$\forall n \geq n_1, \quad \sum_{k=0}^{n_0-1} u_k \leq \frac{\varepsilon}{2} \sum_{k=0}^{n} v_k$$

d'où au final

$$\forall n \geq \max\{n_0, n_1\}, \qquad \sum_{k=0}^{n} u_k \leq (1+\varepsilon) \sum_{k=0}^{n} v_k.$$

Pour l'inégalité de gauche, on utilise à nouveau que $\sum_{k=0}^{n} v_k$ tend vers $+\infty$, et donc il existe n_2 tel que

$$\forall n \geq n_2, \quad \sum_{k=0}^{n_0-1} v_k \leq \frac{\varepsilon}{2} \sum_{k=0}^{n} v_k$$

et donc

$$\begin{aligned} \forall n \geq \max\{n_0, n_2\}, \ \sum_{k=0}^{n} u_k \ &\geq \ \sum_{k=0}^{n_0-1} u_k + (1-\frac{\varepsilon}{2}) \sum_{k=n_0}^{n} v_k \\ &\geq \ (1-\frac{\varepsilon}{2}) \sum_{k=n_0}^{n} v_k = (1-\frac{\varepsilon}{2}) \sum_{k=0}^{n} v_k - (1-\frac{\varepsilon}{2}) \sum_{k=0}^{n_0-1} v_k \\ &\geq \ (1-\frac{\varepsilon}{2}) \sum_{k=0}^{n} v_k - \sum_{k=0}^{n_0-1} v_k \geq (1-\varepsilon) \sum_{k=0}^{n} v_k \end{aligned}$$

Ainsi, pour $n \geq \max\{n_0, n_1, n_2\}$ on a les deux inégalités, d'où le résultat.

$\square$

Remarque 2.38. Si on prend une suite $(u_n)_{n\in\mathbb{N}}$ qui converge vers ℓ et qu'on applique le critère précédent, on retrouve le théorème de Césaro. Distinguons quelques cas :

— si $\ell > 0$, alors $u \sim \ell$ et donc $(u_n)_{n\in\mathbb{N}}$ est positif à partir d'un certain rang. Comme $\sum_{n\in\mathbb{N}} \ell$ est divergente, le résultat précédent montre $\sum_{k=0}^{n} u_k \sim \sum_{k=0}^{n} 1 \sim \ell n$

— si $\ell \leq 0$, on peut par exemple faire une translation pour se ramener au cas précédent : par exemple en étudiant la suite $u + 42 - \ell$ qui est positive à partir d'un certain rang, et équivalente à 42. Ainsi

$$\sum_{k=0}^{n} u_k + n(42-\ell) \sim 42n, \quad \text{ce qui donne bien } \frac{1}{n} \sum_{k=0}^{n} u_k \xrightarrow[n\to+\infty]{} \ell.$$

En réalité, la preuve précédente est très similaire à celle du théorème de Césaro, n'hésitez donc pas à les comparer.

Un autre exemple classique de séries dont on sait étudier la convergence/divergence est donné dans l'exercice 2.49 sur les séries de Bertrand.

On renvoit enfin aux exercices pour les critères classiques de d'Alembert et de Cauchy, voir les exercices 2.50 et 2.51.

2.2.3 Séries à termes réels

Dans ce paragraphe, on aborde les séries à terme général sans signe.

Séries absolument convergentes : Commençons par le premier réflexe à avoir dans cette situation, qui consiste à essayer de se ramener au paragraphe précédent via la proposition suivante :

Proposition 2.39 (Séries absolument convergentes)**.** *Soit $u \in \mathbb{R}^{\mathbb{N}}$. Si la série de terme général $|u|$ est convergente dans $\mathbb{R}$, alors la série de terme général u converge également dans $\mathbb{R}$.*

Dans ce cas, on dira que la série de terme général u est absolument convergente, et on a l'inégalité

$$\left|\sum_{k\in\mathbb{N}} u_k\right| \leq \sum_{k\in\mathbb{N}} |u_k|.$$

Démonstration. Soit donc $u \in \mathbb{R}^{\mathbb{N}}$ telle que la suite des sommes partielles $\forall n \in \mathbb{N},\ \ \Sigma_n := \sum_{k=0}^{n} |u_k|$ a une limite réelle. On s'intéresse à la convergence de $\forall n \in \mathbb{N},\ \ S_n := \sum_{k=0}^{n} u_k$. On propose deux preuves de la convergence de la suite S ; la formule de fin s'obtient ensuite facilement en appliquant l'inégalité triangulaire pour un nombre fini de termes, et par passage à la limite (dont on a justifié l'existence !).

Preuve 1 : Définissons les suites u^+ et u^- par

$$\forall n \in \mathbb{N},\ \ u_n^+ := \max\{u_n, 0\},\quad u_n^- := \max\{-u_n, 0\}.$$

Classiquement, u^+ et u^- sont des suites positives, et $u = u^+ - u^-$, $|u| = u^+ + u^-$. Comme la série de terme général $|u|$ est convergente, et comme $0 \leq u^+ \leq |u|$ et $0 \leq u^- \leq |u|$, par la proposition 2.37 les deux séries de termes généraux u^+ et u^- sont convergentes dans $\mathbb{R}$. Par soustraction, on en déduit que la série de terme général u est aussi convergente dans $\mathbb{R}$.

Preuve 2 : Par la complétude de $\mathbb{R}$, pour montrer que S est convergente, on sait qu'il suffit de montrer qu'elle est de Cauchy. Soit donc $\varepsilon > 0$. Etant convergente, la suite Σ est également de Cauchy, donc il existe $n_0 \in \mathbb{N}$ tel que

$$\forall q \geq p \geq n_0,\ \ |\Sigma_q - \Sigma_p| \leq \varepsilon.$$

On en déduit ainsi que

$$\forall q \geq p \geq n_0, |S_q - S_p| = \left|\sum_{k=p+1}^{q} u_k\right| \leq \sum_{k=p+1}^{q} |u_k| = \Sigma_q - \Sigma_p \leq \varepsilon$$

donc on a montré que la suite S était de Cauchy. Par complétude de $\mathbb{R}$, la suite S est bien convergente. $\square$

Remarque 2.40. * Ces deux preuves sont très intéressantes, et reflètent les remarques du paragraphe 1 : la première preuve se repose sur la propriété de la borne supérieure via le théorème de limite monotone, et la seconde preuve repose sur la complétude. On reverra cette seconde preuve dans un cadre plus général (voir la proposition 2.18 au Chapitre III), et on verra que cela caractérise en fait la complétude. Autrement dit la preuve 1 ci-dessus peut être comprise comme une autre preuve de la complétude de $\mathbb{R}$ à partir de la propriété de la borne supérieure.

Séries semi-convergentes : La proposition 2.39 est une condition suffisante mais non nécessaire ; nous verrons des exemples dans ce paragraphe. Lorsque la série de terme général u n'est pas absolument convergente mais que son terme général tend vers 0, on est pour l'instant démuni pour étudier si elle converge ou non. En passant, on parlera parfois de série semi-convergente pour désigner une série convergente mais non-absolument convergente.

Pour traiter ces séries, on a par exemple le critère spécial des séries alternées :

Proposition 2.41 (Critère spécial des séries alternées)**.** *Soit $(u_n)_{n\in\mathbb{N}}$ une suite réelle, décroissante et qui tend vers 0*[55]*. Alors la série de terme général $((-1)^n u_n)_{n\in\mathbb{N}}$ est convergente. De plus on a l'estimation du reste :*

$$\left|\sum_{k=n+1}^{+\infty} (-1)^k u_k\right| \leq |u_{n+1}|.$$

55. Elle est donc positive, mais inutile de l'indiquer dans les hypothèses.

Démonstration. □

Exemple 2.42. Avec ce critère, on voit que la série $\sum_{n\in\mathbb{N}^*}\frac{(-1)^n}{n}$ est convergente (et comme elle n'est pas absolument convergente d'après l'exemple 2.36, elle est donc semi-convergente).

Plutôt que de démontrer la proposition 2.41, on en donne tout de suite une version plus générale :

Proposition 2.43 (Critère d'Abel)**.** *Soit a et b deux suites réelles telles que*

— *a est une suite réelle décroissante et qui tend vers 0,*

— *b est telle que la suite des sommes partielles $(B_n := \sum_{k=0}^n b_k)_{n\in\mathbb{N}}$ est bornée.*

Alors la série de terme général $(a_n.b_n)_{n\in\mathbb{N}}$ est convergente.

Exemple 2.44. Le critère spécial des séries alternées se déduit de l'énoncé précédent en considérant $b_n = (-1)^n$. En effet dans ce cas, $\forall n \in \mathbb{N},\quad B_n = \begin{cases} 1 \text{ si } n \text{ est pair} \\ 0 \text{ sinon}\end{cases}$.

Démonstration. On détaille la preuve, qui se base sur une transformation d'Abel, qui n'est autre que l'équivalent discret de l'intégration par parties. Dans la situation présente, la méthode de calcul est au moins aussi importante que le résultat lui-même.

Le point de départ de la démonstration est de recalculer $S_n = \sum_{k=0}^n a_k b_k$ avec le constat que $b_n = B_n - B_{n-1}$ pour tout $n \in \mathbb{N}$ (on peut utiliser la formule $B_n = \sum_{k=0}^n b_k$ pour $n < 0$, ce qui donne $B_{-1} = 0$ et donc cette dernière formule reste valable si $n = 0$) :[56]

$$\forall n \in \mathbb{N},\quad S_n = \sum_{k=0}^{n} a_k(B_k - B_{k-1}) = \sum_{k=0}^{n} a_k B_k - \sum_{k=-1}^{n-1} a_{k+1}B_k = a_n B_n + \sum_{k=0}^{n-1}(a_k - a_{k+1})B_k.$$

Etant donné que B est bornée et a tend vers 0, le premier terme aB tend vers 0. Quant au reste, il est absolument convergent (donc convergent par la proposition 2.39) puisque

$$\forall k \in \mathbb{N},\quad |(a_k - a_{k+1})B_k| \le M(a_k - a_{k+1})$$

où M est un majorant de $|b|$ et où on a utilisé la monotonie de a pour se débarasser des valeurs absolues. Cette dernière série majorante est convergente (c'est une série télescopique, et donc sa convergence vient de celle de la suite a), d'où le résultat. □

2.2.4 Pour aller plus loin

On donne ici deux résultats un peu plus délicats, qui interviendront dans la suite du cours. On conclut par quelques pistes non détaillées que le lecteur pourra creuser.

Produit de Cauchy : Il n'est pas difficile de regarder la somme de deux séries de termes généraux $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$. En effet si elles sont toutes les deux convergentes, alors la série de terme général $(a_n + b_n)_{n\in\mathbb{N}}$ sera convergente et on aura

$$\sum_{n\in\mathbb{N}}(a_n + b_n) = \sum_{n\in\mathbb{N}} a_n + \sum_{n\in\mathbb{N}} b_n.$$

56. Etant donné une suite $U \in \mathbb{R}^{\mathbb{N}}$, il est naturel de considérer $U'_n = u_n - u_{n-1}$ (pour tout $n \in \mathbb{N}$; il convient d'étendre u par $u_{-1} = 0$) sa dérivée discrète. Etant donné une suite v, il est aussi naturel de considérer $V_n = \sum_{k=0}^n v_k$ une primitive discrète, et on constate que $V'_n = v_n$ et que $\sum_{k=0}^n U'_k = u_n$ pour tout $n \in \mathbb{N}$, ce qui signifie que les opérations de dérivations discrètes et de primitives discrètes sont bien inverses l'une de l'autre (on n'a pas de problème de constante car on a fixé $u_{-1} = 0$). La transformation d'Abel ressemble donc bien à une intégration par parties, puisqu'avec avec les notations précédentes :

$$\sum_{k=0}^{n} a_k B'_k = a_n B_n - \sum_{k=1}^{n} a'_k B_{k-1}.$$

S'il est bon de se remémorer cette heuristique quand on veut appliquer une transformation d'Abel, on n'apprend pourtant pas la formule précédente par cœur, du fait de la simplicité de sa démonstration, et des problèmes de bord et d'indice qui peuvent varier selon les situations.

Pour le voir il suffit d'écrire cette égalité pour des sommes partielles et d'appliquer la proposition 2.7 sur la limite d'une somme.

Le cas du produit de deux séries est plus délicat, d'une part du point de vue calculatoire, mais surtout du point de vue de la justification de la convergence. Etant donné $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ deux suites réelles, il est naturel en regardant le produit des sommes partielles de poser

$$\forall n\in\mathbb{N},\quad d_n=\sum_{k=0}^{n}a_kb_{n-k}. \tag{2.5}$$

Par contre, dans le cadre des séries absolument convergentes, on a le résultat suivant :

Proposition 2.45. *On suppose que les séries de termes généraux $(a_n)_{n\in\mathbb{N}}\in\mathbb{R}^{\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}\in\mathbb{R}^{\mathbb{N}}$ sont absolument convergentes. Alors la série de terme général défini par (2.5) est absolument convergente et*

$$\sum_{n\in\mathbb{N}}d_n=\sum_{n\in\mathbb{N}}\left(\sum_{k=0}^{n}a_kb_{n-k}\right)=\left(\sum_{n\in\mathbb{N}}a_n\right)\left(\sum_{n\in\mathbb{N}}b_n\right). \tag{2.6}$$

A l'exercice 2.55 on voit que la convergence des séries $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ n'est pas suffisante.

Démonstration. On suit [Pom94, page 144], voir aussi [Gou08b, Page 208].

Supposons d'abord les suites $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ positives. On pose $\forall N\in\mathbb{N},\ D_N=\sum_{n=0}^{N}d_n$. Alors en écrivant

$$\forall N\in\mathbb{N},\ \sum_{n=0}^{N}d_n=\sum_{n=0}^{N}\sum_{k=0}^{n}a_kb_{n-k}=\sum_{(i,j)\in\mathbb{N}^2,i+j\leq N}a_ib_j,\quad\left(\sum_{k=0}^{N}a_k\right)\left(\sum_{k=0}^{N}b_k\right)=\sum_{0\leq i,j\leq N}a_ib_j$$

on obtient que

$$\forall N\in\mathbb{N},\quad D_N\leq\left(\sum_{k=0}^{N}a_k\right)\left(\sum_{k=0}^{N}b_k\right)\leq D_{2N}. \tag{2.7}$$

La première inégalité permet de montrer que la suite $(D_N)_{N\in\mathbb{N}}$ est majorée par $\left(\sum_{n\in\mathbb{N}}a_n\right)\left(\sum_{n\in\mathbb{N}}b_n\right)$ et donc elle est convergente. La suite $(D_{2N})_{N\in\mathbb{N}}$ en est une suite extraite, donc converge vers la même limite, donc par la propriété des encadrements 2.10, on obtient (2.6).

Dans le cas général, on peut appliquer ce qui précède aux suites $(|a_n|)_{n\in\mathbb{N}}, (|b_n|)_{n\in\mathbb{N}}$ qui sont supposées convergentes, ce qui montre que la série de terme général $(\sum_{k=0}^{n}|a_kb_{n-k}|)_{n\in\mathbb{N}}$ est convergente, mais comme

$$\forall n\in\mathbb{N},\ |d_n|\leq\sum_{k=0}^{n}|a_kb_{n-k}|$$

on a bien la convergence absolue de la série de terme général $(d_n)_{n\in\mathbb{N}}$.

Pour obtenir (2.6), on écrit :

$$\begin{aligned}\left|\left(\sum_{k=0}^{N}a_k\right)\left(\sum_{k=0}^{N}b_k\right)-d_N\right|=\left|\sum_{1\leq i,j\leq N,\ i+j>N}a_ib_j\right|&\leq\sum_{1\leq i,j\leq N,\ i+j>N}|a_ib_j|\\&=\left(\sum_{k=0}^{N}|a_k|\right)\left(\sum_{k=0}^{N}|b_k|\right)-\sum_{n=0}^{N}\sum_{k=0}^{n}|a_kb_{n-k}|\end{aligned}$$

et on sait par le point précédent que le terme de droite tend vers 0, d'où le résultat par la proposition 2.10. □

Séries doubles :

Théorème 2.46 (Théorème de Fubini pour les séries doubles). *Soit* $(a_{p,q})_{p\in\mathbb{N},q\in\mathbb{N}}$ *une suites à double indice, c'est-à-dire indéxées par* $\mathbb{N}^2$. *On suppose que la série de terme général* $(\sum_{q\in\mathbb{N}}|a_{p,q}|)_{p\in\mathbb{N}}$ *est convergente*[57]. *Alors les séries*

$$\sum_{p\in\mathbb{N}}\sum_{q\in\mathbb{N}} a_{p,q}, \quad \sum_{q\in\mathbb{N}}\sum_{p\in\mathbb{N}} a_{p,q}$$

sont bien définies, convergentes et de même somme.

Démonstration. [Pom94, 22.4.3] □

2.3 Exercices

Exercice 2.47. Donnons quand même un premier calcul de série simple et utile : montrez que si un réel x a un développement décimal périodique, alors il est rationnel.[58]

Exercice 2.48. Etudiez la convergence de la série géométrique $\left(\sum_{k=0}^{n} a^k\right)_{n\in\mathbb{N}}$ où $a\in\mathbb{R}$, et quand cela est possible, calculez $\sum_{k\in\mathbb{N}} a^k$.

Exercice 2.49 (Séries de Bertrand). Suivant les valeurs de $(\alpha,\beta)\in\mathbb{R}^2$, étudier la convergence de la série de terme général $\left(\frac{1}{n^\alpha \ln(n)^\beta}\right)_{n\in\mathbb{N}^*}$.[59]

Exercice 2.50 (Critère de d'Alembert). Soit $(u_n)_{n\in\mathbb{N}}$ une suite réelle dont les termes sont strictements positifs. On suppose que la suite $\left(\frac{u_{n+1}}{u_n}\right)_{n\in\mathbb{N}}$ a une limite $\ell\in\mathbb{R}_+\cup\{+\infty\}$. Montrez que

— si $\ell<1$ alors la série de terme général $(u_n)_{n\in\mathbb{N}}$ est convergente dans $\mathbb{R}$,

— si $\ell>1$ alors la série de terme général $(u_n)_{n\in\mathbb{N}}$ est divergente.

Exhiber deux exemples pour lesquels $\ell=1$, l'un pour lequel la série est convergente, l'autre pour lesquel la série est divergente.

Exercice 2.51 (Critère de Cauchy). Soit $(u_n)_{n\in\mathbb{N}}$ une suite réelle dont les termes sont strictements positifs. On suppose que la suite $\left((u_n)^{\frac{1}{n}}\right)_{n\in\mathbb{N}}$ a une limite $\ell\in\mathbb{R}_+\cup\{+\infty\}$. Montrez que les mêmes propriétés qu'à l'exercice 2.50 sont valables avec cette valeur de ℓ.[60]

De même, construisez des exemples pour voir qu'on ne peut pas conclure à la nature de la série de terme général u dans le cas $\ell=1$.

Exercice 2.52. Montrez qu'on peut remplacer l'hypothèse de décroissance de la suite $(a_n)_{n\in\mathbb{N}}$ par l'hypothèse (plus faible)

$$\sum_{k=0}^{+\infty}|a_{n+1}-a_n| \;<\; +\infty.$$

---

57. C'est-à-dire

$$\sum_{p\in\mathbb{N}}\sum_{q\in\mathbb{N}}|a_{p,q}|<+\infty.$$

58. La réciproque de cet énoncé est également vraie, mais plus délicate à démontrer.

59. Indication : on pourra appliquer la proposition 2.33 au cas $\alpha=\beta=1$, et déduire tous les autres cas par comparaison. Il est d'ailleurs peut-être plus commode de ne retenir que cette indication et de savoir retrouver le résultat rapidement, plutôt que d'essayer de retenir le résultat par cœur.

60. On peut prouver un critère un peu plus général : s'il existe $r\in[0,1[$ et $n_0\in\mathbb{N}$ tel que $(u_n)^{\frac{1}{n}}\leq r$ pour tout $n\geq n_0$, alors la série de terme général u est convergente.

Exercice 2.53. Etudiez la convergence des séries de termes généraux $(a_n b_n)_{n\in\mathbb{N}}$ où a est une suite réelle, décroissante, qui tend vers 0, et b est de la forme

$$\forall n \in \mathbb{N}, \quad b_n = \sin(\theta n), \qquad \text{ou} \qquad b_n = \cos(\theta n)$$

et où $\theta \not\equiv 0 \; mod \; 2\pi$ (le cas du critère spécial des séries alternées étant inclus ici puisque $\cos(\pi n) = (-1)^n$ pour tout $n \in \mathbb{N}$).

Exercice 2.54 (**Séries et commutativité)**.** Soit $u \in \mathbb{R}^{\mathbb{N}}$.

1. Montrez que si la série de terme général $(u_n)_{n\in\mathbb{N}}$ est absolument convergente et si $\sigma : \mathbb{N} \to \mathbb{N}$ est une bijection, alors la série de terme général $(u_{\sigma(n)})_{n\in\mathbb{N}}$ est absolument convergente et
$$\sum_{k\in\mathbb{N}} u_{\sigma(k)} = \sum_{k\in\mathbb{N}} u_k.$$
2. (Théorème de réarrangement de Riemann). Montrer que si la série de terme général u est semi-convergente, et si $\ell \in \mathbb{R} \cup \{\pm\infty\}$, alors il existe $\sigma : \mathbb{N} \to \mathbb{N}$ une bijection telle que la série de terme général $(u_{\sigma(n)})_{n\in\mathbb{N}}$ a une limite et
$$\lim_{N\to+\infty} \sum_{k=0}^{N} u_{\sigma(k)} = \ell.$$
3. Montrez que si pour tout $\sigma : \mathbb{N} \to \mathbb{N}$ bijective, la série de terme général $(u_{\sigma(n)})_{n\in\mathbb{N}}$ est convergente, alors la série de terme général $(u_n)_{n\in\mathbb{N}}$ est absolument convergente.

Exercice 2.55 (Contre-exemple produit de Cauchy)**.** On suppose $\forall n \in \mathbb{N}^*, a_n = b_n = \frac{(-1)^n}{\sqrt{n}}$, et on définit $(d_n)_{n\in\mathbb{N}}$ par (2.5) (avec $a_0 = b_0 = 0$). Montrer que $\forall n \geq 2, |d_n| \geq 1$ et donc la série de terme général $(d_n)_{n\in\mathbb{N}}$ est grossièrement divergente.

---

3 Fonctions définies sur une partie de $\mathbb{R}$ et à valeurs réelles

On s'intéresse ici à l'étude des fonctions $f : D \to \mathbb{R}$ où $D \subset \mathbb{R}$ [61]. Comme pour les suites, du fait que $\mathbb{R}$ (comme espace d'arrivée) est un corps, on peut considérer une structure d'algèbre sur l'ensemble des fonctions de D dans $\mathbb{R}$. On pourra aussi considérer la composition, quand cela est bien défini.

Aussi, du fait que $\mathbb{R}$ est ordonné, on peut parler de fonction majorée, minorée, bornée, (strictement) croissante, décroissante, monotone. Quand D est symétrique par rapport à 0 ($\forall x \in D, -x \in D$), on peut aussi parler de fonction paire ou impaire.

Pour le moment, on ne fait pas d'hypothèse particulière sur D. Souvent, la littérature présente les énoncés sur des intervalles. C'est un peu vicieux car on les applique souvent pour des ensembles plus généraux. Et surtout, ça cache un peu les moments où il devient vraiment important que l'ensemble de définition devienne un intervalle. On ne cherche donc pas à généraliser pour le plaisir [62], mais bien pour insister sur les moments où les hypothèses rentreront en jeu.

3.1 Limites d'une fonction

Contrairement aux suites, on ne s'intéresse pas qu'aux limites quand la variable tend vers $+\infty$.

61. Dans tout le paragraphe, D sera un sous-ensemble de $\mathbb{R}$ qui désignera l'ensemble de définition des fonctions.
62. Si j'étais amené à parler d'une fonction $f : \mathbb{Q} \to \mathbb{R}$ continue, je garderais une certaine vigilance ; mais honnêtement, ça n'arrive pas très souvent.

3.1.1 Définitions

Il y a plusieurs situations à considérer. Voici la première :

Définition 3.1. [Limite réelle en un point adhérent] Soit $f : D \to \mathbb{R}$, et $a \in \overline{D}$. On dit que f converge vers $\ell \in \mathbb{R}$ en a si

$$\forall \varepsilon > 0, \exists \alpha > 0, \quad \forall x \in]a-\alpha, a+\alpha[\cap D, \quad |f(x) - \ell| \le \varepsilon,$$

et on note alors

$$\lim_{x\to a} f(x) = \ell, \qquad \text{ou } \lim_a f = \ell \qquad \text{ou encore} \qquad f(x) \xrightarrow[x\to a]{} \ell.$$

Comme pour les suites (voir l'exercice 2.17), cette limite est unique, et on invite le lecteur à en rédiger la preuve. D'ailleurs, le lecteur devrait se convaincre de l'intérêt de l'hypothèse "a est adhérent à D" dans la définition précédente.

Remarque 3.2. On peut aussi considérer des limites à gauche ou à droite en remplaçant $]a-\alpha, a+\alpha[\cap D$ par $]a-\alpha, a[\cap D$ et $]a, a+\alpha[\cap D$ respectivement, dans les cas où ces ensembles sont non vides (c'est-à-dire que a est adhérent à $]-\infty, a[\cap D$ ou à $]a, +\infty[\cap D$ respectivement). On peut les noter [63]

$$\lim_{x\to a, x>a} f(x) \text{ ou } \lim_{x\to a^+} f(x) \text{ ou éventuellement } f(a^+).$$

On prendra garde à ce que a a été exclu de ces intervalles, ce qui n'était pas le cas dans la définition 3.1. D'ailleurs, avec cette définition, si $a \in D$ et si $\lim_a f = \ell$, alors nécessairement $f(a) = \ell$. Si on préfère exclure le point a (ce qui ne change que si $a \in D$), on peut considérer la notion de limite épointée : f a pour limite épointée ℓ en a si

$$\forall \varepsilon > 0, \exists \alpha > 0, \quad \forall x \in (]a-\alpha, a+\alpha[\cap D) \setminus \{a\}, \quad |f(x) - \ell| \le \varepsilon.$$

Afin de lier limite et [limites à gauche et à droite], on peut montrer, dans le cas où a est adhérent aux deux ensembles $]-\infty, a[\cap D$ et $]a, +\infty[\cap D$, que

— si $a \in \overline{D} \setminus D$, alors f admet pour limite ℓ en a si et seulement si f admet pour limite ℓ à droite et à gauche en a ;

— si $a \in D$, alors f admet pour limite ℓ en a si et seulement si f admet pour limite ℓ à droite et à gauche en a, *et que* $f(a) = \ell$.

Les autres situations viennent du fait que x peut tendre vers $\pm\infty$ d'une part, et que la limite peut valoir $\pm\infty$ d'autre part. Nous ne détaillons pas tous les cas, mais en précisons deux :

Définition 3.3. Soit $f : D \to \mathbb{R}$.

— si $a \in \overline{D}$, on dit que f tend vers $+\infty$ en a si [64]

$$\forall A \in \mathbb{R}, \exists \alpha > 0, \forall x \in]a-\alpha, a+\alpha[\cap D, \quad f(x) \ge A.$$

— si D n'est pas majoré, on dit que f converge vers $\ell \in \mathbb{R}$ en $+\infty$ si

$$\forall \varepsilon > 0, \exists A \in \mathbb{R}, \forall x \in [A, +\infty[\cap D, \quad |f(x) - \ell| \le \varepsilon.$$

3.1.2 Caractérisation séquentielle

Ce résultat permet de lier les définitions de limites de fonctions à partir de celles sur les suites. Cela peut servir entre autre à transporter des propriétés déjà vues pour les suites au cadre des fonctions. Comme on a vu que la liste des cas pour les fonctions est plus grand que pour les suites, ça peut être bien commode.

63. Les deux dernières notations sont à utiliser avec vigilance ; effectivement l'élément a^+ n'existe pas (et donc f ne peut pas y être définie !), il faut le comprendre comme une notation, et l'éviter si elle nous semble mystérieuse.

64. Remarquez (et démontrez !) que dans les conditions de cette définition, on a nécessairement $a \in \overline{D} \setminus D$.

Proposition 3.4 (Caractérisation séquentielle)**.** *Soit* $f : D \to \mathbb{R}$, $\ell \in \mathbb{R} \cup \{\pm\infty\}$, *et* $a \in \overline{D}$ *ou* $a = \pm\infty$ *si* D *n'est pas majoré/minoré. Alors*

$$\lim_{x \to a} f(x) = \ell \iff \begin{cases} \forall (x_n)_{n\in\mathbb{N}} \in D^{\mathbb{N}} \text{ telle que } x_n \xrightarrow[n\to+\infty]{} a, \\ f(x_n) \xrightarrow[n\to+\infty]{} \ell \end{cases}.$$

Démonstration. Considérons le cas $\ell \in \mathbb{R}$ et $a \in \overline{D}$, les autres cas se traitant de façon similaire.
$\Rightarrow$: Soit $(x_n)_{n\in\mathbb{N}}$ une suite à valeurs dans D et qui converge vers a. Soit $\varepsilon > 0$. Comme f converge vers ℓ en a, il existe $\alpha > 0$ tel que

$$\forall x \in]a-\alpha, a+\alpha[\cap D, \quad |f(x) - \ell| \leq \varepsilon.$$

On applique la définition de la convergence de $(x_n)_{n\in\mathbb{N}}$ avec $\frac{\alpha}{2} > 0$: il existe $n_0 \in \mathbb{N}$ tel que

$$\forall n \geq n_0, |x_n - a| \leq \frac{\alpha}{2}, \quad \text{donc en particulier} \quad a - \alpha < a - \frac{\alpha}{2} \leq x_n \leq a + \frac{\alpha}{2} < a + \alpha.$$

On obtient ainsi

$$\forall n \geq n_0, \quad |f(x_n) - \ell| \leq \varepsilon.$$

$\Leftarrow$: Montrons la contraposée[65] : on suppose donc qu'il existe ε_0 tel que

$$\forall \alpha > 0, \exists x \in]a-\alpha, a+\alpha[\cap D, \quad |f(x) - \ell| > \varepsilon_0.$$

Pour tout $n \in \mathbb{N}$, on applique cette propriété pour $\alpha = \frac{1}{n+1} > 0$, ce qui nous donne un élément $x_n \in D$ tel que $a - \frac{1}{n+1} < x_n < a + \frac{1}{n+1}$ et $|f(x_n) - \ell| > \varepsilon_0$. D'après la proposition 2.10, la suite $(x_n)_{n\in\mathbb{N}}$ ainsi définie converge vers a, mais $(f(x_n))_{n\in\mathbb{N}}$ ne converge pas vers ℓ, ce qu'il fallait démontrer. □

3.1.3 Propriétés

On passe vite sur ce paragraphe, qui se calque aux résultats sur les suites, voir 2.1.2 et 2.1.3.

On peut additionner, multiplier les limites, diviser quand le dénominateur est non nul sur un voisinage du point a où on étudie la limite. Comme cette situation n'existe par pour les suites, faisons l'effort de préciser un énoncé dans le cas de la composition des limites :

Proposition 3.5 (Composition des limites)**.** *Soit* D, E *deux ensembles de* $\mathbb{R}$, $f : D \to \mathbb{R}$, $g : E \to \mathbb{R}$ *telles que* $f(D) \subset E$[66], *et* $a \in \overline{D}$. *On suppose*

$$\lim_{x\to a} f(x) = b, \qquad b \in \overline{E}, \qquad \text{et } \lim_{y \to b} g(y) = \ell \in \mathbb{R} \cup \{\pm\infty\}.$$

Alors $g \circ f$ *a pour limite* ℓ *en* a.

65. *Autant que possible, je conseille de faire la différence entre une preuve par contraposée et un raisonnement par l'absurde. Si on veut montrer $P \Rightarrow Q$, dans le premier cas on va supposer $non(Q)$ et montrer $non(P)$. Dans le second cas, on va supposer à la fois P et $non(Q)$, et chercher une contradiction. Il arrive souvent que les étudiants foncent sur le raisonnement par l'absurde : ils supposent P et $non(Q)$, et ils montrent $non(P)$ sans jamais utiliser P (en faisant exactement ce qu'on aurait fait pour le raisonnement par contraposée). Ils concluent à une contradiction simplement du fait qu'ils ont P et $non(P)$. Ce raisonnement n'est pas incorrect, mais je pense qu'il prête à confusion. Selon moi, un raisonnement par l'absurde est plus élaboré car on part avec plus d'informations, ce qui est plus délicat à gérer quand on cherche où aller dans notre raisonnement. Tout au moins dans l'apprentissage du cours, faites bien attention qu'un raisonnement par l'absurde est assez rare, et qu'il n'est pas anodin. J'en profite aussi pour dire qu'un raisonnement direct, quand il est possible, me paraît souvent (je n'ai pas dit "toujours" !) plus élégant qu'un raisonnement par l'absurde. C'est peut-être une question de goût, mais il m'est d'avis que si vous produisez un raisonnement par l'absurde pour un résultat, vous devriez vous demander si c'était nécessaire, et le cas échéant, vous entraîner à aussi écrire le raisonnement direct. Pour progresser, on doit s'efforcer à trouver les raisonnements les plus naturels et les plus simples. Même si au final, le bon raisonnement restera le vôtre, puisque c'est bien celui que vous produirez pendant les épreuves.

66. Cette condition assure que la fonction $g \circ f$ est bien définie.

Démonstration. On propose deux preuves, la seconde permettant de voir la mise en œuvre de la propriété de caractérisation séquentielle (Proposition 3.4). La première preuve sera donnée seulement dans le cas $a \in \mathbb{R}$, $b \in \mathbb{R}$ et $\ell \in \mathbb{R}$, alors que la seconde preuve s'applique à tous les cas[67].

Preuve 1 : Soit $\varepsilon > 0$. Par définition de la limite de g, il existe $\alpha > 0$ tel que

$$\forall y \in]b-\alpha, b+\alpha[\cap E,\ |g(y)-\ell| \leq \varepsilon. \tag{3.1}$$

On applique désormais la définition de la limite de f en a pour la valeur $\frac{\alpha}{2} > 0$, et on obtient l'existence de $\beta > 0$ tel que

$$\forall x \in]a-\beta, a+\beta[\cap D,\ |f(x)-b| \leq \frac{\alpha}{2}.$$

Ainsi, si $x \in]a-\beta, a+\beta[\cap D$, $f(x) \in]b-\alpha, b+\alpha[$ et $f(x) \in f(D) \subset E$, donc on peut appliquer (3.1) à $y = f(x)$, ce qui donne

$$\forall x \in]a-\beta, a+\beta[\cap D,\ |g(f(x))-\ell| \leq \varepsilon,$$

d'où le résultat.

Preuve 2 : Soit $(x_n)_{n\in\mathbb{N}} \in D^{\mathbb{N}}$ une suite qui converge vers a. Nous allons appliquer successivement 3 fois la proposition 3.4, deux fois dans le sens direct, une fois dans le sens réciproque.

Comme f converge vers b en a, on a $f(x_n) \xrightarrow[n\to+\infty]{} b$. Comme g converge vers ℓ en b, on a $g \circ f(x_n) \xrightarrow[n\to+\infty]{} \ell$. Comme on a montré cette propriété pour toute suite $(x_n)_{n\in\mathbb{N}} \in D^{\mathbb{N}}$, on conclut bien que $g \circ f(x) \xrightarrow[x\to a]{} \ell$. □

On conclut par deux exercices qui peuvent être vus comme des adaptations du théorème 2.8 et de la proposition 2.10 au cas des fonctions :

3.1.4 Exercices

Exercice 3.6. Soit $f : D \to \mathbb{R}$ monotone, et $a \in \overline{D\cap]-\infty, a[}$[68]. Montrer que f admet une limite (dans $\mathbb{R} \cup \{\pm\infty\}$) à gauche en a.

Si $a \in \overline{D\cap]a, +\infty[}$, montrer que f admet une limite à droite en a.

Exercice 3.7. Soit f, g, h trois fonctions réelles définies sur D telles que $f \leq g \leq h$, et $a \in \overline{D}$. Montrer que si f et h ont pour limite $\ell \in \mathbb{R}$ en a, alors g aussi.

3.2 Continuité

3.2.1 Définition

Définition 3.8 (Continuité en un point, sur un ensemble). Soit $f : D \to \mathbb{R}$ une fonction.

— Etant donné $a \in D$, on dit que f est continue en a si

$$\lim_{x\to a} f(x) = f(a).$$

— Etant donné $E \subset D$, on dit que f est continue sur E si f est continue en tout point $x \in E$.

67. Pour la bonne raison que les cas ont été traités dans la proposition 3.4.
68. On pourrait autoriser $a = \pm\infty$ si D n'est pas majoré/minoré.

Quand on dit que la fonction f est continue, on sous-entend qu'elle est continue sur tout son ensemble de définition D, et on notera $f \in C^0(D, \mathbb{R})$ ou plus simplement $C^0(D)$ s'il n'y a pas d'ambiguité sur l'espace d'arrivée. Il n'empêche que je conseille de toujours préciser l'ensemble dont on est en train de parler ; cela évitera les confusions, et nous rappellera de vérifier qu'on ne dit pas de bêtise.

On peut également parler de l'ensemble des points de continuité de f.

On dira que f est discontinue en $a \in D$ si elle n'est pas continue en a, et on peut également parler de l'ensemble des points de discontinuité de f.

Remarque 3.9 (Prolongement par continuité)**.** Il est important de noter qu'on a demandé $a \in D$ et non $a \in \overline{D}$. En effet, pour que la continuité en a ait un sens, il faut au préalable que f soit définie en a.

Néanmoins, si $f : D \to \mathbb{R}$ et a une limite $\ell \in \mathbb{R}$ en $a \in \overline{D} \setminus D$, alors on peut définir

$$\begin{array}{rccl} \widetilde{f}: & D \cup \{a\} & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \begin{cases} f(x) \text{ si } x \in D \\ \ell \text{ si } x = a \end{cases} \end{array}$$

qui définit une fonction qui prolonge f et qui est continue en a. On l'appelle prolongement par continuité de f en a. Par exemple, on note souvent sinc la fonction sinus cardinal, qui est le prolongement par continuité en 0 de $x \in \mathbb{R}^* \mapsto \frac{\sin(x)}{x}$.

Remarque 3.10. Comme pour les limites, on peut parler de continuité à droite et/ou à gauche en un point. On dira par exemple que $f : D \to \mathbb{R}$ est continue à droite en $a \in D$ si

$$\lim_{x \to a, x > a} f(x) = f(a).$$

On peut montrer que f est continue en a si et seulement si elle est continue à gauche et à droite en a.

3.2.2 Théorèmes issus de la continuité

Dans ce paragraphe, on énonce et démontre 3 résultats fondamentaux de l'analyse réelle, qui reposent sur la notion de continuité.

Théorème 3.11 (Théorème des valeurs intermédiaires)**.** *Soit I un intervalle de $\mathbb{R}$, et $f : I \to \mathbb{R}$ une fonction continue sur I. Alors $f(I)$ est un intervalle.*

Démonstration. Soit $(u, v) \in f(I)$. Il existe $(a, b) \in I$ tels que $u = f(a)$ et $v = f(b)$. Supposons par exemple $a \leq b$ et $u \leq v$, les autres cas étant similaires[69]. Soit enfin $w \in]u, v[$. On veut montrer que $w \in f(I)$. On considère

$$A = \{x \in [a, b], \quad f(x) \leq w\}.$$

L'ensemble A est non vide (il contient a) et majoré, il possède donc une borne supérieure qu'on note M. On veut montrer que $f(M) = w$. Comme souvent, on procède avec les deux propriétés de M :

— M est un majorant de A, donc

$$\forall x \in]M, b], \quad x \notin A \text{ donc } f(x) > w.$$

De plus, $f(b) > w$, donc par continuité, il existe $\alpha > 0$ tel que $\forall x \in]b - \alpha, b], \quad f(x) > w$, d'où le fait que $M < b$ et que l'intervalle $]M, b]$ est non vide. On peut alors faire tendre x vers M par la droite, et par continuité de f, on obtient $f(M) \geq w$.

69. Mieux, on peut en fait déduire la propriété sur les "petits o" à partir de la propriété sur les équivalents, et vice versa, puisque l'on peut définir l'un avec l'autre. Ce type d'astuce est bon à repérer pour éviter de reproduire des preuves, mais ce n'est pas une catastrophe, à mon avis, de passer à côté, tant qu'on arrive à produire une preuve non "astucieuse" en un temps limité.

— Pour $n \in \mathbb{N}^*$, $M - \frac{1}{n}$ n'est pas un majorant de A, donc il existe $x_n \in A$ tel que $x_n \geq M - \frac{1}{n}$ [70]. Ainsi
$$\forall n \in \mathbb{N}^*, \quad M - \frac{1}{n} \leq x_n \leq M, \qquad \text{et} \qquad f(x_n) \leq w.$$
Par la proposition 2.10, la suite $(x_n)_{n\in\mathbb{N}}$ converge vers M, et par la proposition 3.4 et la continuité de f, $(f(x_n))_{n\in\mathbb{N}}$ converge vers $f(M)$, d'où $f(M) \leq w$.

Au final $f(M) = w$ et $w \in f(I)$, d'où le résultat. □

Corollaire 3.12. *Soit $f : I \to \mathbb{R}$ continue sur l'intervalle I, et $a < b$ deux éléments de I. Si $f(a) \leq 0$ et $f(b) \geq 0$, alors il existe $c \in [a, b]$ tel que $f(c) = 0$.*

Théorème 3.13 (Théorème des bornes)**.** *Soit $a < b$ deux réels, et $f : [a, b] \to \mathbb{R}$ une fonction continue sur $[a, b]$. Alors f est bornée sur $[a, b]$, et atteint ses bornes, c'est-à-dire qu'il existe $(c_1, c_2) \in [a, b]^2$ tels que*
$$\min_{[a,b]} f = f(c_1), \qquad \textit{et} \qquad \max_{[a,b]} f = f(c_2).$$

Démonstration. Montrons que f est majorée et atteint son supremum. On pourra appliquer cela à $-f$ pour obtenir le cas de l'infimum. En fait, nous allons simplement montrer qu'elle atteint son supremum, qui sera alors nécessairement réel [71].

Comme $[a, b]$ est non vide, f admet un supremum $M \in \mathbb{R} \cup \{+\infty\}$. Par propriété de la borne supérieure (via l'exercice 2.18), il existe $(x_n)_{n\in\mathbb{N}}$ une suite de $[a, b]$ telle que $f(x_n) \xrightarrow[n\to+\infty]{} M$. Par le théorème de Bolzano-Weierstrass 2.13, il existe une extraction φ et $x_\infty \in [a, b]$ [72] tels que $x_{\varphi(n)} \xrightarrow[n\to+\infty]{} x_\infty$. Par continuité de f, on a donc $f(x_{\varphi(n)}) \xrightarrow[n\to+\infty]{} f(x_\infty) \in \mathbb{R}$, et par unicité de la limite [73], on en déduit $M = f(x_\infty)$, ce qu'il fallait démontrer. □

Pour terminer ce paragraphe, on s'intéresse à la question de la bijectivité des fonctions réelles, et de la continuité des applications réciproques.

Théorème 3.14 ("Théorème" de la bijection)**.** *Soit I un intervalle, et $f : I \to \mathbb{R}$. Si f est continue et strictement croissante, alors $J := f(I)$ est un intervalle, $f : I \to J$ est une bijection [74], et $f^{-1} : J \to I$ est continue et strictement croissante.*

Remarque 3.15. J'ai respecté l'usage en énonçant ce résultat classique et en le nommant théorème, mais je mets en garde sur les incompréhensions qu'il amène. La première partie du théorème, en effet, est essentiellement une redite du théorème des valeurs intermédiaires, et les différentes assertions (f bijective et $f(I)$ intervalle) n'utilisent chacune qu'une des hypothèses (la stricte monotonie et la continuité respectivement). En effet, commençons par remarquer que le résultat
$$\Big[f : D \to \mathbb{R} \text{ strictement croissante}\Big] \Longrightarrow \quad f \text{ injective sur } D \tag{3.2}$$
est élémentaire et ne nécessite pas d'hypothèse de continuité. Ainsi, si f est seulement supposée strictement monotone [75] alors $f : D \to f(D)$ est clairement une bijection.

Le fait que $f(I)$ est un intervalle découle du Théorème des valeurs intermédiaire, et n'utilise pas l'hypothèse de stricte monotonie.

70. On a refait ici l'exercice 2.18.

71. Attention, ici on s'autorise à parler de supremum dans $\mathbb{R} \cup \{+\infty\}$. Si vous n'aimez pas, vous pouvez tout à fait montrer d'abord que f est majorée, par exemple en raisonnant par l'absurde, et en utilisant les mêmes ingrédients que dans la suite de la preuve.

72. Le fait que x_∞ soit dans $[a, b]$ vient d'un passage à la limite dans $\forall n \in \mathbb{N}, a \leq x_{\varphi(n)} \leq b$. Dit autrement, on utilise le caractère fermé de $[a, b]$.

73. Et le fait qu'une suite extraite d'une suite convergente est également convergente, et converge vers la même limite, voir Exercice 2.20.

74. Attention, on fait l'abus ici de dénoter f à la fois pour la fonction $f : I \to \mathbb{R}$ et pour la fonction $f : I \to J$; n'oubliez pas que la donnée d'une fonction contient a priori la donnée de ses ensembles de départ et d'arrivée, donc la rigueur devrait nous inciter à donner des noms différents à ces deux fonctions. D'ailleurs, seule la seconde est bijective, et f^{-1} désigne son inverse.

75. Et D n'est pas nécessairement un intervalle, ni f nécessairement continue.

La seule nouveauté du théorème 3.14 réside donc dans le caractère continu de f^{-1}[76], le caractère strictement croissant étant pour sa part élémentaire et n'utilisant pas l'hypothèse de continuité.

Démonstration. Comme on le note à la Remarque 3.15, il reste à montrer que f^{-1} est continue et strictement croissante. Commençons par ce second point : si $(u, v) \in J$ avec $u < v$, alors il existe $(x, y) \in I$ tels que $u = f(x)$ et $v = f(y)$. Par contraposée de la stricte croissance de f, on obtient $x \leq y$, et comme $u \neq v$, $x \neq y$. Ainsi $x = f^{-1}(u) < y = f^{-1}(v)$, d'où la stricte croissance de f^{-1}. Soit désormais $u \in J$, et $x \in I$ son antécédent. On va montrer que f^{-1} est continue en u.

Preuve 1 : Supposons que x n'est pas une extrêmité de I. Alors il existe $\varepsilon_0 > 0$ tel que $[x - \varepsilon_0, x + \varepsilon_0] \subset I$. Soit désormais $\varepsilon \in]0, \varepsilon_0]$. Par la stricte croissance de f et le fait que J est un intervalle, J contient $[f(x-\varepsilon), f(x+\varepsilon)]$. De plus $f(x-\varepsilon) < f(x) < f(x+\varepsilon)$. On pose $\alpha = \min\left\{\frac{f(x)-f(x-\varepsilon)}{2}, \frac{f(x+\varepsilon)-f(x)}{2}\right\}$, et alors

$$f(x-\varepsilon) < f(x) - \alpha < f(x) < f(x) + \alpha < f(x+\varepsilon).$$

Ainsi

$$\forall v \in]u - \alpha, u + \alpha[, f(x-\varepsilon) < v < f(x+\varepsilon)$$

ce qui donne en passant par f^{-1} strictement croissante :

$$f^{-1}(u) - \varepsilon = x - \varepsilon < f^{-1}(v) < x + \varepsilon = f^{-1}(u) + \varepsilon$$

d'où la continuité de f^{-1} en u. Si x est une extrêmité de I, on fait une preuve similaire en remplaçant $]x - \alpha, x + \alpha[$ par $]x - \alpha, x]$ ou $[x, x + \alpha[$.

Preuve 2 : Par l'absurde, si f^{-1} n'est pas continue en u, par le théorème de caractérisation séquentielle, il existe $(u_n)_{n\in\mathbb{N}}$ une suite d'éléments de J qui converge vers u mais telle que $(x_n := f^{-1}(u_n))_{n\in\mathbb{N}}$ ne converge pas vers x. Ainsi il existe $\varepsilon > 0$ et φ une extraction telle que

$$\forall n \in \mathbb{N}, |x_{\varphi(n)} - x| > \varepsilon.$$

Au moins un des ensembles $\{n \in \mathbb{N}, x_{\varphi(n)} > x + \varepsilon\}$ et $\{n \in \mathbb{N}, x_{\varphi(n)} < x - \varepsilon\}$ est de cardinal infini. Supposons par exemple que le premier est infini : il existe donc en particulier $n_0 \in \mathbb{N}$ tel que $x_{\varphi(n_0)} > x + \varepsilon$ et comme $(x_{\varphi(n_0)}, x)$ sont dans I et I est un intervalle, $x + \varepsilon \in I$. Par monotonie stricte de f, il vient

$$\forall n \in \mathbb{N}, \ \ u_{\varphi(n)} > f(x+\varepsilon) \text{ d'une part, et d'autre par } f(x+\varepsilon) > f(x).$$

On en déduit une contradiction avec le fait que $(u_{\varphi(n)})_{n\in\mathbb{N}}$ converge vers $u = f(x)$. □

3.2.3 Uniforme continuité

Définition 3.16. Etant donnée une fonction $f : D \to \mathbb{R}$, on dit que f est uniformément continue sur D si

$$\forall \varepsilon > 0, \exists \alpha > 0, \ \forall (x, y) \in D, \ |x - y| < \alpha \ \Rightarrow \ |f(x) - f(y)| \leq \varepsilon.$$

Remarque 3.17. Il convient de bien réécrire les deux phrases définissant la continuité de f sur D, et l'uniforme continuité de f sur D, qui s'écrivent respectivement :

$$\forall x \in D, \ \forall \varepsilon > 0, \ \exists \alpha > 0, \ \forall y \in]x - \alpha, x + \alpha[\cap D, \quad |f(x) - f(y)| \leq \varepsilon.$$

et

$$\forall \varepsilon > 0, \ \exists \alpha > 0, \ \forall x \in D, \ \forall y \in]x - \alpha, x + \alpha[\cap D, \quad |f(x) - f(y)| \leq \varepsilon.$$

La différence entre ces deux formules tient uniquement en la place de $\forall x \in D$; dans le premier cas, ce quantificateur arrive avant le quantificateur $\exists \alpha > 0$, ce qui signifie que α a le droit de dépendre de x.

76. Même si, me semble-t-il, seule la première partie de l'énoncé figure dans la plupart des livres de lycée. La seconde partie a sans doute disparu des programmes car elle était considérée trop compliquée. Mais du coup, il semble que personne ne se rende compte que l'énoncé est devenu creux et incohérent...

Alors que dans la seconde version, α est introduit avant x, et donc le même α doit convenir pour tous les x choisis [77].

De cette remarques, on comprend facilement que la continuité uniforme implique la continuité.

Dans l'exercice 3.28 on donne des fonctions continues non uniformément continues. Il n'est pas anodin que dans cet exercice on ait cherché des contre-exemples définis sur $\mathbb{R}_+$ ou $]0,1]$. En fait, si la fonction est définie sur un segment de $\mathbb{R}$, on a bien équivalence entre continuité et continuité uniforme :

Théorème 3.18 (Théorème de Heine). *Soit $a<b$ deux réels, et $f:[a,b]\to\mathbb{R}$ une fonction continue. Alors f est uniformément continue sur $[a,b]$.*

Démonstration. On présente deux preuves, la première étant plus élémentaire, la second utilisant l'approche topologique de la compacité qu'on abordera au paragraphe 1.5 du Chapitre III.

Preuve 1 : Par contraposée, on va montrer que si f n'est pas uniformément continue, alors on peut construire un point de discontinuité dans $[a,b]$. On suppose donc qu'il existe $\varepsilon_0>0$ tel que

$$\forall \alpha>0, \exists (x,y)\in[a,b]^2, \quad |x-y|<\alpha \;\text{ et }\; |f(x)-f(y)|>\varepsilon_0.$$

Pour tout $n\in\mathbb{N}^*$, on applique la phrase précédente à $\alpha=\frac{1}{n}$, ce qui nous donne deux suites $(x_n)_{n\in\mathbb{N}^*}$ et $(y_n)_{n\in\mathbb{N}^*}$ telles que :

$$\forall n\in\mathbb{N}^*, \quad |x_n-y_n|<\frac{1}{n}, \quad \text{ et } \quad |f(x_n)-f(y_n)|>\varepsilon_0. \tag{3.3}$$

D'après le théorème de Bolzano-Weierstrass 2.13, il existe une extraction φ et $x_\infty\in[a,b]$ [78] tels que

$$x_{\varphi(n)} \xrightarrow[n\to+\infty]{} x_\infty.$$

Mais d'après la première partie de (3.3), on a aussi $(y_{\varphi(n)})_{n\in\mathbb{N}}$ qui converge vers x_∞, ce qui montre d'après la deuxième partie de (3.3) que f ne peut pas être continue en x_∞.

Preuve 2 : Soit $\varepsilon>0$. Par continuité de f, on a

$$\forall x\in[a,b], \exists \alpha_x>0, \;\; \forall y\in]x-\alpha_x, x+\alpha_x[\cap[a,b], \;\; |f(y)-f(x)|\le\frac{\varepsilon}{2}. \tag{3.4}$$

On considère

$$[a,b]\subset \bigcup_{x\in[a,b]} \left]x-\frac{\alpha_x}{2}, x+\frac{\alpha_x}{2}\right[$$

qui est un recouvrement de $[a,b]$ par des ouverts, donc par compacité, on peut en extraire un sous-recouvrement fini

$$[a,b]\subset \bigcup_{i=1}^{N} \left]x_i-\frac{\alpha_{x_i}}{2}, x_i+\frac{\alpha_{x_i}}{2}\right[\tag{3.5}$$

où $N\in\mathbb{N}^*$ et $(x_1,\dots,x_N)\in[a,b]^N$. On pose

$$\alpha=\frac{1}{2}\min\{\alpha_{x_i}\}_{i\in[\![1,N]\!]}.$$

Soit $(x,y)\in[a,b]$ tels que $|x-y|<\alpha$. Par (3.5), il existe $i\in[\![1,N]\!]$ tel que $x\in\left]x_i-\frac{\alpha_{x_i}}{2}, x_i+\frac{\alpha_{x_i}}{2}\right[$, et on constate que

$$|y-x_i|\le|y-x|+|x-x_i|<\alpha+\frac{\alpha_{x_i}}{2}\le\alpha_{x_i}$$

et donc on peut appliquer la continuité en x_i, c'est-à-dire l'équation (3.4) à la fois à x et à y qui sont dans $]x_i-\alpha_{x_i}, x_i+\alpha_{x_i}[$, ce qui donne

$$|f(y)-f(x)|\le|f(y)-f(x_i)|+|f(x)-f(x_i)|\le\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon$$

et achève la preuve. □

77. Par contre, à l'intérieur de chaque formule, on pourrait échanger les quantificateurs $\forall$ qui se côtoient. Par exemple, la continuité sur D peut aussi s'écrire :

$$\forall\varepsilon>0, \; \forall x\in D, \exists\alpha>0, \; \forall y\in]x-\alpha, x+\alpha[\cap D, \; |f(x)-f(y)|\le\varepsilon.$$

78. Voir note 72.

Voyons une condition suffisante d'uniforme continuité qui est souvent utilisée en pratique, et qui a son intérêt propre :

Définition 3.19. Soit $f : D \to \mathbb{R}$ une fonction. On dit que f est lipschitzienne sur D si

$$\exists k \in \mathbb{R},\ \ \forall (x,y) \in D^2,\ |f(y) - f(x)| \leq k|y - x|. \tag{3.6}$$

On dira que f est k-lipschitzienne si la constante k convient dans (3.6).

3.2.4 Exercices

Exercice 3.20. Soit $f : D \to \mathbb{R}$ une fonction monotone. Montrer que l'ensemble des points de discontinuité de f est dénombrable.[79]

Exercice 3.21. Trouvez un contre-exemple au Théorème 3.11 si on ne suppose plus I un intervalle.

Exercice 3.22. Montrer que toute fonction polynomiale de degré impair admet une racine réelle.

Exercice 3.23. Soit $f : I \to \mathbb{R}$ où I est de la forme $[a, +\infty[$ ou $]-\infty, b]$ ou $\mathbb{R}$. Montrer que si f est continue et a une limite réelle en $+\infty$, $-\infty$ ou les deux (respectivement), alors f est bornée sur I.

Exercice 3.24. Lorsqu'une fonction $f : E \to F$ (où E, F sont deux espaces métriques, voir Chapitre III[80].) est bijective, continue et que son inverse $f^{-1} : F \to E$ est continu, on dit que f est un homéomorphisme[81].

Ainsi, le théorème de la bijection affirme que si I est un intervalle, et $f : I \to \mathbb{R}$ est strictement monotone, alors $f : I \to f(I)$ est un homéomorphisme.

Trouvez un contre-exemple à ce résultat quand I n'est plus un intervalle, c'est-à-dire une fonction $f : D \to \mathbb{R}$ continue et strictement monotone, mais dont la réciproque $f^{-1} : f(D) \to D$ n'est pas continue.

Pour conclure, on laisse en exercice un résultat qui exprime que dans le cas d'une fonction continue sur un intervalle, l'hypothèse de stricte monotonie est équivalente à l'hypothèse d'injectivité.

Exercice 3.25 (*).

1. Soit $f : I \to \mathbb{R}$ continue et injective sur un intervalle I. Montrer que f est strictement monotone sur I.
   Ainsi, du fait de (3.2), il y a bien équivalence entre les deux notions de continuité et d'injectivité pour une fonction strictement monotone[82]. Donnez un contre-exemple à ce résultat quand I n'est plus un intervalle.
2. Soit $g : D \to J$ une application surjective et monotone, où $D \subset \mathbb{R}$ est "quelconque" et J est un intervalle. Montrer que g est continue.

Exercice 3.26. Montrer que si $f : D \to \mathbb{R}$ est lipschitzienne sur D, alors elle est uniformément continue sur D.

79. Rappelons qu'il y a deux conventions possibles pour le mot dénombrable ; certains appellent dénombrable tout ensemble qui s'injecte dans $\mathbb{N}$, et d'autres appellent dénombrable les ensembles qui sont en bijection avec $\mathbb{N}$, ce qui exclut les ensembles finis. On a choisi la première convention ; on aurait pu dire "au plus dénombrable" pour éviter toute confusion.

80. Si le lecteur n'est pas déjà familier avec cette notion, il peut penser sans perdre l'intérêt de la remarque à E, F deux sous-ensembles de $\mathbb{R}$

81. Il s'agit juste d'un mot ; ce qui compte, c'est l'idée fréquente en algèbre qu'on demande à l'inverse d'avoir les mêmes propriétés que la bijection initiale ; c'est parfois automatique (si φ est un morphisme de groupe bijectif, son inverse φ^{-1} est aussi un morphisme de groupe ; on l'appellera isomorphisme de groupe), mais pas toujours, comme le montre l'exercice. Voir également le théorème d'isomorphisme de Banach (Chapitre IX) sur cette question.

82. Mais notez bien qu'un sens n'utilise pas la continuité, alors que sa réciproque oui !

Exercice 3.27 (Caractérisation séquentielle de l'uniforme continuité)**.** Soit $f : D \to \mathbb{R}$. Montrer que

$$\begin{array}{c} f \text{ est uniformément} \\ \text{continue sur } D \end{array} \iff \left\{ \begin{array}{r} \forall (x_n)_{n\in\mathbb{N}} \in D^{\mathbb{N}}, (y_n)_{n\in\mathbb{N}} \in D^{\mathbb{N}} \text{ telle que } (x_n - y_n) \xrightarrow[n\to+\infty]{} 0, \\ \big(f(x_n) - f(y_n)\big) \xrightarrow[n\to+\infty]{} 0. \end{array} \right.$$

Exercice 3.28. 1. Exhiber une fonction $f : \mathbb{R}_+ \to \mathbb{R}$ qui est continue sur $\mathbb{R}$, mais pas uniformément continue. Exhiber un autre contre-exemple défini sur $]0, 1]$.

2. Montrer que si $f : \mathbb{R}_+ \to \mathbb{R}$ est uniformément continue, alors il existe $(a, b) \in \mathbb{R}^2$ tels que

$$\forall x \in \mathbb{R}_+, |f(x)| \leq ax + b.$$

3.3 Dérivabilité

Dans ce paragraphe, on suppose les fonctions définies sur un ensemble $D \subset \mathbb{R}$. Le lecteur qui souhaite simplifier les définitions peut supposer que $D = I$ est un intervalle. Dans le cas général, on fera les hypothèses qu'il faut pour que les limites aient un sens.

3.3.1 Définition et premières propriétés

Définition 3.29. Soit $f : D \to \mathbb{R}$, et $a \in D$ tel que $a \in \overline{D \setminus \{a\}}$. On dit que f est dérivable en a si la fonction [83]

$$x \in D \setminus \{a\} \longmapsto \frac{f(x) - f(a)}{x - a} \tag{3.7}$$

admet une limite réelle en a. On notera alors $f'(a)$ cette limite.

Si $E \subset D$ est tel que f est dérivable en tout point de E, on dit que f est dérivable sur E et on peut définir la fonction $f' : E \to \mathbb{R}$.

Remarque 3.30. Bien sûr, si on préfère regarder des limites en 0, on peut faire un simple changement de variable :

$$f'(x) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \lim_{h\to 0} \frac{f(a + h) - f(a)}{h}. \tag{3.8}$$

Remarque 3.31. On a bien précisé que la limite devait être réelle. Quand la limite existe et vaut $\pm\infty$, on dit que la fonction n'est pas dérivable, mais qu'elle possède une tangente verticale. C'est le cas de la fonction $x \in \mathbb{R}_+ \mapsto \sqrt{x}$ en 0.

Remarque 3.32. *Si E est un espace vectoriel normé et si $f : I \to E$ où I est un intervalle de $\mathbb{R}$, alors on peut définir de la même manière la dérivabilité de f, la limite du taux de variation étant prise pour la norme de E. Par exemple, si $E = \mathbb{R}^n$, on peut voir que $f : I \to \mathbb{R}^n$ sera dérivable en $x_0 \in I$ si et seulement si les fonctions coordonnées de f sont dérivables en x_0.

Remarque 3.33. En considérant l'exemple $\forall x \in \mathbb{R},\ f(x) = |x|$, on observe facilement que la fonction n'est pas dérivable en 0. On peut néanmoins considérer la dérivabilité à droite et à gauche en regardant les limites à droite et à gauche en a de la fonction (3.7), limites qu'on notera $f'_d(a)$ et $f'_g(a)$ respectivement (ou éventuellement $f'(a^+), f'(a^-)$, voir la note 63), quand celles-ci existent.

Comme la fonction (3.7) n'est pas définie en a le point en lequel on calcule la limite, on déduit (voir Remarque 3.2) que f est dérivable en a si et seulement si elle est dérivable à gauche et à droite en a et que $f'_d(a) = f'_g(a)$.

Concluons par la condition nécessaire (mais non suffisante comme le montre la valeur absolue en 0) classique :

83. Appelée taux de variation.

Proposition 3.34. *Si $f : D \to \mathbb{R}$ est dérivable en $a \in D$, alors f est continue en a.*

Remarque 3.35. On peut donner une version gauche/droite : si f est dérivable à gauche (resp. droite) en a, alors f est continue à gauche (resp. droite) en a.

Démonstration. Ecrivons

$$\forall x \in D \setminus \{a\}, \quad f(x) - f(a) = \frac{f(x) - f(a)}{x - a}.(x - a)$$

et comme le premier terme a une limite réelle et le second une limite nulle, le produit tend bien vers 0, et donc $f(x)$ converge vers $f(a)$ quand x tend vers a. □

Pour conclure, on peut donner une interprétation de la dérivabilité en terme de développement limité :

Remarque 3.36. La dérivabilité de $f : D \to \mathbb{R}$ en a est en fait équivalente à l'existence d'un développement limité de f à l'ordre 1 en a [84], c'est-à-dire que f est dérivable en a si et seulement s'il existe $(\alpha, \beta) \in \mathbb{R}^2$ tel que

$$f(a + h) = \alpha + \beta h + o(h)$$

et alors $\alpha = f(a)$ et $\beta = f'(a)$.

3.3.2 Opérations

Il peut devenir rapidement compliqué de calculer des dérivées à partir de la définition. On a souvent recours à une décomposition de la fonction étudiée à l'aide des opérations usuelles, afin de se ramener à des fonctions dont on a appris par cœur la dérivée.

Proposition 3.37. *Soit $f, g : D \to \mathbb{R}$ deux fonctions dérivables en $a \in D$, et $\lambda \in \mathbb{R}$. Alors*

— *$(f + g)$, λf et $f.g$ sont dérivables en a, et*

$$(f + g)'(a) = f'(a) + g'(a), \quad (\lambda f)'(a) = \lambda f'(a), \quad \textit{et} \quad (fg)'(a) = f'(a)g(a) + f(a)g'(a).$$

— *on suppose de plus que $g(a) \neq 0$; alors $\frac{f}{g}$ est bien définie sur un voisinage de a, et y est dérivable, avec*

$$\left(\frac{f}{g}\right)'(a) = \frac{g(a)f'(a) - g'(a)f(a)}{g(a)^2}.$$

Démonstration. Le calcul pour la somme est élémentaire. Pour le produit, on écrit simplement :

$$\begin{aligned}\frac{f(x)g(x) - f(a)g(a)}{x - a} &= \frac{f(x)g(x) - f(a)g(x) + f(a)g(x) - f(a)g(a)}{x - a} \\ &= g(x).\frac{f(x) - f(a)}{x - a} + f(a).\frac{g(x) - g(a)}{x - a}.\end{aligned}$$

Pour le quotient, on commence par constater que comme g est dérivable en a, par la proposition 3.34, g est continue en a. Comme de plus $g(a) \neq 0$, g ne s'annule pas sur un voisinage de a, ce qui implique que le quotient $\frac{f}{g}$ est bien défini sur ce voisinage.

Pour le calcul, on écrit

$$\frac{\frac{f(x)}{g(x)} - \frac{f(a)}{g(a)}}{x - a} = \frac{\frac{f(x)g(a) - f(a)g(x)}{g(x)g(a)}}{x - a} = \frac{1}{g(x)g(a)}.\frac{\big[f(x)g(a) - f(a)g(a)\big] + \big[f(a)g(a) - f(a)g(x)\big]}{x - a}$$

d'où le résultat. □

Proposition 3.38. *Soit $f : D \to \mathbb{R}$, $g : E \to \mathbb{R}$ et $f(D) \subset E$. On suppose f dérivable en a et g dérivable en $f(a)$. Alors $g \circ f$ est dérivable en a et*

$$(g \circ f)'(a) = g'(f(a)).f'(a).$$

84. Attention, on verra au paragraphe 3.3.4 que ce n'est pas le cas pour les dérivées d'ordre supérieur.

Je laisse la preuve en exercice, en conseillant d'utiliser la remarque 3.36.

On conclut avec la dérivée de la fonction réciproque :

Proposition 3.39. *Soit I un intervalle qui n'est pas restreint à un singleton, $f : I \to J$ continue et bijective, $a \in I$ et $b = f(a)$. Si f est dérivable en a et $f'(a) \neq 0$, alors f^{-1} est dérivable en b et*

$$(f^{-1})'(b) = \frac{1}{f'(a)} = \frac{1}{f'(f^{-1}(b))}.$$

Démonstration. Pour tout $y \in J$, il existe un unique $x \in I$ [85] tel que $y = f(x)$. On peut alors écrire

$$\forall y \neq b, \quad \frac{f^{-1}(y) - f^{-1}(b)}{y - b} = \frac{f^{-1}(f(x)) - f^{-1}(f(a))}{f(x) - f(a)} = \frac{1}{\frac{f(x)-f(a)}{x-a}}$$

où on a utilisé dans la dernière formule que $x \neq a$ puisque $y \neq b$ et que f est bijective.

Pour conclure, il reste à voir que x converge vers a quand y tend vers b ; comme $x = f^{-1}(y)$, cela découle de la continuité de f^{-1} donnée par le théorème 3.14. □

3.3.3 Étude des variations d'une fonction

Vous le savez depuis le lycée, les dérivées sont un outil très performant pour étudier les variations d'une fonction. Nous allons le démontrer dans ce paragraphe. Chaque résultat a une preuve assez simple, mais c'est plutôt l'enchaînement des résultats qui est non trivial et sur lequel il convient de prendre un peu de recul. On retiendra en particulier l'importance des idées de l'optimisation via l'utilisation du théorème des bornes.

Théorème 3.40 (Théorème de Rolle)**.** *Soient $a < b$, $f : [a, b] \to \mathbb{R}$ une fonction continue sur $[a, b]$, dérivable sur $]a, b[$, et telle que $f(a) = f(b)$. Alors il existe $c \in]a, b[$ tel que $f'(c) = 0$.*

Démonstration. Comme f est supposée continue, on peut appliquer le théorème des bornes 3.13, et donc f réalise son minimum et son maximum en $c \in [a, b]$ et $d \in [a, b]$ respectivement. Faisons deux cas :

— c ou d est dans $]a, b[$. Supposons par exemple qu'il s'agit de c, l'autre situation étant similaire. On a alors

$$\forall x \in [a, b], \quad f(x) \geq f(c),$$

et comme $c \in]a, b[$, il existe $h_0 > 0$ tel que

$$\forall h \in]0, h_0[, \qquad \begin{array}{lll} f(c+h) \geq f(c) & \text{donc} & \frac{f(c+h)-f(c)}{h} \geq 0 \\ f(c-h) \geq f(c) & \text{donc} & \frac{f(c-h)-f(c)}{-h} \leq 0, \end{array}$$

ce qui donne quand h tend vers 0, $f'(c) \geq 0$ et $f'(c) \leq 0$, donc c convient.

— si c et d sont tous les deux au bord de l'intervalle $[a, b]$, du fait que $f(a) = f(b)$, la fonction f est constante sur $[a, b]$. Et donc $c' = \frac{a+b}{2}$ convient.

□

Corollaire 3.41 (Egalité des accroissements finis)**.** *Soient $a < b$, $f : [a, b] \to \mathbb{R}$ continue sur $[a, b]$ et dérivable sur $]a, b[$. Alors il existe $c \in]a, b[$ tel que*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

85. Attention, x est bien ici une fonction de y, en l'occurence $x = f^{-1}(y)$; on a préféré alléger la notation, mais cela ne doit pas exempter le lecteur qui souhaite faire des approximations similaires d'une grande vigilance quand il voudra "passer à la limite en x" ; n'oubliez jamais qu'une variable introduite par un "il existe" dépend des variables qui ont été introduites précédemment.

Démonstration. Il suffit d'appliquer le théorème de Rolle à la fonction [86]

$$\forall x \in [a,b], \quad g(x) = f(x) - \frac{f(b)-f(a)}{b-a}(x-a).$$

□

Remarque 3.42. Ce résultat est parfois appelé théorème des Accroissements finis, et il est en fait une généralisation du théorème de Rolle. Personnellement je préfère parler d'égalité des accroissements finis, par distinction avec l'inégalité des accroissements finis (voir ci-dessous), résultat plus faible, mais qui est le seul à persister quand l'espace d'arrivée est de dimension supérieure à 1, voir le paragraphe 1.5 du Chapitre VII.

Corollaire 3.43 (Inégalité des accroissements finis)**.** *Soit $f : [a,b] \to \mathbb{R}$, continue sur $[a,b]$ et dérivable sur $]a,b[$. Alors*

$$\forall (x,y) \in [a,b], \quad |f(y)-f(x)| \leq \left(\sup_{c\in]a,b[} |f'(c)|\right) |y-x|.$$

Remarque 3.44. Le "nombre" $\sup_{c\in]a,b[} |f'(c)|$ est a priori un élément de $\mathbb{R}_+ \cup \{+\infty\}$, car on n'a pas supposé que f' était bornée sur $]a,b[$. Evidemment dans ce cas l'énoncé n'est pas faux, mais inutile.

Démonstration. Il suffit d'appliquer l'égalité des accroissements finis sur chaque intervalle $[x,y]$ si $x<y$ sont deux éléments de $[a,b]$. □

On conclut cette partie avec le résultat le plus notable sur l'utilisation des dérivées :

Théorème 3.45 (Variation et signe de la dérivée)**.** *Soit I un intervalle, et $f : I \to \mathbb{R}$ une fonction dérivable.*

— *f est croissante sur I si et seulement si $f' \geq 0$ sur I.*

— *f est strictement croissante sur I si et seulement si $\Big[$ $f' \geq 0$ sur I et l'ensemble $\{x \in I, f'(x) = 0\}$ est d'intérieur vide* [87] $\Big]$.

On peut également caractériser le caractère (strictement) décroissant et le caractère constant en remplaçant $f' \geq 0$ par $f' \leq 0$ et $f' = 0$ respectivement.

Remarque 3.46. Attention au cas de la stricte monotonie. En effet on a effectivement

$$f' > 0 \text{ sur } I \quad \Longrightarrow \quad f \text{ strictement croissante sur } I$$

mais la réciproque est fausse, comme vous pourrez vous en convaincre en exhibant un contre-exemple.

Démonstration. — Supposons f croissante. Etant donné $x \in I$, on a

$$\forall y \in I \text{ tel que } y > x, \quad f(y) \geq f(x), \quad \text{donc} \quad \frac{f(y)-f(x)}{y-x} \geq 0.$$

Par dérivabilité de f en x, en prenant la limite $y \to x^+$, on obtient bien $f'(x) \geq 0$.

— Supposons $f' \geq 0$. Etant donnés $a < b$ dans I, l'égalité des accroissements finis donne effectivement que $f(a) \leq f(b)$.

— On raisonne par contraposée. Supposons f croissante mais non strictement croissante. Alors il existe $x < y$ dans I tel que $f(x) = f(y)$, et f est donc constante sur $[x,y] \subset I$. On en déduit que f' est nulle sur $]x,y[$.

— Réciproquement, si $f' \geq 0$ mais f' est nulle sur un intervalle $]x,y[\subset I$ avec $y > x$, alors f est constante sur cet intervalle, donc non strictement croissante.

□

86. Qui, à une constante près, mesure la différence entre la fonction f et la droite qui joint $(a, f(a))$ et $(b, f(b))$. Pensez à toujours faire un dessin plutôt que d'essayer de retenir par cœur un nombre incalculable de formules !

87. C'est-à-dire qu'il n'existe pas d'intervalle $J \subset I$ non vide et non restreint à un singleton sur lequel f' est nulle, voir aussi le paragraphe 1.2.

3.3.4 Dérivées d'ordre supérieur

Si une fonction $f : D \to \mathbb{R}$ est dérivable, on peut considérer la fonction $f' : D \to \mathbb{R}$, et on peut se demander si elle est continue, voire dérivable. Dans le cas où f' est dérivable (disons sur D, sinon on doit restreindre l'ensemble de définition), on dit que f est deux fois dérivable sur l'ensemble D, et on peut considérer $(f')'$, qu'on note plus commodément f'' ou encore $f^{(2)}$, et qu'on appellera dérivée seconde de f. On peut itérer ce processus autant de fois que possible, et considérer la fonction $f^{(n)}$ pour $n \geq 2$ quand celle-ci est bien définie. Il convient de définir par convention $f = f^{(0)}$.

Remarque 3.47. Attention, on n'a pas défini de dérivabilité seconde d'une fonction $f : D \to \mathbb{R}$ en un point $a \in D$. Cela n'est pas interdit, mais alors cela signifie que f est dérivable sur un voisinage de a, ce qui permet de considérer la fonction f' sur ce voisinage, et que cette dernière est dérivable en a.

Il convient souvent de considérer une classe de fonctions un peu meilleures que n-fois dérivables :

Définition 3.48. Pour $n \in \mathbb{N}$, on dira qu'une fonction $f : D \to \mathbb{R}$ est de classe C^n si elle est n-fois dérivable sur D et si la fonction $f^{(n)}$ est continue sur D.
On dira que $f : D \to \mathbb{R}$ est de classe C^∞ si elle est de classe C^n pour tout $n \in \mathbb{N}$.[88]

Il est facile de vérifier (par exemple par une récurrence élémentaire) que la somme, le produit, le quotient, la composée[89] de deux fonctions n-fois dérivables sont n-fois dérivables.

Pour obtenir des formules, cette fois, il est également facile de voir que la dérivée n-ième d'une somme de fonctions n'est autre que la somme des dérivées n-ième. Pour le produit, on a la formule de Leibniz, qui se retient grâce à son parallèle avec la formule du binôme de Newton :

Proposition 3.49 (Formule de Leibniz). *Soit $n \in \mathbb{N}$, et $f, g : D \to \mathbb{R}$ deux fonctions n-fois dérivables sur D. Alors (fg) est n-fois dérivable sur D et*

$$(fg)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(k)} g^{(n-k)}$$

La preuve est laissée au lecteur et se fait par une simple récurrence, qui ressemble fortement à la preuve de la formule du binôme de Newton[90].

Remarque 3.50. Pourquoi s'arrêter là et ne pas se demander la dérivée n-ième d'un quotient ou d'une composée ? En fait, cela est beaucoup moins connu, mais il y a bien une formule, dite de Faà di Bruno, et qui donne la dérivée n-ième d'une composée. On ne l'écrit pas ici, elle est assez indigeste, et rarement utile[91], mais comme c'est une question naturelle que vous devriez vous être posée, on voulait vous prévenir.

Quant au quotient, comme on a facilement accès à la dérivée n-ième de la fonction $i(x) = \frac{1}{x}, \forall x \in \mathbb{R}^*$, on peut la déduire des formules de Leibniz et Faà di Bruno on constatant que $\frac{f}{g} = f \times (i \circ g)$.

3.3.5 Formules de Taylor

L'idée de la dérivée est de pouvoir approcher une fonction par l'équation d'une droite autour d'un point donné, autrement dit par un polynôme de degré 1. En réitérant cette idée, grâce au paragraphe précédent, on va pouvoir approcher une fonction par un polynôme de plus haut degré, à condition que la fonction soit suffisamment régulière.

Nous allons voir 3 formules de Taylor : la première est purement locale et sert à calculer des développements limités, alors que les deux autres donnent une estimation de l'erreur commise. Dans ce paragraphe, les fonctions sont définies sur un intervalle I.

---

88. Comme une fonction dérivable est continue, cela est équivalent à dire que f est n-fois dérivable sur D pour tout $n \in \mathbb{N}$.

89. Pour ces deux dernières opérations, il convient de mettre les hypothèses ad hoc, qui sont celles permettant de dire que la fonction quotient ou la fonction composée est bien définie.

90. Ne vous exemptez pas de l'effort d'écrire cette démonstration, qui doit être expéditive et vous fera réviser si besoin la relation sur les coefficients binomiaux qu'on appelle le "triangle de Pascal".

91. On l'ajoutera si la suite de l'écriture de ces notes montre que ça nous servira par la suite ; si vous lisez ces mots, c'est que ça n'est pas arrivé dans la version actuelle de ce document...

Proposition 3.51 (Formule de Taylor-Young)**.** *Soit $f : I \to \mathbb{R}$ de classe C^n avec $n \in \mathbb{N}^*$, et a à l'intérieur de I. Alors*

$$f(a+h) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} h^k + \underset{h \to 0}{o}(h^n)$$

Remarque 3.52. Comme pour la définition de la dérivée, on peut préférer tout écrire avec une variable qui tend vers 0 (ce qu'on a fait ici) ou avec une variable qui tend vers a (ce qu'on avait fait dans la définition de la dérivée). On invite le lecteur à s'entrainer à passer de l'une à l'autre des formules. Il n'est pas rare de voir un étudiant mis à mal parce qu'on a eu le malheur de lui demander de faire un développement limité autour d'un autre point que 0...

Pour mémoire, la formule peut également s'écrire :

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (x-a)^k + \underset{x \to a}{o}((x-a)^n)$$

Proposition 3.53 (Formule de Taylor-Lagrange)**.** *Soit $f : I \to \mathbb{R}$ de classe C^{n+1}* [92] *avec $n \in \mathbb{N}$, et a à l'intérieur de I. Alors pour tout $h \in \mathbb{R}$ tel que $a+h \in I$, il existe $\theta \in]0,1[$ tel que*

$$f(a+h) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} h^k + \frac{h^{n+1}}{(n+1)!} f^{(n+1)}(a+\theta h)$$

Remarque 3.54. Notez que si $n = 0$, on retrouve le théorème des accroissements finis [93]. En effet,

$$f(a+h) = f(a) + hf'(a+\theta h)$$

correspond bien à la formule

$$\frac{f(b)-f(a)}{b-a} = f'(c)$$

où $b = a+h$, et où l'existence d'un élément c dans l'intervalle $]a,b[$ (on a choisi $b > a$ (i.e. $h > 0$) pour fixer les idées, mais ça n'est pas nécessaire) correspond bien à l'existence d'un $\theta \in]0,1[$.

D'ailleurs, à l'image de la Remarque 3.52, on aurait pu écrire que si $f : [a,b] \to \mathbb{R}$ est de classe C^{n+1}, alors il existe $c \in]a,b[$ tel que

$$f(b) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (b-a)^k + \frac{f^{(n+1)}(c)}{(n+1)!} (b-a)^{n+1}. \tag{3.9}$$

Proposition 3.55 (Formule de Taylor avec reste intégral)**.** *Soit $f : I \to \mathbb{R}$ de classe C^{n+1} avec $n \in \mathbb{N}$, et a à l'intérieur de I. Alors pour tout $h \in \mathbb{R}$ tel que $a+h \in I$,*

$$f(a+h) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} h^k + \frac{h^{n+1}}{n!} \int_0^1 (1-t)^n f^{(n+1)}(a+th) dt.$$

Remarque 3.56. Comme en Remarque 3.54, on invite le lecteur à toujours tester le cas particulier $n = 0$ quand il écrit cette formule ; ce cas correspond d'ailleurs à l'initialisation de la récurrence proposée en démonstration.

Remarque 3.57. La formule peut également s'écrire :

$$f(b) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (b-a)^k + \int_a^b \frac{f^{(n+1)}(t)}{n!} (b-t)^n dt.$$

92. Comme pour le théorème de Rolle, on peut affaiblir l'hypothèse en f de classe C^n sur $[a,b]$ et $n+1$ fois dérivable sur $]a,b[$.

93. On invite le lecteur à constater ce cas particulier $n = 0$ à chaque fois qu'il écrira une telle formule sujette à de nombreuses erreurs d'indice. C'est un réflexe à avoir constamment : notre mémoire est faillible, et pour une formule complexe, il faut savoir la tester intelligemment sur des cas particuliers. Evidemment, ça ne démontrera pas que la formule est correcte, mais ça pourra vous permettre de détecter des erreurs grossières.

Démonstration. On laisse cette preuve, qui s'écrit simplement par récurrence, en exercice. L'initialisation correspond à ce qui est souvent appelé "Théorème fondamentale de l'analyse" (voir théorème 5.12), et l'hérédité repose sur une simple intégration par partie.[94] □

3.3.6 Exercices

Exercice 3.58. Calculer à partir de la définition la dérivée de $f(x) = x^n$ pour tout $x \in \mathbb{R}$.

Exercice 3.59. Comme à l'exercice 3.24, on invite le lecteur à produire un contre-exemple à cette proposition quand I n'est pas un intervalle.

Exercice 3.60 (Conditions d'optimalité d'ordre 1)**.** Dans la preuve du théorème de Rolle, on a prouvé que si c est un minimum de $f : [a, b] \to \mathbb{R}$ et que $c \in]a, b[$, alors c est un point critique de f, c'est-à-dire $f'(c) = 0$. On peut naturellement se demander ce qui persiste lorsque f réalise son minimum au bord de l'intervalle.

Si $f : [a, b] \to \mathbb{R}$ est une fonction dérivable sur $[a, b]$, et c qui réalise le minimum de f. Montrer que :

— si $c = a$, alors $f'(c) \geq 0$,

— si $c = b$, alors $f'(c) \leq 0$.

Faites une interprétation graphique.

Exercice 3.61. Soit $f : I \to \mathbb{R}$ dérivable sur un intervalle I. Montrer que

$$f \text{ est lipschitzienne sur } I \iff f' \text{ est bornée sur } I.$$

Exercice 3.62. En rapport à la Remarque 3.36, exhiber une fonction $f : \mathbb{R} \to \mathbb{R}$ qui est dérivable sur $\mathbb{R}$, possède un développement limité à l'ordre 2 en 0 (ou même à tout ordre), mais n'est pas deux fois dérivable en 0. Par contre, voir l'exercice 3.63 pour la réciproque.

Exercice 3.63 (Preuve de Taylor-Young)**.** On trouve des preuves dans tous les livres[95], mais je les trouve souvent un peu parachutées. Je propose une démonstration en plusieurs étapes[96], sans doute un peu plus longue, mais probablement un peu plus compréhensible.

1. On suppose $\forall k \in [\![1, n]\!],\ f^{(k)}(a) = 0$. Montrer que[97]

$$\lim_{t \to a} \frac{f(t) - f(a)}{(t - a)^n} = 0.$$

2. En considérant

$$x \mapsto f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x - a)^k, \tag{3.10}$$

montrer la formule de Taylor-Young.

Exercice 3.64 (Preuve de Taylor-Lagrange)**.** Comme ci-dessus, on propose la preuve en quelques étapes, qui créent un parallèle naturel avec le paragraphe 3.3.3. Dans tout l'exercice $f : [a, b] \to \mathbb{R}$ est de classe C^{n+1}.

94. Le lecteur qui n'est pas efficace pour rédiger une récurrence simple mais un peu technique est évidemment invité à rédiger cette preuve sans aller la lire quelque part !

95. Attention, certaines preuves (par exemple [Gou08b, Dan07]), un chouia plus élaborées, sont faites pour des fonctions à valeurs dans un espace plus général que ℝ ; on n'a alors plus le droit à l'égalité des accroissements finis, ce qui explique la différence de démarche de ces références. Il n'est évidemment pas interdit d'aller prioritairement travailler ces preuves, mais il faut certainement avoir conscience de leur raison d'être. On reviendra sur ces preuves au Chapitre VII.

96. Trouvée ici.

97. On pourra procéder par récurrence, appliquer l'hypothèse de récurrence à f', et l'égalité des accroissements finis à f pour conclure à l'étape d'hérédité.

1. On commence par supposer $f(a) = f(b) = 0$, et $\forall k \in [\![1, n]\!],\ f^{(k)}(a) = 0$. Montrer qu'il existe $c \in]a, b[$ tel que $f^{(n+1)}(c) = 0$.
2. On suppose maintenant $\forall k \in [\![1, n]\!],\ f^{(k)}(a) = 0$. En considérant une fonction de la forme $x \mapsto f(x) - \lambda(x-a)^{n+1}$ pour un $\lambda \in \mathbb{R}$ bien choisi, montrer qu'il existe $c \in]a, b[$ tel que
$$f(b) - f(a) = \frac{f^{(n+1)}(c)}{(n+1)!}(b-a)^{n+1}.$$
3. Finalement dans le cas général, en considérant (3.10), montrer (3.9)

3.4 Fonctions usuelles

Dans ce paragraphe, on construit et étudie quelques fonctions classiques.

3.4.1 Fonctions puissances

On cherche à construire les fonctions
$$f_\alpha : \begin{array}{ccc} D_\alpha & \longrightarrow & \mathbb{R} \\ x & \longmapsto & x^\alpha \end{array}$$
pour des $\alpha \in \mathbb{Q}$, et où on verra que D_α sera soit $]0, +\infty[$, soit $[0, +\infty[$, soit $\mathbb{R}$ suivant les cas. Les cas $\alpha \in \mathbb{R}$ sont plus complexes, et reposent sur la construction des fonctions exponentielles que l'on aborde au paragraphe suivant.

— Si $\alpha \in \mathbb{N}$, on peut définir x^α par répétition du produit de x par lui-même, et ceci est valable pour tout $x \in D_\alpha = \mathbb{R}$. On peut facilement montrer que ces fonctions sont continues, dérivables et même C^∞. En effet, avec la formule du binôme de Newton, on montre, si $\alpha \in \mathbb{N}^*$:
$$\forall x \in \mathbb{R},\ \ f'_\alpha(x) = \alpha x^{\alpha-1}.$$

— Si $\alpha = \frac{1}{q}$ où $q \in \mathbb{N}^*$ est pair, alors on applique le théorème 3.14 à la fonction
$$f_q : \begin{array}{ccc} \mathbb{R}_+ & \longrightarrow & \mathbb{R} \\ x & \longmapsto & x^q \end{array}$$
qui est strictement croissante et continue, et donc bijective. On en déduit que la fonction définie par[98]
$$f_{\frac{1}{q}} := (f_q)^{-1} : \begin{array}{ccc} \mathbb{R}_+ & \longrightarrow & \mathbb{R} \\ x & \longmapsto & x^{\frac{1}{q}} \end{array}$$
est bijective, continue et strictement croissante. Avec la proposition 3.39, on montre que $f_{\frac{1}{q}}$ est dérivable sur $\mathbb{R}_+^*$ et
$$\forall x \in]0, +\infty[,\ \ f'_{\frac{1}{q}}(x) = \frac{1}{q} x^{\frac{1}{q}-1}.$$
Elle a une tangente verticalle en 0.

— Si $\alpha = \frac{1}{q}$ où $q \in \mathbb{N}^*$ est impair, alors on applique le théorème 3.14 à la fonction
$$f_q : \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & x^q \end{array}$$

98. Ceci est une définition : la définition de $x^{\frac{1}{q}}$ est $(f_q)^{-1}(x)$.

qui est strictement croissante et continue, et donc bijective. On en déduit que la fonction définie par

$$f_{\frac{1}{q}} := (f_q)^{-1} : \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & x^{\frac{1}{q}} \end{array}$$

est bijective, continue et strictement croissante. Avec la proposition 3.39, on montre que $f_{\frac{1}{q}}$ est dérivable sur $\mathbb{R}^*$ et

$$\forall x \in \mathbb{R}^*, \;\; f'_{\frac{1}{q}}(x) = \frac{1}{q} x^{\frac{1}{q}-1}.$$

Elle a une tangente verticalle en 0.

— Si $\alpha = \frac{p}{q}$ avec $p \in \mathbb{N}$ et $q \in \mathbb{N}^*$ et $\operatorname{pgcd}(p,q) = 1$, on peut poser $D_\alpha = D_{\frac{1}{q}}$ [99] et $\forall x \in D_\alpha, x^{\frac{p}{q}} := (x^{\frac{1}{q}})^p$.

— Si enfin $\alpha = \frac{p}{q}$ avec $p \in \mathbb{Z} \setminus \mathbb{N}$, on pose $D_\alpha = D_{|\alpha|} \setminus \{0\}$ et $\forall x \in D_\alpha, x^{\frac{p}{q}} := \frac{1}{x^{-\frac{p}{q}}}$.

Il faudrait vérifier que les dernières notations sont légitimes (par exemple indépendantes de l'écriture de la fraction $\frac{p}{q}$ au moins si $x > 0$) et indépendantes de l'ordre dans lequel on fait les opérations, par exemple a-t-on $(x^p)^{\frac{1}{q}} = (x^{\frac{1}{q}})^p$? Il faut aussi vérifier :

$$\forall (x,y) \in]0,+\infty[^2, \;\; \forall (\alpha,\beta) \in \mathbb{Q}^2, \;\; x^{\alpha+\beta} = x^\alpha . x^\beta, \;\; (x^\alpha)^\beta = x^{\alpha\beta}, \;\; x^\alpha y^\alpha = (xy)^\alpha$$

On renvoit à [God98, IV.1] pour plus de détails.

3.4.2 Fonctions exponentielle et logarithme

Commençons par remarquer que ces fonctions ne sont pas si simples à construire, et qu'il existe plusieurs constructions possibles. Remarquons ensuite que si l'une de ces deux fonctions est construite, alors via le théorème 3.14, on peut en déduire assez facilement une construction de l'autre en tant que réciproque. On commence par s'intéresser à la fonction exponentielle en précisant deux constructions, et on conclura par une construction plus adaptée à la fonction logarithme.

***Première construction élémentaire de la fonction exponentielle :** On peut montrer le résultat suivant, principalement avec la théorie des suites et des fonctions réelles, ce qui s'intègre bien dans ce chapitre.

Théorème 3.65. *Il existe une unique fonction $f : \mathbb{R} \to \mathbb{R}$ dérivable sur $\mathbb{R}$ et telle que $f' = f$ et $f(0) = 1$.*

On donne cette construction en exercice ci-dessous : c'est en fait un des objets de l'épreuve écrite du Capes 2002. Voir aussi ce document de D. Perrin ou cet autre document de G. Eguether.

***Deuxième construction élémentaire de la fonction exponentielle :** La deuxième construction se concentre sur l'idée d'étendre de $\mathbb{Q}$ à $\mathbb{R}$ les fonctions puissances du paragraphe 3.4.1, et permet d'obtenir une caractérisation de la fonction exponentielle par une équation fonctionnelle. En effet, on a regardé les fonctions $x \mapsto x^\alpha$ pour $x \in ,+\infty[$ et $\alpha \in \mathbb{Q}$. On échange les rôles de x et α, en fixant $a \in]0,+\infty[$ et en regardant la fonction

$$\begin{array}{ccc} \mathbb{Q} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & a^x \end{array} \tag{3.11}$$

On peut montrer (sans avoir recours au théorème 3.65) :

99. C'est-à-dire $[0,+\infty[$ si q est pair, et $\mathbb{R}$ si q est impair.

Théorème 3.66. *Pour tout $a \in]0,+\infty[$, il existe une unique fonction $f : \mathbb{R} \to \mathbb{R}^*$ continue telle que*

$$\forall (x,y) \in \mathbb{R}^2, \quad f(x+y) = f(x).f(y) \tag{3.12}$$

et $f(1) = a$. On appelle f l'exponentielle de base a, qu'on note $\exp_a$.

On omet ici la preuve de ce théorème qui repose sur un prolongement par densité de la fonction (3.11) ; deux approches sont possibles (mais très proches), utiliser la complétude (à l'image de la preuve du théorème 1.87 au Chapitre III), ou la monotonie (à l'image du théorème 6.15). On renvoit à [God98, IV.2 et IV.6].

Ici, on a construit *les* fonctions exponentielles, sans donner de place prépondérante au nombre e. Pour conclure sur une construction de l'exponentielle, soit on connait le nombre e et on choisit simplement $a = e$, soit on montre (voir [God98, IV.7]) qu'il existe $L(a)$ tel que[100]

$$\forall x \in \mathbb{R}, \quad \exp_a'(x) = L(a) \exp_a'(x)$$

et alors e peut être défini comme l'unique réel strictement positif tel que $L(e) = \exp_e'(0) = 1$. Alors on note plus simplement $\exp = \exp_e$.

Au final, on a la formule

$$\forall x \in]0,+\infty[, \forall a \in \mathbb{R}, \quad x^a = \exp(a\ln(x)) = e^{a\ln(x)}$$

qui peut être vue comme une propriété ou une définition selon l'approche choisie.

Remarque 3.67. Comme on l'a dit en introduction de ce paragraphe, on peut déduire de cette construction l'existence de la fonction $\ln_a$ le logarithme en base a, comme inverse de la fonction $\exp_a$, à condition que $a \neq 1$ (en montrant que $\exp_a$ est continu et strictement monotone). De plus, si on choisit $a = e$, $\ln = \ln_e$ s'appelle le logarithme népérien.

Remarquons également qu'on aurait pu penser dans l'autre sens, et appliquer cette approche par densité en construisant d'abord les fonctions logarithmes, voir le paragraphe 5 de ce document.

Troisième construction de l'exponentielle : cette construction est plus classique et probablement plus commode une fois qu'on s'y connaît un peu en série de réels et de fonctions. En effet on peut poser

$$\exp(x) := \sum_{n\in\mathbb{N}} \frac{x^n}{n!}.$$

On détaillera cette approche au Chapitre II dans le cadre de l'analyse complexe.

Propriétés de la fonction $\exp$: On montre avec des études de fonctions que exp est strictement croissante, tend vers $+\infty$ en $+\infty$, vers 0 en $-\infty$, et même que

$$\lim_{x\to\infty} \frac{\exp(x)}{x^n} = +\infty$$

Une construction de la fonction logarithme : Une approche radicalement différente peut être appliquée ici, via la théorie de l'intégration de Riemann. En effet, on peut définir :

$$\forall x \in]0,+\infty[, \quad \ln(x) = \int_1^x \frac{1}{t} dt.$$

C'est aussi une approche qu'on utilisera au Chapitre II.

3.4.3 Fonctions trigonométriques et leurs réciproques

On suppose connues les fonctions sinus et cosinus. On les construira dans le Chapitre II grâce à l'exponentielle complexe qu'on construira comme une série entière. Néanmoins, comme on l'a proposé au paragraphe précédent pour l'exponentielle, il existe d'autres constructions de sin et cos : voir par exemple [AM04, Page 74].

100. $L(a)$ n'est autre que $\ln(a)$, mais on ne veut pas tourner en rond !

Fonction arcsin **:** On s'intéresse à la restriction de la fonction sin sur $I := [-\frac{\pi}{2}, \frac{\pi}{2}]$, qui est strictement croissante et d'image $[-1,1]$. On définit

$$\begin{array}{rccl} \arcsin := (\sin_{|I})^{-1} : & [-1,1] & \longrightarrow & \mathbb{R} \\ & y & \longmapsto & x \text{ l'unique élément de } [-\frac{\pi}{2}, \frac{\pi}{2}] \text{ tel que } \sin(x) = y \end{array}$$

Proposition 3.68. *La fonction* arcsin *est continue, strictement croissante, dérivable sur* $]-1,1[$ *et*

$$\forall x \in]-1,1[, \ \ \arcsin'(x) = \frac{1}{\sqrt{1-x^2}}.$$

De plus, elle a des tangentes verticales en -1 *et en* 1.

Démonstration. On applique le théorème 3.14 et la proposition 3.39, ce qui donne

$$\forall x \in]-1,1[, \ \ \arcsin'(x) = \frac{1}{\sin'(\arcsin(x))} = \frac{1}{\cos(\arcsin(x))} = \frac{1}{\sqrt{1-\sin^2(\arcsin(x)}} = \frac{1}{\sqrt{1-x^2}}.$$

où on a utilisé que $\cos > 0$ sur $]-\frac{\pi}{2}, \frac{\pi}{2}[$ pour pouvoir écrire $\cos = \sqrt{1-\sin^2}$. □

Fonction arccos **:** On s'intéresse à la restriction de la fonction cos sur $I := [0, \pi]$, qui est strictement décroissante et d'image $[-1,1]$. On définit

$$\begin{array}{rccl} \arccos : (\cos_{|I})^{-1} : & [-1,1] & \longrightarrow & \mathbb{R} \\ & y & \longmapsto & x \text{ l'unique élément de } [0,\pi] \text{ tel que } \cos(x) = y \end{array}$$

Proposition 3.69. *La fonction* arccos *est continue, strictement décroissante, dérivable sur* $]-1,1[$ *et*

$$\forall x \in]-1,1[, \ \ \arccos'(x) = -\frac{1}{\sqrt{1-x^2}}.$$

De plus, elle a des tangentes verticales en -1 *et en* 1.

La preuve est similaire à la précédente, on la laisse en exercice.

Fonction arctan **:** On s'intéresse à la restriction de la fonction tan sur $I :=]-\frac{\pi}{2}, \frac{\pi}{2}[$, qui est strictement croissante et d'image $\mathbb{R}$. On définit

$$\begin{array}{rccl} \arctan := (\tan_{|I})^{-1} : & \mathbb{R} & \longrightarrow & \mathbb{R} \\ & y & \longmapsto & x \text{ l'unique élément de }]-\frac{\pi}{2}, \frac{\pi}{2}[\text{ tel que } \tan(x) = y \end{array}$$

Proposition 3.70. *La fonction* arctan *est continue, strictement croissante, telle que*

$$\lim_{x\to+\infty} \arctan(x) = \frac{\pi}{2}, \quad \lim_{x\to-\infty} \arctan(x) = -\frac{\pi}{2}$$

et elle est dérivable sur $\mathbb{R}$ *avec*

$$\forall x \in \mathbb{R}, \ \ \arctan'(x) = \frac{1}{1+x^2}.$$

La preuve est une nouvelle fois laissée au lecteur.

3.4.4 Fonctions hyperboliques

Par analogie avec l'écriture complexe des fonctions sin et cos, on définit

$$\forall x \in \mathbb{R}, \quad \cosh(x) := \frac{e^x + e^{-x}}{2}, \quad \sinh(x) := \frac{e^x - e^{-x}}{2}, \quad \tanh(x) = \frac{\sinh(x)}{\cosh(x)}.$$

Le comportement de ces fonctions est radicalement différent de celui des fonctions trigonométrique. Néanmoins, on trouve un fort parallèle dans les formulaires (par exemple dans les dérivées).

3.4.5 Exercices

Exercice 3.71. 1. Montrez l'inégalité de Bernoulli

$$\forall y \geq -1, (1+y)^n \geq 1+ny.$$

2. Soit $x \in \mathbb{R}$. Montrer que les suites $\forall n \in \mathbb{N}^*, u_n(x) = (1+\frac{x}{n})^n$ et $v_n(x) := \frac{1}{u_n(-x)} = (1-\frac{x}{n})^{-n}$ sont adjacentes dès que $n > |x|$.

Ainsi on peut poser

$$\exp(x) := \lim_{n\to+\infty} \left(1+\frac{x}{n}\right)^n.$$

3. En remarquant que $\forall x \in \mathbb{R}, \forall n \geq 1, u_n'(x) = (1+\frac{x}{n})^{n-1}$, montrer [101]

$$\forall x \in \mathbb{R}, \forall h \in \mathbb{R}, \quad h\exp(x) \leq \exp(x+h) - \exp(x) \leq h\exp(x+h).$$

4. En déduire que exp est continue, dérivable, et $\exp' = \exp$. Ceci achève la preuve de l'existence.
5. On cherche maintenant à prouver l'unicité :
   (a) Montrer
   $$\forall x \in \mathbb{R}, \exp(x).\exp(-x) = 1. \tag{3.13}$$
   (b) En déduire que si $f : \mathbb{R} \to \mathbb{R}$ est une fonction dérivable telle que $f' = f$ et $f(0) = 1$, alors $f \equiv \exp$ [102].
6. À partir de là, on peut essayer de déduire l'équation fonctionnelle satisfaite par la fonction exp.
   (a) En reprenant l'idée de la preuve de l'unicité, montrer l'existence et l'unicité des solutions de $f' = \lambda f, f(0) = a$ où $(\lambda, a) \in \mathbb{R}^2$.
   (b) Montrer [103]
   $$\forall (x,y) \in \mathbb{R}^2, \exp(x+y) = \exp(x).\exp(y).$$

Exercice 3.72. 1. Montrer

$$\forall x \in [-1,1], \quad \arccos(x) + \arcsin(x) = \frac{\pi}{2}.$$

2. Montrer

$$\forall x \in \mathbb{R}^*, \quad \arctan(x) + \arctan\left(\frac{1}{x}\right) = \text{signe}(x).\frac{\pi}{2}.$$

Exercice 3.73. 1. Etudier et tracer les courbes représentatives des fonctions $\cosh, \sinh, \tanh$, et montrer

$$\cosh^2 - \sinh^2 = 1, \quad \cosh' = \sinh, \quad \sinh' = \cosh, \quad \tanh' = 1 - \tanh^2 = \frac{1}{\cosh^2}.$$

2. Montrer qu'on peut définir

$$\text{argch} : [1, +\infty[\to \mathbb{R}, \quad \text{argsh} : \mathbb{R} \to \mathbb{R}, \quad \text{argth} :]-1,1[\to \mathbb{R}$$

les bijections réciproques de $\cosh, \sinh, \tanh$ sur des intervalles bien choisis.

3. Montrer

$$\forall x \in [1,+\infty[, \quad \text{argch}(x) = \ln(x+\sqrt{x^2-1}), \quad \forall x \in]1,+\infty[, \text{argch}'(x) = \frac{1}{\sqrt{x^2-1}}$$
$$\forall x \in \mathbb{R}, \quad \text{argsh}(x) = \ln(x+\sqrt{x^2+1}), \quad \forall x \in \mathbb{R}, \text{argsh}'(x) = \frac{1}{\sqrt{1+x^2}}$$
$$\forall x \in]-1,1[, \quad \text{argth}(x) = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right), \quad \forall x \in]-1,1[, \text{argth}'(x) = \frac{1}{1-x^2}.$$

---

101. On pourra utiliser l'inégalité des accroissements finis.
102. On pourra dériver $g : x \mapsto f(x)\exp(-x)$
103. Après avoir fixé y, on pourra dériver la fonction $x \mapsto g(x) := \exp(x+y) - \exp(x)\exp(y)$

4 Suites et séries de fonctions

On a vu dans le paragraphe précédent l'étude de $f : D \to \mathbb{R}$ une fonction, et on a étudié son comportement vis-à-vis de sa variable $x \in D \subset \mathbb{R}$. Ici, on s'intéresse à voir f comme une variable, ce qui explique que ce paragraphe annonce les prémices de l'analyse fonctionnelle, qui sera développée plus en détails aux Chapitres III et IX.

4.1 Convergence simple d'une suite de fonctions

Voir mail de Lichère

4.2 Convergence uniforme d'une suite de fonctions

Etant donné D un sous-ensemble de $\mathbb{R}$, on considère ici $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions réelles définies sur D. Autrement dit, pour tout $n \in \mathbb{N}$, $f_n : D \to \mathbb{R}$ est une fonction. On peut bien sûr fixer n et étudier chaque fonction f_n avec les outils du paragraphe 3, mais on veut ici s'intéresser au comportement asymptotique de la suite de fonctions, c'est-à-dire par exemple, se demander si $(f_n)_{n\in\mathbb{N}}$ peut se rapprocher [104] d'une fonction à la limite $n \to +\infty$, comme on l'a fait au paragraphe 2 pour des suites de réels.

La première réponse, la plus naïve et naturelle, consiste à dire que la suite $(f_n)_{n\in\mathbb{N}}$ se rapproche d'une fonction f si c'est le cas pour chaque évaluation en $x \in D$. Précisément, on a la définition :

Définition 4.1 (Convergence simple)**.** Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions réelles définies sur $D \subset \mathbb{R}$, et $f : D \to \mathbb{R}$. On dit que $(f_n)_{n\in\mathbb{N}}$ converge simplement vers f si

$$\forall x \in D, \quad \lim_{n\to\infty} f_n(x) = f(x). \tag{4.1}$$

On notera éventuellement

$$f_n \xrightarrow[n\to+\infty]{} f \text{ simplement.}$$

Comment étudier la convergence simple d'une suite de fonction $(f_n)_{n\in\mathbb{N}}$? Rien de vraiment nouveau ici, on fixe $x \in D$, et on est ramené à l'étude de la suite réelle $(f_n(x))_{n\in\mathbb{N}}$; on étudie la convergence de cette suite, et quand celle-ci existe (dans $\mathbb{R}$), on note $f(x)$ sa limite. Si ce processus a fonctionné pour tout $x \in D$, on a alors trouvé le candidat limite f, et prouvé la convergence simple de (f_n) vers f [105].

Remarque 4.2 (**Topologie de la convergence simple ?)**.** Les étudiants déjà un peu versés dans la topologie, devraient se demander si cette convergence correspond à une topologie (ou une métrique) sur l'espace des fonctions de D dans $\mathbb{R}$. La réponse est : oui, cela correspond bien à une topologie, mais non, cette dernière n'est pas issue d'une métrique [106], en tout cas si D est non dénombrable. Pour cette raison, ces considérations ne sont pas au programme de l'agrégation [107].

Une fois cette convergence introduite, on va devoir faire face au fait que cette dernière est trop faible, au sens où on ne peut pas déduire grand chose sur la limite f à partir de ce qu'on sait sur $(f_n)_{n\in\mathbb{N}}$ [108]. On sera donc amené à manipuler d'autres types de convergence. Il en existe plein, ce qu'on abordera notamment dans le chapitre IX, mais la première convergence qu'il convient d'apprendre à manipuler et qui reste simple à étudier, est la convergence uniforme :

Définition 4.3 (Convergence uniforme)**.** Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions réelles définies sur $D \subset \mathbb{R}$, et $f : D \to \mathbb{R}$. On dit que $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers f si

$$\lim_{n\to\infty}\left(\sup_{x\in D} |f_n(x) - f(x)|\right) = 0. \tag{4.2}$$

104. En en sens à définir ; on verra qu'il en existe plusieurs.
105. Soit dit en passant, ces quelques mots élémentaires permettent de conclure à l'unicité de la limite simple d'une suite de fonctions.
106. Autrement dit, non métrisable.
107. Comme quoi, pas besoin d'aller chercher bien loin pour trouver des topologies non métrisables.
108. On peut quand même savoir des choses : à titre d'exemple, si les $(f_n)_{n\in\mathbb{N}}$ sont (toutes) positives, ou croissantes, alors ce sera aussi le cas pour f. On ne fait pas de liste et on n'énonce pas de proposition, ce sont des résultats assez évidents.

On notera éventuellement

$$\left[f_n \xrightarrow[n\to+\infty]{} f \text{ uniformément}\right], \text{ ou encore } f_n \xrightarrow[n\to+\infty]{\|\cdot\|_\infty} f$$

Remarque 4.4 (*Convergence issue d'une norme ?)**.** Pour $f : D \to \mathbb{R}$, on peut introduire

$$\|f\|_\infty := \sup_{x\in D} |f(x)|, \tag{4.3}$$

qu'on appelle la norme infinie [109]. Si cela s'avère nécessaire, on peut alourdir la notation en $\|f\|_{\infty,D}$ pour faire référence à l'espace sur lequel on prend le sup, qui peut s'avérer être un sous-ensemble de l'espace de départ des fonctions.

La convergence uniforme de $(f_n)_{n\in\mathbb{N}}$ vers f s'écrit alors

$$\|f_n - f\|_\infty \xrightarrow[n\to+\infty]{} 0. \tag{4.4}$$

Attention, dans ce cadre $\|\cdot\|_\infty$ n'est pas une norme, car tant qu'on ne sait pas que les fonctions considérées sont bornées, le nombre défini en (4.3) existe seulement dans $\mathbb{R}_+ \cup \{+\infty\}$. Néanmoins, on peut introduire

$$\forall (f,g) \in \mathbb{R}^D, \quad d(f,g) := \min\{1, \|f-g\|_\infty\}$$

et on peut vérifier qu'il s'agit d'une métrique sur l'ensemble des fonctions de D dans $\mathbb{R}$, et que la convergence pour cette métrique est la convergence uniforme.

Si par contre on se restreint aux fonctions bornées de D dans $\mathbb{R}$, alors (4.3) définit bien une norme, et la convergence uniforme correspond bien à la convergence associée à cette norme, voir le Chapitre III.

Remarque 4.5 (Comparaison convergence simple/uniforme)**.** Il convient de se familiariser avec ces deux modes de convergence avec des exemples. Mais il convient également de faire l'effort d'écrire la phrase complète de quantificateurs mis en jeu dans les définitions et d'en saisir la différence, comme on l'a fait pour la continuité uniforme à la Remarque 3.17. Une réécriture de (4.1) est :

$$\forall x \in D,\ \forall \varepsilon > 0,\ \exists n_0 \in \mathbb{N},\ \forall n \geq n_0,\ \ |f_n(x) - f(x)| \leq \varepsilon,$$

alors que la réécriture de (4.2) est [110] :

$$\forall \varepsilon > 0,\ \exists n_0 \in \mathbb{N},\ \forall n \geq n_0,\ \forall x \in D,\ \ |f_n(x) - f(x)| \leq \varepsilon.$$

La différence entre ces deux formules tient uniquement en la place de $\forall x \in D$; dans le premier cas, ce quantificateur arrive avant le quantificateur $\exists n_0$, ce qui signifie que n_0 a le droit de dépendre de x. Alors que dans la seconde version, n_0 est introduit avant x, et donc le même n_0 doit convenir pour tous les x choisis [111].

De ces remarques, on comprend facilement que la convergence uniforme implique la convergence simple :

$$\left[f_n \xrightarrow[n\to+\infty]{} f \text{ uniformément}\right] \implies \left[f_n \xrightarrow[n\to+\infty]{} f \text{ simplement}\right]$$

et suggère que la réciproque n'est sans doute pas valable. Bien sûr, il faut produire un contre-exemple pour s'en convaincre, ce que je vous invite à faire.

109. Mais attention à ce mot ! Voir le reste de la remarque...

110. En effet, étant donnés A un ensemble de réels et M un réel,

$$\sup A \leq M \iff \forall x \in A,\ x \leq M.$$

111. Par contre, à l'intérieur de chaque formule, on pourrait échanger les quantificateurs $\forall$ qui se côtoient. Par exemple, la convergence uniforme peut aussi s'écrire :

$$\forall \varepsilon > 0,\ \exists n_0 \in \mathbb{N},\ \forall x \in D,\ \forall n \geq n_0,\ \ |f_n(x) - f(x)| \leq \varepsilon.$$

4.3 Propriétés de la convergence uniforme

Comme dans le cas de $\mathbb{R}$, nous allons voir qu'une suite de fonctions qui est de Cauchy au sens uniforme, est uniformément convergente.

Définition 4.6. Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions définies sur D. On dit que $(f_n)_{n\in\mathbb{N}}$ est de Cauchy au sens uniforme si

$$\forall \varepsilon > 0,\ \ \exists n_0 \in \mathbb{N},\ \forall p \geq n_0,\ \forall q \geq n_0,\ \ \sup_{x\in D} |f_q(x) - f_p(x)| \leq \varepsilon.$$

Proposition 4.7 (Complétude de la convergence uniforme)**.** *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions définies sur D. Si $(f_n)_{n\in\mathbb{N}}$ est de Cauchy au sens uniforme, alors il existe $f : D \to \mathbb{R}$ telle que $(f_n)_{n\in\mathbb{N}}$ converge vers f uniformément.*

Remarque 4.8. La réciproque de cet énoncé est également valable, mais comme on le verra de façon détaillée au paragraphe 1.4 du Chapitre III, cette dernière est élémentaire, et valable pour toute métrique. Une suite convergente est toujours de Cauchy ; c'est la réciproque qui est non triviale, et qu'on appelle complétude.

Démonstration. Comme $(f_n)_{n\in\mathbb{N}}$ est de Cauchy, on peut écrire :

$$\forall \varepsilon > 0,\ \ \exists n_0 \in \mathbb{N}, \forall (p,q) \geq n_0,\ \ \forall x \in D,\ |f_q(x) - f_p(x)| \leq \varepsilon. \tag{4.5}$$

Si on considère donc $x \in D$ fixé, la suite $(f_n(x))_{n\in\mathbb{N}}$ est de Cauchy dans $\mathbb{R}$, donc convergente d'après la complétude de $\mathbb{R}$. Ainsi on peut noter $f(x)$ sa limite, et on a prouvé que $(f_n)_{n\in\mathbb{N}}$ converge simplement vers f.

Pour montrer la convergence uniforme, on reprend (4.5) et on fait tendre $q \to \infty$, ce qui donne

$$\forall \varepsilon > 0,\ \ \exists n_0 \in \mathbb{N}, \forall p \geq n_0,\ \ \forall x \in D,\ |f(x) - f_p(x)| \leq \varepsilon,$$

d'où la convergence uniforme de $(f_n)_{n\in\mathbb{N}}$ vers f. □

Remarque 4.9. Cette preuve est un grand classique de la complétude. Comme presque toujours pour une preuve de complétude, on se ramène à la complétude d'un autre espace dont on sait déjà qu'il est complet, ici l'espace d'arrivée $\mathbb{R}$. Ceci permet de construire la limite. Le reste de la preuve consiste à montrer que la convergence a bien lieu en le sens souhaité, ce qui se fait en essayant de passer à la limite dans la définition initiale d'une suite de Cauchy.

Il faut parfois une étape supplémentaire, qui consiste à montrer aussi que la limite construite est bien dans le bon espace. Ici, on ne demandait à la limite f que d'être une fonction de D dans $\mathbb{R}$, donc il n'y avait rien à ajouter.

On peut également chercher à savoir les propriétés qui seront préservées par la limite uniforme d'une suite de fonctions. La propriété la plus célèbre de la convergence uniforme est sans doute qu'elle préserve la continuité :

Théorème 4.10. *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions réelles définies sur D, $a \in D$. On suppose*

$$\forall n \in \mathbb{N},\ \ f_n \text{ est continue en } a$$

et

$$f_n \xrightarrow[n\to+\infty]{} f \text{ uniformément.}$$

Alors f est continue en a.

C'est l'exemple typique de démonstration qui doit devenir un automatisme : on a différentes informations (ici la convergence uniforme et la continuité de chaque $f_n, n \in \mathbb{N}$), dont il va falloir combiner les définitions pour obtenir la propriété souhaitée. Au final, la seule subtilité éventuelle de la preuve consiste à écrire les informations dans le bon ordre, même si je trouve qu'ici il n'y a pas vraiment d'ambiguité.

Démonstration. Soit $\varepsilon > 0$. Comme (f_n) converge uniformément vers f, il existe $n_0 \in \mathbb{N}$ tel que

$$\forall x \in D, \forall n \geq n_0, \quad |f_n(x) - f(x)| \leq \varepsilon.$$

En particulier [112]

$$\forall x \in D, |f_{n_0}(x) - f(x)| \leq \varepsilon$$

Or on sait que f_{n_0} est continue en a. Donc il existe $\alpha > 0$ tel que

$$\forall x \in]a-\alpha, a+\alpha[\cap D, \quad |f_{n_0}(x) - f_{n_0}(a)| \leq \varepsilon.$$

Au final, on a

$$\forall x \in]a-\alpha, a+\alpha[\cap D, \quad |f(x) - f(a)| \leq |f(x) - f_{n_0}(x)| + |f_{n_0}(x) - f_{n_0}(a)| + |f_{n_0}(a) - f(a)| \leq 3\varepsilon,$$

d'où le résultat. □

On donne désormais un résultat élémentaire sur le passage à la limite dans les intégrales.

Proposition 4.11 (Convergence uniforme et intégration)**.** *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions continues sur $[a,b]$, et qui converge uniformément vers $f : [a,b] \to \mathbb{R}$ (qui est donc aussi continue). Alors*

$$\int_a^b f_n(t)dt \xrightarrow[n\to+\infty]{} \int_a^b f(t)dt.$$

Remarque 4.12. Ce résultat est élémentaire, et peu puissant puisqu'il fait des hypothèses fortes (que la convergence soit uniforme, et que l'intervalle d'intégration soit un segment). On verra au Chapitre IV des théorèmes beaucoup plus performants ; il n'empêche qu'il peut être raisonnable de se rendre compte que vous vous trouvez dans le cadre d'application de ce résultat.

Démonstration. Les fonctions étant continues sur un segment, elles y sont bornées et intégrables. On peut écrire, grâce à la linéarité et à la croissance de l'intégrale,

$$\forall n \in \mathbb{N}, \quad \left|\int_a^b f_n(t)dt - \int_a^b f(t)dt\right| \leq \int_a^b |f_n(t) - f(t)|\, dt \leq \int_a^b \|f_n - f\|_\infty dt = (b-a)\|f_n - f\|_\infty$$

ce qui permet de conclure puisque le terme de droite dans l'inégalité converge vers 0. □

Afin de montrer qu'une limite de fonctions C^1 est dérivable, on a donc recours à l'énoncé suivant :

Proposition 4.13 (Convergence uniforme et dérivation)**.** *Soit I un intervalle, $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions C^1. On suppose :*

— *il existe $a \in I$ tel que la suite $(f_n(a))_{n\in\mathbb{N}}$ converge vers $\ell \in \mathbb{R}$,*

— *la suite des dérivées $(f'_n)_{n\in\mathbb{N}}$ converge uniformément vers une fonction $g : I \to \mathbb{R}$.*

Alors la suite $(f_n)_{n\in\mathbb{N}}$ converge simplement vers la fonction

$$\forall x \in I, \quad f(x) = \int_a^x g(t)dt + \ell, \tag{4.6}$$

qui est bien définie, de classe C^1, et telle que $f' = g$. De plus la convergence de $(f_n)_{n\in\mathbb{N}}$ vers f est uniforme sur tout segment inclus dans I.

Remarque 4.14. En pratique, on utilise souvent la version suivante, plus faible mais plus concise : si $(f_n)_{n\in\mathbb{N}}$ est une suite de fonctions C^1 sur I, qui converge uniformément vers f et telles que $(f'_n)_{n\in\mathbb{N}}$ converge uniformément vers g, alors f est de classe C^1 et $f' = g$.

112. Comme on avait fait un choix de réel en note 26, on choisit ici un entier $n \geq n_0$. Notez qu'on aurait pu choisir $n_0 + 42$ pour rester proche de *la* réponse. Plus sérieusement, si vous avez rédigé cette preuve sans choisir de valeur de n, et donc en disant qu'on utilise la continuité de f_n pour tout $n \geq n_0$, alors il y a quelque chose que vous n'avez pas compris, et vous devez retravailler la note 26.

Démonstration. Soit $x \in I$. Il y a convergence uniforme de $(f_n')_{n\in\mathbb{N}}$ vers g sur $[a,x]$ (ou $[x,a]$) donc d'après le théorème 4.10 g est continue (donc on peut l'intégrer) et d'après la proposition 4.11,

$$\int_a^x f_n'(t)dt \xrightarrow[n\to+\infty]{} \int_a^x g(t)dt$$

Mais comme

$$\forall n \in \mathbb{N}, \quad f_n(x) = \int_a^x f_n'(t)dt + f_n(a)$$

on déduit bien de la convergence de $(f_n(a))_{n\in\mathbb{N}}$ vers ℓ la convergence simple de $(f_n)_{n\in\mathbb{N}}$ vers la fonction f définie par (4.6), qui est bien C^1 puisque g est continue.

Supposons désormais que $J = [c,d] \subset I$. Alors[113]

$$\forall x \in J, \ \ |f_n(x)-f(x)| \leq |f_n(a)-f(a)|+\left|\int_a^x |f_n'(t) - g(t)|dt\right| \leq |f_n(a)-f(a)|+\max\{|d-a|,|c-a|\}\|f_n'-g\|_\infty$$

d'où la convergence uniforme de $(f_n)_{n\in\mathbb{N}}$ vers f sur J. □

4.4 Séries de fonctions

Pour les séries de fonctions, on peut considérer pas moins de 4 formes de convergences ; deux qui sont une simple traduction des convergences pour les suites de fonctions, et deux qui sont spécifiques aux séries :

Définition 4.15. Soit $(f_n)_{n\in\mathbb{N}}$ définies sur $D \subset \mathbb{R}$. On peut considérer la suite $(F_n)_{n\in\mathbb{N}}$ des sommes partielles

$$\forall n \in \mathbb{N}, \ \ F_n = \sum_{k=0}^{n} f_k$$

pour désigner la série de fonctions de terme général $(f_n)_{n\in\mathbb{N}}$.

- On dit que la série de fonctions de terme général $(f_n)_{n\in\mathbb{N}}$ converge simplement vers $S : D \to \mathbb{R}$ si pour tout $x \in D$, $F_n(x)$ converge vers $S(x)$.
- On dit que la série de fonctions de terme général $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers $S : D \to \mathbb{R}$ si
$$\sup_{x\in D} |F_n(x) - S(x)| \xrightarrow[n\to+\infty]{} 0.$$
- On dit que la série de fonctions de terme général $(f_n)_{n\in\mathbb{N}}$ converge absolument si pour tout $x \in D$, la série de terme général $(f_n(x))_{n\in\mathbb{N}}$ est absolument convergente, c'est-à-dire si
$$\forall x \in D, \ \ \sum_{n\in\mathbb{N}} |f_n(x)| < +\infty$$
- On dit que la série de fonctions de terme général $(f_n)_{n\in\mathbb{N}}$ converge normalement si la série de terme général $(\sup_{x\in D} |f_n(x)|)_{n\in\mathbb{N}}$ est convergente, c'est-à-dire si
$$\sum_{n\in\mathbb{N}} \sup_{x\in\mathbb{N}} |f_n(x)| < +\infty$$

Remarque 4.16. Attention à l'écriture $\sum_{n\in\mathbb{N}} f_n$, qui peut prêter à confusion sur la nature de la convergence. Vous devez avoir au moins prouvé la convergence simple de la série de fonctions pour avoir le droit d'écrire cette expression (comme on le voit ci-dessous, la convergence simple est impliquée par toutes les autres convergences).

113. Attention, si vous manipulez des bornes d'intégrales non ordonnées,

$$\text{il faut bien écrire } \left|\int_a^b f(t)dt\right| \leq \left|\int_a^b |f(t)|dt\right| \qquad \text{et non} \qquad \left|\int_a^b f(t)dt\right| \leq \int_a^b |f(t)|dt$$

car si $b < a$ cette seconde formule est fausse puisque son terme de gauche est positif alors que le terme de droite est négatif (strictement négatif sauf si f est nulle).

Avec la proposition 2.39, on voit que la convergence absolue d'une série de fonctions implique sa convergence simple, c'est-à-dire que

$$\left[\forall x \in D,\ \sum_{n\in\mathbb{N}} |f_n(x)| < +\infty\ \right] \implies \left[\exists S : D \to \mathbb{R},\ \ \lim_{n\to+\infty}\left(\sum_{k=0}^{n} f_k(x)\right) = S(x)\right].$$

Du fait de l'inégalité $\forall x \in D, |f_n(x)| \leq \sup_{y\in D}|f_n(y)|$, on voit également facilement que la convergence normale implique la convergence absolue de la série.

Enfin, on a vu en Remarque 4.5 que la convergence uniforme implique la convergence simple. On va maintenant voir un résultat similaire entre la convergence normale et la convergence uniforme [114] :

Proposition 4.17. *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions. Si la série de terme général $(f_n)_{n\in\mathbb{N}}$ converge normalement, alors elle converge uniformément, c'est-à-dire qu'il existe $S : D \to \mathbb{R}$ telle que*

$$\sup_{x\in D}\left|\sum_{k=0}^{n} f_k(x) - S(x)\right| \xrightarrow[n\to+\infty]{} 0.$$

Démonstration. Comme pour la proposition 2.39, on propose deux preuves, dont la seconde rentre dans le cadre général des espaces complets.

Preuve 1 : Pour tout $x \in D$, $|f_n(x)| \leq \sup_{y\in D}|f_n(y)|$ donc la série de terme général $(f_n(x))_{n\in\mathbb{N}}$ est absolument convergente, donc convergente par la proposition 2.39. On peut ainsi définir

$$\forall x \in D,\ \ S(x) = \sum_{n\in\mathbb{N}} f_n(x)$$

et il reste à voir que la convergence est uniforme. Pour cela, on constate que

$$\forall n \in \mathbb{N}, \forall x \in D,\ \ |F_n(x) - S(x)| \leq \sum_{k=n+1}^{+\infty} |f_k(x)| \leq \sum_{k=n+1}^{+\infty} \sup_{y\in D}|f_k(y)|.$$

Intéressons-nous au terme de droite : d'une part il est indépendant de x donc on peut passer au sup en x, et d'autre part il s'agit du reste de la série de terme général $(\sup_{y\in D}|f_n(y)|)_{n\in\mathbb{N}}$, qui est convergente par hypothèse, donc il tend vers 0, ce qui conclut la preuve.

Preuve 2 : Pour montrer que $(F_n)_{n\in\mathbb{N}}$ est uniformément convergente, on sait d'après la proposition 4.7 qu'il suffit de montrer qu'elle est de Cauchy au sens uniforme. Soit donc $\varepsilon > 0$. Étant convergente, la suite $\left(\sum_{k=0}^{n}\sup_{y\in D}|f_k(y)|\right)_{n\in\mathbb{N}}$ est également de Cauchy, donc il existe $n_0 \in \mathbb{N}$ tel que

$$\forall q \geq p \geq n_0,\ \ \left|\sum_{k=p+1}^{q}\sup_{y\in D}|f_k(y)|\right| \leq \varepsilon.$$

On en déduit ainsi que

$$\forall q \geq p \geq n_0, \forall x \in D,\ \ |F_q(x) - F_p(x)| \leq \sum_{k=p+1}^{q}|f_k(x)| \leq \sum_{k=p+1}^{q}\sup_{y\in D}|f_k(y)| \leq \varepsilon$$

donc on a montré que la suite $(F_n)_{n\in\mathbb{N}}$ était de Cauchy au sens uniforme, elle est donc bien convergente. □

Remarque 4.18. Les deux preuves paraissent différentes, mais elles reposent sur les mêmes outils. En effet la première preuve n'utilise pas le critère de Cauchy uniforme, mais elle utilise via la proposition 2.39 la complétude de $\mathbb{R}$, ou une propriété équivalente. La seconde utilise le critère de Cauchy uniforme,

114. C'est le seul résultat non trivial de toute l'affaire. Les étudiants apprennent souvent par cœur une sorte de diagramme liant toutes les convergences précédemment citées. Pourquoi pas, mais il me semble plus judicieux de savoir "reconstruire" un tel diagramme en sachant redémontrer très rapidement toutes les implications faciles, et retenir lesquelles ne sont pas simples (et elles, les retenir par cœur).

qui a été prouvé également grâce à la complétude de $\mathbb{R}$ [115]. D'ailleurs, on peut considérer que la première preuve est une preuve alternative de la complétude de l'espace de fonctions de D dans $\mathbb{R}$ pour la métrique définie en Remarque 4.4 ; en effet, on verra (Proposition 2.18 au Chapitre III) qu'il s'agit d'un fait général que la propriété que la convergence de la série des normes implique la convergence de la série est une condition qui caractérise la complétude [116].

On peut désormais s'entraîner à adapter les énoncés du paragraphe 4.3 au cadre des séries. On donne deux exemples, le lecteur fera le reste :

On trouvera également un critère d'Abel ou un critère des séries alternées pour la convergence uniforme ici, Page 32. À mon avis il n'est pas nécessaire d'apprendre ces choses (en tout cas pas par cœur). Je laisse cela à l'appréciation de chacun. Par contre, il est clair que la démarche de chercher et prouver ces énoncés constitue un bon test de la compréhension précédemment évoquée et sera très bénéfique pour le lecteur qui s'y attèle.

4.5 Exercices

Exercice 4.19 (Opérations)**.** Encore une fois, on préfère ne pas donner d'énoncé pour ne pas encombrer la mémoire, mais on pourrait se demander comment la convergence uniforme se comporte avec les opérations algébriques. Le lecteur se convaincra que la somme ou la multiplication par un scalaire ne pose pas de problème. Pour le produit par contre, il faut faire un peu attention.

Montrer que si $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers f et $(g_n)_{n\in\mathbb{N}}$ converge uniformément vers g, *et que* f et g sont bornées sur D, alors $(f_n g_n)_{n\in\mathbb{N}}$ converge uniformément vers fg.

Bien sûr, on fournira un contre-exemple dans le cas où on ne suppose plus f et g bornées.

Exercice 4.20. Comme d'habitude, il faut se convaincre de la pertinence d'un tel résultat en exhibant un contre-exemple où la convergence n'est pas uniforme, mais seulement simple.

Aussi, si vous n'avez pas encore beaucoup d'aisance avec ce type de démonstration, il est à mon avis très instructif d'essayer de démontrer le résultat (faux !) qu'une limite simple de fonctions continues en a est continue en a. Il est en effet intéressant de saisir où le raisonnement "buggue" en comparaison de celui ci-dessus. Ces expérimentations doivent vous permettre de construire une intuition sur l'importance d'avoir une propriété "uniforme", et même vous permettre d'anticiper quand cela sera nécessaire d'y avoir recours. Car c'est bien là que réside la subtilité de cette notion.

Les propriétés suivantes sont aussi très classiques, et leurs preuves, proches de la précédente, sont laissées en exercice :

Exercice 4.21 (Interversions des limites)**.** Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions définies sur D qui converge uniformément vers $f : D \to \mathbb{R}$.

1. Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de D qui converge vers $x \in D$, et on suppose f continue en x. Montrer que
$$f_n(x_n) \xrightarrow[n\to+\infty]{} f(x).$$
2. Soit $a \in \overline{D}$. On suppose que
$$\forall n \in \mathbb{N}, \quad \lim_{x\to a} f_n(x) \text{ existe et vaut } \ell_n \in \mathbb{R}.$$
   (a) Montrer que $(\ell_n)_{n\in\mathbb{N}}$ est une suite de Cauchy.

115. Ce type de considération doit permettre aux étudiants de prendre du recul et de se conforter dans la véracité (ou le caractère erroné) de leurs démonstrations. Si vous produisez une preuve différente de ce que vous avez pu lire ici ou là, ça peut être une très bonne chose, à condition bien sûr que la-dite preuve soit correcte ; si vous utilisez "beaucoup moins" d'outils que dans la preuve originale, c'est un signal d'alarme à prendre en compte.

116. Attention, en toute rigueur, on ne l'abordera que dans le cadre des espaces vectoriel normé, alors qu'on a déjà vu que $\mathbb{R}^D$ n'était pas normé. Comme quoi les cadres abstraits ne sont pas toujours à propos pour traiter les cas particuliers.

(b) En déduire que

$$\lim_{x\to a} f(x) \quad \text{et} \quad \lim_{n\to+\infty} \ell_n \quad \text{existent et sont égales.}$$

Ainsi on a montré, dans le cas de la convergence uniforme :

$$\lim_{n\to+\infty}\lim_{x\to a} f_n(x) = \lim_{x\to a}\lim_{n\to+\infty} f_n(x).$$

3. Exhibez des contre-exemples aux questions précédentes quand la convergence est seulement simple.

Exercice 4.22. Toujours la même rengaine ! Trouvez des contre-exemples quand la convergence n'est pas uniforme, ou quand elle l'est mais qu'on essaie d'intégrer sur un ensemble non borné.

Pour conclure, on revient à la question de la régularité, et on cherche à voir comment la dérivabilité peut passer à la limite. Commençons par un contre-exemple :

Exercice 4.23. Montrer que la suite de fonctions $\forall n \in \mathbb{N}, \forall x \in [-1,1], f_n(x) = \sqrt{x^2 + \frac{1}{n}}$ converge uniformément vers la fonction valeur absolue.

Ainsi les fonctions $(f_n)_{n\in\mathbb{N}}$ sont C^1 (et même C^∞) sur $[-1,1]$ mais la limite n'est pas dérivable en 0. On verra via le théorème de Weierstrass que cet exemple n'a rien d'exceptionnel, puisqu'on peut approcher uniformément toute fonction continue par des fonctions polynomiales (et donc C^∞).

Exercice 4.24 (*Version dérivable au lieu de C^1 de la proposition 4.13)**.** [117] La preuve précédente a été particulièrement simple car on a eu recours à l'intégration des fonctions $(f_n')_{n\in\mathbb{N}}$ (et de la limite g), ce qui était possible car on a supposé les fonctions $(f_n)_{n\in\mathbb{N}}$ de classe C^1. Si on les supposait seulement dérivables, on ne pourrait pas le faire. Pourtant on peut avoir un énoncé encore valable : Soit I un intervalle de $\mathbb{R}$, $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions dérivables sur I et telles que

— il existe $a \in I$ tel que la suite $(f_n(a))_{n\in\mathbb{N}}$ converge vers $\ell \in \mathbb{R}$,

— la suite des dérivées $(f_n')_{n\in\mathbb{N}}$ converge uniformément vers une fonction $g : I \to \mathbb{R}$.

1. Montrez que si $x \in I$ alors la suite $(f_n(x))_{n\in\mathbb{N}}$ est de Cauchy. Par complétude de $\mathbb{R}$, on en déduit que $(f_n(x))_{n\in\mathbb{N}}$ converge vers un réel qu'on note $f(x)$.
2. Montrez que f est dérivable sur I et que $f' = g$.

Solution Pages 29-30 sur ce lien

5 Intégration de Riemann

On construit dans ce paragraphe l'intégrale de Riemann ; pour certains aspects calculatoires (on a déjà utilisé certains dans les paragraphes précédents), cette intégrale convient tout à fait. Au final, les fonctions que l'on manipule explicitement sont souvent suffisamment régulières. Mais on verra d'une

117. Personnellement je ne suis pas friand de ces généralisations ; à vrai dire, je ne suis pas sûr d'avoir vu un jour une situation où cette généralisation s'applique mais pas la version précédente. Je ne doute pas que ça existe, mais j'imagine que ça reste plutôt rare. Néanmoins, je pense qu'il est bon d'avoir connaissance de ces variantes, au moins afin de mieux comprendre la littérature (qui ne donne souvent qu'une version) ; aussi, l'exercice est un bon entrainement, et il est à mon avis important d'avoir l'habitude que les situations C^1 sont souvent plus simples que les situations "dérivables" ; notamment, c'est un bon réflexe de penser à l'intégration quand les dérivées sont continues, et il peut être utile de voir comment on peut s'en sortir sans cela. Ces subtilités reviendront dans le Chapitre VII, en particulier dans l'étude de la convexité.

part que les théorèmes de convergence sont souvent peu performants [118], mais surtout que dès qu'on aborde des aspects plus fins liés à l'analyse fonctionnelle, et dont les applications les plus naturelles dans le programme de l'agrégation seront l'analyse de Fourier et la théorie des probabilités, on verra que cette théorie amène des limitations non négligeables. On développera donc au Chapitre IV la théorie de l'intégration de Lebesgue, qui est plus délicate, mais bien plus commode sur ces deux aspects. On renvoit à la note 125 pour d'autres commentaires instructifs sur cette question Riemann/Lebesgue.

5.1 Intégration sur un segment

Dans ce paragraphe, on se donne $a < b$ deux réels, et on considère des fonctions $f : [a, b] \to \mathbb{R}$.

5.1.1 Construction de l'intégrale des fonctions continues par morceaux

Commençons par définir les fonctions que nous allons pouvoir intégrer :

Définition 5.1. — On dit qu'une fonction $f : [a, b] \to \mathbb{R}$ est en escalier [119] s'il existe une subdivision $x_0 = a < x_1 < x_2 < \ldots < x_N = b$ finie ($N \in \mathbb{N}^*$) de $[a, b]$ telle que f est constante sur chaque intervalle $]x_i, x_{i+1}[$ où $i \in [\![0, N-1]\!]$. On note $Esc([a, b])$ l'ensemble des fonctions en escalier.

— On dit qu'une fonction $f : [a, b] \to \mathbb{R}$ est continue par morceaux s'il existe une subdivision $x_0 = a < x_1 < x_2 < \ldots < x_N = b$ finie ($N \in \mathbb{N}^*$) de $[a, b]$ telle que pour tout $i \in [\![0, N-1]\!]$,

$$f_{|]x_i, x_{i+1}[} \text{ est continue et prolongeable par continuité sur } [x_i, x_{i+1}]$$ [120].

On note $C^0_m([a, b])$ l'ensemble des fonctions continues par morceaux.

On peut commencer par donner la définition de l'intégrale d'une fonction en escalier dont l'interprétation géométrique est élémentaire (la somme d'aires (signées) de rectangles) :

Définition 5.2 (Intégrale d'une fonction en escalier)**.** Si $f : [a, b] \to \mathbb{R}$ est en escalier, alors il existe une subdivision $x_0 = a < x_1 < x_2 < \ldots < x_N = b$ de $[a, b]$ et $(\alpha_i)_{i \in [\![0, N-1]\!]}$ telle que pour tout $i \in [\![0, N-1]\!]$, $f = \alpha_i$ sur $]x_i, x_{i+1}[$. On pose alors

$$\int_a^b f(t)dt = \sum_{i=0}^{N-1} \alpha_i (x_{i+1} - x_i).$$

Remarque 5.3. Attention, il faut vérifier (ça n'est pas dur, mais il faut le faire !) que cette définition ne dépend pas de la subdivision choisie, qui n'est pas unique (une autre approche consiste à ajouter des conditions pour la rendre unique, par exemple en demandant aux intervalles $]x_i, x_{i+1}[$ d'être maximaux au sens où f n'est pas constante sur un intervalle strictement plus grand).

---

118. La situation dans la littérature est d'ailleurs plutôt confuse : en effet, depuis quelques années, le théorème de convergence dominée (issu de la théorie de Lebesgue) et ses conséquences (notamment sur les intégrales à paramètre) sont apparus notamment dans les programmes de CPGE, sous des formes adaptées (comme on ne parle pas de tribu ni de mesurabilité, on suppose les fonctions continues par morceaux, afin d'utiliser l'intégrale de Riemann ; ce qui ne coûte pas grand chose à ce niveau là). Leurs démonstrations sont admises, l'idée étant surtout que les étudiants apprennent à les utiliser, et ne soient pas obligés à des circonvulations parfois farfelues dues aux limitations de l'intégrale de Riemann. C'est une idée défendable et attaquable, je ne me prononcerai pas là-dessus. Pour un étudiant de l'agrégation, il est effectivement indispensable de savoir utiliser le théorème de convergence dominée et ses conséquences (disons au moins dans le cadre des fonctions continues par morceaux, comme cela est fait en CPGE), mais il faut quand même avoir conscience du cadre qui se trouve derrière. Il peut, à mon avis, être de bon ton d'utiliser des énoncés de la théorie de Riemann quand c'est possible et quand ceux-ci restent simples (voir par exemple la Remarque 4.12), tout en ayant le réflexe d'avoir recours au théorème de convergence dominée dans la majorité des cas. En tout cas, on donnera ici seulement les résultats naturels dans la théorie de Riemann, et on renvoit au Chapitre IV pour tout ce qui concerne le théorème de convergence dominée. Quant à certains aspects du programme, ils peuvent devenir assez scabreux quand on décide de se cantonner à la théorie de Riemann. Bref, tout ça pour mettre en garde l'étudiant agrégatif dans sa façon d'utiliser des livres de CPGE : à titre d'exemple, les différences entre [Gou98] et [Gou08b] sont instructives. Voir aussi les commentaires de [God03, page 18].

119. Attention à ne pas mélanger avec les fonctions étagées du Chapitre IV...

120. C'est-à-dire que f a des limites réelles en x_i à droite et en x_{i+1} à gauche.

Nous souhaitons maintenant construire l'intégrale des fonctions continues par morceaux : on va procéder par approximation par des fonctions en escalier. Commençons donc par le lemme préliminaire suivant :

Lemme 5.4. *Si $f \in C^0_m([a,b])$ alors il existe une suite $(f_n)_{n\in\mathbb{N}}$ de fonctions en escaliers telles que*

$$f_n \xrightarrow[n\to+\infty]{} f \quad \textit{uniformément.}$$

Démonstration. Supposons déjà que f est continue sur $[a,b]$. Soit $n \in \mathbb{N}^*$ et considérons la subdivision uniforme $x_i = a + i.\frac{b-a}{n}$ pour $i \in [\![0, n-1]\!]$ [121]. On définit f_n la fonction en escalier qui vaut $f(x_i)$ sur chaque intervalle $[x_i, x_{i+1}[$, pour chaque $i \in [\![0, n-1]\!]$. Ainsi définie, f_n n'est définie que sur $\cup_{i=0}^{n-1}[x_i, x_{i+1}[= [a,b[$; on peut donc compléter cette définition en posant $f_n(b) = f(b)$. [122] Nous allons voir que cette suite de fonctions converge uniformément vers f. Soit donc $\varepsilon > 0$. Par le théorème de Heine 3.18, f est uniformément continue sur $[a,b]$ donc il existe $\alpha > 0$ telle que

$$\forall (x,y) \in [a,b]^2, \ \ |x-y| \leq \alpha \Rightarrow |f(y) - f(x)| \leq \varepsilon.$$

On pose $n_0 = E\left(\frac{b-a}{\alpha}\right) + 1 \in \mathbb{N}^*$ [123]. Alors pour tout $n \geq n_0$ et tout $x \in [a,b[$ (le cas $x = b$ ne posant pas de problème puisque $f_n(b) = f(b)$), il existe un unique $i \in [\![0, n-1]\!]$ tel que $x \in [x_i, x_{i+1}[$, et alors $|x - x_i| \leq \frac{b-a}{n} \leq \alpha$, d'où

$$|f(x) - f_n(x)| = |f(x) - f(x_i)| \leq \varepsilon,$$

d'où la convergence uniforme de $(f_n)_{n\in\mathbb{N}}$ vers f.

Pour le cas général où f est continue par morceaux, on peut considérer une subdivision $(a_j)_{j\in[\![0,N-1]\!]}$ telle que f est prolongeable par continuité sur chaque $[a_j, a_{j+1}]$, ce qui implique que f est uniformément continue sur $]a_j, a_{j+1}[$, et on peut alors reproduire l'argument ci-dessus en considérant des subdivisions uniformes de chacun de ces intervalles, et en faisant attention (comme on l'a fait ci-dessus en b) de définir $f_n(a_j) = f(a_j)$ pour gérer les possibles discontinuités. On laisse les détails au lecteur [124]. □

On peut désormais définir l'intégrale de Riemann d'une fonction continue par morceaux :

Définition-Proposition 5.5. Etant donnée une fonction $f \in C^0_m([a,b])$, il existe une suite $(f_n)_{n\in\mathbb{N}}$ de fonctions en escaliers qui converge uniformément vers f : la suite $\left(\int_a^b f_n(t)dt\right)_{n\in\mathbb{N}}$ est convergente, et on définit

$$\int_a^b f(t)dt := \lim_{n\to+\infty} \int_a^b f_n(t)dt.$$

Cette définition est cohérente car si g_n est une autre suite de fonctions en escaliers qui converge uniformément vers f, alors

$$\lim_{n\to+\infty} \int_a^b f_n(t)dt = \lim_{n\to+\infty} \int_a^b g_n(t)dt.$$

Démonstration. Pour montrer que la suite $\left(\int_a^b f_n(t)dt\right)_{n\in\mathbb{N}}$ est convergente, on montre qu'elle est de Cauchy en constatant que

$$\forall (p,q) \in \mathbb{N}, \quad \left|\int_a^b f_q(t)dt - \int_a^b f_p(t)dt\right| \leq (b-a)\|f_q - f_p\|_\infty$$

121. Attention, ne perdez pas de vue que cette subdivision dépend de n ; on aurait pu noter $(x_i^{(n)})_{i\in[\![0,n-1]\!]}$.

122. Pour ceux qui aiment les indicatrices, la définition de f_n s'écrit :

$$f_n = \sum_{i=0}^{n-1} f(x_i)\mathbb{1}_{[x_i,x_{i+1}[} + f(b)\mathbb{1}_{\{b\}}$$

123. Constatez que ce choix de n_0 ne dépend pas de x ; c'est ce qui donne une convergence uniforme.

124. Quant au lecteur assidu qui a fait l'exercice 5.23, il pourra conclure plus rapidement en écrivant $f = g + h$ où g est continue et h en escalier ; par le cas des fonctions continues, il existe $(g_n)_{n\in\mathbb{N}}$ suite de fonctions en escalier qui converge uniformément vers g, et la suite $(g_n + h)_{n\in\mathbb{N}}$ fournit bien une suite de fonctions en escaliers qui converge vers f uniformément.

et en se rappelant que la suite $(f_n)_{n\in\mathbb{N}}$ étant uniformément convergente, elle est de Cauchy au sens uniforme. Par complétude de $\mathbb{R}$, on a donc bien la convergence de $\left(\int_a^b f_n(t)dt\right)_{n\in\mathbb{N}}$.

Pour l'assertion finale, on constate juste que

$$\forall n \in \mathbb{N}, \quad \left|\int_a^b g_n(t)dt - \int_a^b f_n(t)dt\right| \leq (b-a)\|g_n - f_n\|_\infty \leq (b-a)\big[\|f_n - f\|_\infty + \|g_n - f\|_\infty\big].$$

□

Concluons en énonçant les propriétés élémentaires de l'intégrale ainsi contruite.

Proposition 5.6. *L'intégrale de Riemann*

— *est linéaire : si $(f,g) \in C^0_m([a,b])$ et si $\alpha \in \mathbb{R}$, alors*

$$\int_a^b (\alpha f + g)(t)dt = \alpha \int_a^b f(t)dt + \int_a^b g(t)dt$$

— *est croissante : si $f \in C^0_m([a,b])$ est telle que $f \geq 0$, alors $\int_a^b f(t)dt \geq 0$.*

— *satisfait la relation de Chasles : si $f \in C^0_m([a,b])$ et si $c \in]a,b[$, alors*

$$\int_a^b f(t)dt = \int_a^c f(t)dt + \int_c^b f(t)dt$$

On ne détaille pas la démonstration, qui consiste à vérifier ces propriétés d'abord pour les fonctions en escaliers, puis à voir qu'elles passent à la limite.

5.1.2 *Généralisations

Dans ce paragraphe, on va détailler deux classes de fonctions plus générales que les fonctions continues par morceaux, et pour lesquelles on pourra définir une intégrale. La conviction de l'auteur, au vu du programme de l'agrégation externe, est que l'intérêt de ces considérations ne réside pas vraiment dans ces deux classes en elles-mêmes, mais plutôt dans le lien de la construction de l'intégrale avec le théorème de prolongement des applications linéaires continues qu'on verra au Chapitre III et qu'on a en fait redémontré sans trop le dire au paragraphe précédent.

Il est en effet à mon avis peu intéressant d'essayer d'obtenir et de comprendre en détail la classe la plus vaste de fonctions qu'on peut intégrer au sens de Riemann, pour la bonne raison que de toute façon, les applications à l'analyse fonctionnelle qui seront abordées dans ce document et qui sont à l'image du programme de l'agrégation, nécessitent la théorie de Lebesgue. Dit autrement, à chaque fois qu'on aura recours à l'intégrale de Riemann, on intégrera des fonctions qui seront au moins continues par morceaux ; et dès qu'on aura besoin de considérer l'intégrale de fonctions très peu régulières a priori, on utilisera le cadre de la théorie de Lebesgue, qui est le "bon" d'un point de vue manipulation et analyse fonctionnelle. [125]

125. Il me semble qu'une partie non négligeable des candidats à l'agrégation ont une phobie pas toujours rationnelle de la théorie de l'intégration de Lebesgue. Je peux comprendre les origines d'une telle phobie, la construction de l'intégrale de Lebesgue requiert effectivement des circonvolutions nettement plus effrayantes que la construction de l'intégrale de Riemann qui prend tout au plus une ou deux pages et quelques dessins assez parlants. Mais je vois souvent des étudiants qui en arrivent à la contradiction suivante : ils font une impasse assez drastique sur tout ce qui a recours à l'intégrale de Lebesgue, au point de devoir construire des leçons bancales sur tel ou tel aspect du programme (par exemple l'analyse de Fourier), mais ils cherchent à compenser en ayant une connaissance un peu démesurée et à mon avis inutile des contours les plus fins de l'intégrale de Riemann. Il m'est d'avis que ces étudiants feraient mieux de laisser de côté les notions de fonctions réglées ou Riemann-intégrables (qui d'ailleurs ne sont pas au programme), et de consacrer le temps ainsi gagné à reconsidérer leur impasse, à laisser de côté les aspects les plus délicats de la théorie de l'intégration de Lebesgue (notamment la construction de la mesure de Lebesgue n'est pas au programme, mais au-delà de ça, je pense qu'il n'est pas attendu une fine maîtrise des subtilités de la théorie de la mesure, de ce qu'est un ensemble borélien, etc.) et à apprendre à manipuler les théorèmes simples et efficaces de convergence qui en découlent, à apprivoiser son langage dans le cadre de la théorie des probabilités, et si le temps le permet, à prendre un peu de recul sur son importance du point de vue de l'analyse fonctionnelle.

Fonctions réglées : Si on relit la construction du paragraphe précédent à la lumière du Chapitre III, on a en effet montré au Lemme 5.4 que l'espace $Esc([a,b],\mathbb{R})$ est dense dans $(C^0_m([a,b]),\|\cdot\|_\infty)$, et on en a déduit à la Définition-Proposition 5.5 qu'on pouvait prolonger l'application

$$\begin{array}{rccc} I: & (Esc([a,b]),\|\cdot\|_\infty) & \longrightarrow & \mathbb{R} \\ & f & \longmapsto & \int_a^b f(t)dt \end{array} \tag{5.1}$$

à $C^0_m([a,b])$. Remarquons qu'on a utilisé dans la démonstration la complétude de $\mathbb{R}$, et même si on ne l'a pas dit ainsi, que l'application I était linéaire continue. On aurait donc pu appliquer directement le corollaire 2.51 qu'on verra au Chapitre IX.

On peut alors se demander naturellement quel est l'espace naturel sous-jacent à cette construction, et qui se trouve être strictement plus gros que $C^0_m([a,b])$:

Définition 5.7 (Fonctions réglées)**.** On dit qu'une fonction $f:[a,b]\to\mathbb{R}$ est réglée si elle est limite uniforme de fonctions en escaliers sur $[a,b]$. On notera $Reg([a,b])$ l'espace des fonctions réglées sur $[a,b]$.

On a ainsi la suite d'inclusion

$$Esc([a,b])\subset C^0_m([a,b])\subset Reg([a,b])\subset B([a,b]),\quad \text{et}\quad \overline{Esc([a,b])}=Reg([a,b])$$

où $B([a,b])$ désigne l'espace des fonctions bornées sur $[a,b]$, et où l'adhérence est prise pour la norme uniforme.

On peut donc prouver

Proposition 5.8. *L'application I définie en (5.1) se prolonge de manière unique en une application linéaire continue sur $(Reg([a,b]),\|\cdot\|_\infty)$.*

Démonstration. On applique donc le corollaire 2.51 après avoir constaté que

$$\forall f\in Esc([a,b]),\ \ |I(f)|\le (b-a)\|f\|_\infty,$$

et invoqué la complétude de $\mathbb{R}$ comme espace d'arrivée de I. □

Remarque 5.9. Citons un avantage non négligeable de la construction précédente : il est facile de la généraliser au cas où les fonctions f sont à valeurs dans un espace de Banach E. Toute la construction est similaire, mais on ne pourra plus parler de croissance de l'intégrale. On pourra par contre montrer l'inégalité

$$\left\|\int_a^b f(t)dt\right\|_E\le\int_a^b\|f(t)\|_E dt.$$

Ces généralisations sont beaucoup plus délicates dans le cadre de la mesure de Lebesgue.

****Fonctions Riemann intégrables :** juste précédemment, on a tiré parti du fait qu'on avait montré qu'une fonction continue par morceaux était limite uniforme de fonctions en escaliers. Mais la convergence uniforme n'était qu'un moyen suffisant et non nécessaire pour pouvoir définir l'intégrale des fonctions considérées.

Voici une définition plus générale de fonction intégrable au sens de Riemann :

Définition 5.10. Etant donné une subdivision $x_0=a<x_1<\ldots<x_N=b$, on note $\sigma=\{x_i\}_{i\in[\![0,N]\!]}$ et on considère pour $f:[a,b]\to\mathbb{R}$ une fonction bornée[126] :

$$I^-_\sigma(f)=\sum_{i=0}^{N-1}(x_{i+1}-x_i)\left[\inf_{[x_i,x_{i+1}]}f\right],\qquad\text{et}\qquad I^+_\sigma(f)=\sum_{i=0}^{N-1}(x_{i+1}-x_i)\left[\sup_{[x_i,x_{i+1}]}f\right].$$ [127]

126. On a besoin que la fonction soit bornée pour éviter d'écrire $+\infty-\infty$
127. Ces sommes sont parfois appelées sommes de Darboux.

On dit que f est Riemann intégrable sur $[a,b]$ si

$$\sup_\sigma I_\sigma^-(f) = \inf_\sigma I_\sigma^+(f) \tag{5.2}$$

où les supremum et infimum sont pris sur l'ensemble des subdivisions de $[a,b]$. Dans ce cas, on note $\int_a^b f(t)dt$ la valeur commune donnée par (5.2).

Remarque 5.11. Dans [Mar09, Chapitre 10.III] il est expliqué comment on peut voir cette construction en utilisant également le corollaire 2.51.

5.1.3 Propriétés de l'intégrale

"Théorème fondamental de l'analyse" : ce résultat établit que les opérations de dérivation et d'intégration sont réciproques l'une de l'autre. Cela donne un outil très performant pour calculer des intégrales.

Théorème 5.12. *On suppose que f est continue sur $[a,b]$, et on pose*

$$\forall x \in [a,b], \quad F(x) = \int_a^x f(t)dt. \tag{5.3}$$

Alors F est de classe C^1 sur $[a,b]$ et $F' = f$.

Remarque 5.13. Comme en témoignera la démonstration, on peut localiser l'énoncé précédent : si f est continue par morceaux ou même Riemann-intégrable, alors F définie par (5.3) est dérivable en tout point x en lequel f est continue, et alors $F'(x) = f(x)$.

Démonstration. Soit $\varepsilon > 0$, et $x \in [a,b]$. Par continuité de f en x, il existe $\alpha > 0$ tel que

$$\forall t \in [x-\alpha, x+\alpha] \cap [a,b], \quad |f(t) - f(x)| \leq \varepsilon.$$

Alors

$$\forall h \in [-\alpha,\alpha] \text{ tel que } x+h \in [a,b], \quad \begin{aligned} \left|\frac{F(x+h)-F(x)}{h} - f(x)\right| &= \left|\frac{1}{h}\int_x^{x+h} \Big[f(t)-f(x)\Big]dt\right| \\ &\leq \left|\frac{1}{h}\int_x^{x+h} \Big|f(t)-f(x)\Big|dt\right| \leq \varepsilon \end{aligned}$$

d'où la dérivabilité de F en x et le fait que $F'(x) = f(x)$. Le caractère C^1 de F vient de la continuité de sa dérivée f. □

On invite le lecteur à prouver les deux corollaires suivants :

Corollaire 5.14. *Soit f de classe C^1 sur $[a,b]$. Alors*

$$\forall (x,y) \in [a,b]^2, \quad \int_x^y f'(t)dt = f(y) - f(x).$$

Corollaire 5.15. *Si $f : [a,b] \to \mathbb{R}$ est continue et $F : [a,b] \to \mathbb{R}$ est dérivable et telle que $F' = f$, alors*

$$\forall (x,y) \in [a,b]^2, \quad \int_x^y f(t)dt = F(y) - F(x).$$

Sommes de Riemann : Si on revient à la définition-proposition 5.5 et au lemme 5.4, on se rend compte qu'on a défini, au moins pour une fonction continue :

$$\int_a^b f(t)dt = \lim_{n\to+\infty} \frac{b-a}{n}\sum_{k=0}^{n-1} f\left(a+k\frac{b-a}{n}\right).$$

Autrement dit, on a sommé les aires des rectangles "à gauche" de largeur $\frac{1}{n}$ et on a montré que cela avait une limite quand $n \to +\infty$.

Le résultat suivant donne une généralisation, avec plus de latitude sur les largeurs et hauteurs des rectangles considérés :

Proposition 5.16. *Soit $f \in C^0_m([a,b])$. Pour tout $n \in \mathbb{N}$, on suppose que σ_n est une subdivision de $[a,b]$ de pas δ_n, c'est-à-dire*[128]

$$\sigma_n : x_0 = a < x_1 < x_2 < \ldots < x_p = b, \quad \textit{où} \quad p \in \mathbb{N}, \quad \textit{et} \quad \delta_n := \max_{i\in[\![0,p-1]\!]} |x_{i+1}-x_i|,$$

et on se donne pour tout $i \in [\![0,p-1]\!], \theta_i \in [x_i, x_{i+1}]$. On pose alors

$$S_n = \sum_{i=0}^{p-1}(x_{i+1}-x_i)f(\theta_i).$$

$$\textit{Si} \quad \delta_n \xrightarrow[n\to+\infty]{} 0 \quad \textit{alors} \quad S_n \xrightarrow[n\to+\infty]{} \int_a^b f(t)dt.$$

Démonstration. On détaille rapidement la preuve (voir aussi [Gou08b, page 124] ou [Dan07, page 196]), assez instructive vis-à-vis de la construction de l'intégrale de Riemann : notamment on utilisera une démarche très similaire à plusieurs reprises pour l'intégrale de Lebesgue, en remplaçant l'indicatrice d'un segment par l'indicatrice d'un ensemble mesurable. Notons qu'il convient de voir S_n comme une "fonction dont la variable est f", pour $n \in \mathbb{N}$, on note donc $S_n(f) = \sum_{i=0}^{p-1}(x_{i+1}-x_i)f(\theta_i)$.

Étape 1 : Supposons que $f = \alpha \mathbb{1}_{[c,d]}$. Alors pour $n \in \mathbb{N}$ fixé, en utilisant les notations de l'énoncé, on a

$$\left|S_n(f) - \int_a^b f(t)dt\right| = \left|\sum_{k=0}^{p-1}\int_{x_i}^{x_{i+1}} \big[f(t)-f(\theta_i)\big]dt\right| \le \sum_{k=0}^{p-1}\int_{x_i}^{x_{i+1}} \Big|f(t)-f(\theta_i)\Big|dt$$

Pour $i \in [\![0,n-1]\!]$, si $[x_i,x_{i+1}]$ ne contient pas c ou d, alors f est constante égale à $f(\theta_i)$ sur $[x_i,x_{i+1}]$. Donc au plus deux termes de la somme de droite sont non nuls, et ces termes sont alors chacun majorés par $|\alpha|\delta_n$, d'où le résultat.

Étape 2 : Si maintenant f est une fonction en escalier, alors elle est combinaison linéaire de fonction de la forme $\alpha\mathbb{1}_{[c,d]}$[129]. Par linéarité de l'intégrale et de S_n vis-à-vis de la variable f (pour tout $n \in \mathbb{N}$), le résultat reste donc valable pour les fonctions en escalier.

Étape 3 : Enfin, soit $f \in C^0_m([a,b])$. Soit $\varepsilon > 0$. D'après le lemme 5.4, il existe g en escalier telle que $\|f-g\|_\infty \le \frac{\varepsilon}{3(b-a)}$. D'après l'étape 2, il existe $n_0 \in \mathbb{N}$ tel que $|S_n(g) - \int_a^b g(t)dt| \le \frac{\varepsilon}{3}$. On peut alors écrire

$$\begin{aligned}\left|S_n(f) - \int_a^b f(t)dt\right| &\le |S_n(f)-S_n(g)| + \left|S_n(g) - \int_a^b g(t)dt\right| + \left|\int_a^b g(t)dt - \int_a^b f(t)dt\right| \\ &\le \sum_{i=0}^{p-1}(x_{i+1}-x_i)|f(\theta_i)-g(\theta_i)| + \frac{\varepsilon}{3} + (b-a)\frac{\varepsilon}{3(b-a)} \le \varepsilon.\end{aligned}$$

□

128. Attention, les quantitiés p, x_i, θ_i dépendent de n.

129. Attention à cet énoncé, qui ne fait pas beaucoup attention à la nature des intervalles considérés ; l'énoncé n'est pas pour autant incorrect, mais il oblige à écrire

$$\mathbb{1}_{]c,d[} = \mathbb{1}_{[c,d]} - \mathbb{1}_{\{c\}} - \mathbb{1}_{\{d\}},$$

la fonction de gauche étant effectivement en escalier d'après la définition. De façon plus commode, on pourrait aussi dire que l'étape 1 est valable pour des intervalles ouverts ou fermés à gauche et à droite.

Remarque 5.17. *Le raisonnement précédent est très important à connaître ; c'est un raisonnement typique en analyse, on l'appelle raisonnement par densité. On a déjà évoqué en effet au paragraphe 5.1.2 que le lemme 5.4 établissait la densité des fonctions en escaliers dans l'espace des fonctions continues par morceaux muni de la norme uniforme. On a en fait d'abord prouvé que la propriété 5.16 était vraie pour les fonctions en escalier, pour ensuite montrer que ce fait s'étendait par approximation aux fonctions continues par morceaux dans l'étape 3.

Comparons au résultat classique : si F et G sont deux fonctions définies et continues sur $\mathbb{R}$ et si $F = G$ sur $\mathbb{Q}$ alors $F = G$. Ce résultat simple vaut dans un contexte beaucoup plus général, où $\mathbb{Q}$ joue le rôle d'un ensemble dense dans l'espace de départ $\mathbb{R}$. Dans la preuve précédente, ce sont $f \mapsto \lim_{n\to+\infty} S_n(f)$ et $f \mapsto \int_a^b f(t)dt$ qui jouent le rôle de F et G. On a montré que ces fonctions sont égales sur l'espace des fonctions en escalier (qui joue le rôle de $\mathbb{Q}$ dans l'exemple précédent). Et au milieu de la preuve, sans le dire ainsi (on aurait pu), on a montré que ces applications étaient continues pour la norme uniforme. Il me semble important d'avoir conscience de cela, sans chercher à se ramener forcément à un énoncé du type $F = G$ sur un ensemble dense et F, G sont continues, donc $F = G$. On reproduit bien souvent le raisonnement à la main comme on l'a fait ci-dessus. J'y vois deux raisons : des fois on se perd dans l'abstraction avec des notations encombrantes. Des fois, il y a un jeu entre plusieurs topologies (F et G peuvent être a priori continues pour des normes différentes, et l'ensemble sur lequel on sait l'égalité peut être dense pour une autre norme encore), ce qui fait qu'une preuve redétaillée devient souvent plus lisible.

On est dans le monde de l'analyse fonctionnelle, car la fonction est devenue la variable de travail.

Intégrales à paramètre : On appelle intégrale à paramètre une fonction définie par

$$\forall x \in I, \quad \varphi(x) := \int_a^b f(t,x)dt \tag{5.4}$$

où $f : [a,b] \times I \to \mathbb{R}$ est telle que pour tout $x \in I$, $t \mapsto f(t,x)$ est continue par morceaux. Une question naturelle est de savoir si la régularité de f se communique à celle de φ.

On verra au Chapitre IV que le cadre de l'intégrale de Lebesgue permet, notamment via le théorème de converge dominée, de fournir un cadre beaucoup plus flexible pour l'étude de la régularité des fonctions ainsi définies. Néanmoins, comme cela peut être commode dans les applications, on cite deux résultats qui peuvent être obtenus de manière élémentaire. Précisons quand même que lorsque le domaine d'intégration n'est pas un segment (voir le paragraphe 5.2), de tels théorèmes ne sont plus valables, et on renvoit au Chapitre IV pour de telles situations.

5.1.4 Méthodes de calcul explicite d'intégrale

Comme on l'a dit en introduction, on néglige ce paragraphe. On se contente de rappeler que la formule d'intégration par partie est une conséquence simple de $(uv)' = u'v + uv'$ et du Corollaire 5.14, et est a priori donné pour des fonctions C^1. On donne quand même le théorème de changement de variable :

Proposition 5.18 (Changement de variable)**.** *Soient I un intervalle, $\phi : [a,b] \to I$ de classe C^1, et $f : I \to \mathbb{R}$ continue. Alors*

$$\int_a^b f(\phi(t))\phi'(t)dt = \int_{\phi(a)}^{\phi(b)} f(x)dx. \tag{5.5}$$

La preuve vient simplement de l'égalité $f' \circ \phi.\phi' = (f \circ \phi)'$.

Remarque 5.19. On a écrit cet énoncé classique surtout pour saisir l'occasion de quelques remarques : il n'est pas demandé à ϕ d'être bijective (ou même injective), ou un C^1-difféomorphisme entre $[a,b]$ et son image, contrairement à ce qu'on entend (ou lit) souvent. La confusion a plusieurs origines : d'une part le cas général des intégrales multiples, qui requiert des hypothèses plus fortes du type C^1-difféomorphisme (voir le Chapitre IV), en second lieu le fait qu'on utilise souvent la formule de droite à gauche, auquel cas il peut paraître plus naturel de l'écrire :

$$\int_\alpha^\beta f(x)dx = \int_{\phi^{-1}(\alpha)}^{\phi^{-1}(\beta)} f(\phi(t))\phi'(t)dt$$

et de supposer ϕ bijective (même si au fond, on peut remplacer $\phi^{-1}(\alpha), \phi^{-1}(\beta)$ par deux antécédents de α et β, et ça marchera sans supposer ϕ bijective ; il n'empêche qu'en pratique, on inverse souvent l'écriture $x = \phi(t)$).

Enfin, on peut lire (voir par exemple [Dan07, Corollaire 13.16 et Remarque 13.2]) des tentatives d'affaiblissement des hypothèses de régularité, par exemple si f est seulement continue par morceaux, car alors il n'est pas clair sans hypothèse sur ϕ que $f \circ \phi$ sera continue par morceaux, et donc suivant le cadre dans lequel on se place, il se peut qu'on n'ait pas le droit d'écrire l'intégrale de $f \circ \phi$ (on peut peut-être y voir un avantage aux cadres plus généraux évoqués au Paragraphe 5.1.2, mais honnêtement, je doute que les énoncés qu'on discute ici ait un grand intérêt).

5.1.5 Intégration numérique

On aborde brièvement la question du calcul approché d'une intégrale. Rappelons qu'on est bien en train de traiter le programme de tronc commun, qui spécifie deux méthodes, à savoir la méthode des rectangles (on devrait dire les méthodes des rectangles), et la méthode de Monte-Carlo. Le programme de l'option B va un peu plus loin.

Notez que ce sont des sujets souvent mal maîtrisés (ou même connus) des candidats, alors qu'ils mettent bien en valeurs les différents aspects du programme, à la fois d'un point de vue de l'algèbre linéaire, et du point de vue de l'analyse fonctionnelle, avec des calculs de norme d'application linéaire etc. Ils peuvent également illustrer le théorème des accroissements finis et ses généralisations via les formules de Taylor.

On renvoit à [Dem16, Chapitre III], à ce polycopié (chapitre IV) pour beaucoup plus d'informations sur le sujet.

Méthode des rectangles : Étant donnée une fonction $f \in C^0_m([a,b])$, on pose $x_k = a + k\frac{b-a}{n}$ pour $k \in [\![0,n]\!]$, et

$$R^g_n(f) := \frac{b-a}{n}\sum_{k=0}^{n-1} f(x_k), \quad R^d_n(f) := \frac{b-a}{n}\sum_{k=1}^{n} f(x_k),$$

qu'on appelle respectivement méthodes des rectangles à gauche, à droite, et méthode du point milieu. On a déjà vu au paragraphe sur les sommes de Riemann, que ces formules convergent vers $\int_a^b f(x)dx$. Mais on s'intéresse ici à évaluer la vitesse de convergence, car d'un point de vue pratique, on aimerait avoir une idée de la valeur de n pour laquelle on peut considérer que le calcul approché est pertinent. C'est ce qu'on appelle une estimation d'erreur, qu'on propose à l'exercice suivant :

Remarque 5.20. À très faible coût, on peut construire deux méthodes d'ordre 2 :

$$R^m_n(f) := \frac{b-a}{n}\sum_{k=1}^{n} f\left(\frac{x_k + x_{k+1}}{2}\right), \quad T_n(f) := \frac{b-a}{n}\sum_{k=1}^{n} \frac{f(x_i) + f(x_{i+1})}{2}$$

respectivement appelées méthode du point milieu et méthode des trapèzes (pour des raisons claires sur un dessin. On peut montrer que si f est de classe C^2, alors

$$\forall n \in \mathbb{N}^*, \quad \left|\int_a^b f(x)dx - R^m_n(f)\right| \leq \frac{\|f''\|_\infty (b-a)^3}{24n^2}, \qquad \left|\int_a^b f(x)dx - T_n(f)\right| \leq \frac{\|f''\|_\infty (b-a)^3}{12n^2},$$

voir par exemple [Dan07, 24.4].

Méthode de Monte-Carlo : cette méthode est probabiliste, on n'en détaille pas l'analyse ici, on se contente de l'expliquer brièvement, et de faire quelques remarques culturelles. Disons qu'on souhaite calculer

$$I = \int_0^1 g(x)dx.$$

On va voir I comme l'espérance d'une variable alétoire. Supposons que U est une variable aléatoire de loi uniforme sur $[0,1]$ [130]. Alors par le théorème de transfert,

$$I = \mathbb{E}(g(U)).$$

130. On renvoit au Chapitre IV pour une définition de ces outils.

Si $(U_k)_{k\in\mathbb{N}}$ sont des variables aléatoires réelles indépendantes et identiquement distribuées (v.a.r.i.i.d.) de même loi que U, la loi des grands nombres affirme que

$$\frac{1}{n}\sum_{k=1}^{n} g(U_i) \xrightarrow[n\to+\infty]{} I, \quad \text{presque sûrement.}$$

Dit autrement, si $u_1, u_2, \ldots, u_n$ sont tirés au hasard dans $[0,1]$ [131], alors $\frac{1}{n}\sum_{k=1}^{n} g(u_i)$ est une approximation de I.

Remarquons qu'on peut généraliser les idées précédentes pour calculer

$$I = \int_{\mathbb{R}} g(x) f_X(x)dx, \qquad \text{où } f_X \text{ est la densité de probabilité de la variable } X$$

(donc en particulier f_X est positive et d'intégrale 1 sur $\mathbb{R}$). Si on tire $(X_i)_{i\in\mathbb{N}^*}$ i.i.d. de même loi que X, alors

$$\frac{1}{n}\sum_{k=1}^{n} g(X_i) \xrightarrow[n\to+\infty]{} \mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x) f_X(x)dx, \quad \text{presque sûrement.}$$

Comme pour les méthodes des rectangles, on se demande la vitesse de convergence. Contrairement à précédemment, on ne va pas pouvoir donner une estimation d'erreur "déterministe", car il y aura toujours une probabilité non nulle que les réalisations de (X_i) tirées au hasard ne soient pas bonnes. Mais le théorème central limite permet d'évaluer cette probabilité. En effet, si $\mathbb{E}(g(X)^2) < +\infty$, alors

$$\sqrt{n}\left[\sum_{k=1}^{n} g(X_i) - I\right] \xrightarrow[n\to+\infty]{loi} \mathcal{N}(0,\sigma^2)$$

où la convergence est en loi, et $\sigma^2 = \mathrm{Var}(g(X))$. Cela signifie que la convergence est de l'ordre de $\frac{1}{\sqrt{n}}$, mais pour être plus précis, il faut parler d'intervalle de confiance, et on renvoit à un cours de probabilité. Remarquons aussi qu'en pratique, on ne connaît pas σ^2, on l'approche donc lui même par la méthode de Monte-Carlo.

Listons pour conclure quelques avantages et inconvénients de cette méthode : d'abord, la vitesse de convergence n'est pas très bonne, en comparaison des méthodes déterministes. Par contre, il y a deux avantages majeurs : d'abord, on n'a pas fait d'hypothèse de régularité de la fonction à intégrer, contrairement à ce qu'on avait fait pour les méthodes déterministes. Ensuite, et ça là sans doute que réside le réel succès de la méthode de Monte-Carlo, la dépendance en la dimension de l'espace de départ (qu'on a choisi 1 dans la présentation, mais qui pourrait être adaptée à des dimension plus grandes) se comporte beaucoup mieux pour la méthode de Monte-Carlo que pour les méthodes déterministes.

5.2 Intégrales généralisées

Dans ce paragraphe, on aborde la définition d'une intégrale sur un intervalle non borné (par exemple $[a,+\infty[$) ou qui ne contient pas ses extrêmités (par exemple $]a,b]$). En effet, l'approche de Riemann a utilisé à plusieurs reprises le caractère compact de l'intervalle de travail ; et à juste titre, car si dans le paragraphe précédent, on n'a demandé qu'un peu de régularité aux fonctions pour pouvoir les intégrer sur des segments, nous devrons ici systématiquement nous inquiéter de l'existence des intégrales (on parlera d'intégrabilité de la fonction considérée). À ce titre, nous verrons comme dans le cas des séries, des méthodes pour déterminer cette intégrabilité.

Définition 5.21. — Soit $a \in \mathbb{R}$, $b \in \mathbb{R}\cup\{+\infty\}$ et $f : [a,b[$ une fonction continue par morceaux. On dit que l'intégrale $\int_a^b f(t)dt$ converge si la fonction $A \in [a,b[\ \mapsto \ \int_a^A f(t)dt$ admet une limite réelle quand $A \to b$. Dans ce cas on définit

$$\int_a^b f(t)dt := \lim_{A\to b}\int_a^A f(t)dt.$$

On parle d'intégrale impropre ou d'intégrale généralisée. Dans le cas contraire, on dira que l'intégrale est divergente.

131. Ce qui pose la question naturelle de savoir comment simuler un tel tirage, ce qui semble être dans le programme de l'option A.

— De la même manière et en échangeant les rôles de a et b, on peut définir la convergence d'une intégrale sur $]a, b]$ avec $a \in \mathbb{R} \cup \{-\infty\}$ et $b \in \mathbb{R}$.

— Si maintenant $a \in \mathbb{R} \cup \{-\infty\}$ et $b \in \mathbb{R} \cup \{+\infty\}$ et $f :]a, b[\to \mathbb{R}$ est continue par morceaux, on dit que l'intervalle $\int_a^b f(t)dt$ converge s'il existe $c \in]a, b[$ tel que les deux intégrales $\int_a^c f(t)dt$ et $\int_c^b f(t)dt$ sont convergentes, et alors on définit :

$$\int_a^b f(t)dt := \int_a^c f(t)dt + \int_c^b f(t)dt,$$

ce qui d'après la relation de Chasles, ne dépend pas du choix de c.

Si par contre l'une ou l'autre des intégrales diverge, on dit que l'intégrale est divergente.

Comme ceci n'a rien de bien passionnant ni difficile, nous passons vite sur les propriétés des intégrales impropres, qui sont linéaires, satisfont la relation de Chasles, etc. Tout se passe bien, le seul danger étant de bien être sûr que toutes les intégrales qu'on manipule sont bien définies. Comme pour les limites, il convient de ne pas écrire $\int_a^b f(t)dt$ sans avoir établi au préalable que cette intégrale était convergente. À l'image de la mise en garde (2.4), on n'écrira pas, par exemple :

$$\int_0^{+\infty} (t - t)dt = \int_0^{+\infty} tdt - \int_0^{+\infty} tdt.$$

Aussi, de nombreux critères du paragraphe sur les séries se transposent ici dans le cadre des intégrales impropres, notamment les théorèmes de comparaison pour les intégrales de fonctions positives, ainsi que le caractère convergent des intégrales absolument convergentes. On ne répète pas ces énoncés ici et on invite le lecteur à les produire et les montrer par lui-même, ou s'il en ressent vraiment le besoin, à aller trouver sa référence favorite sur le sujet.

Enfin, on ne cherche pas à donner des propriétés d'intégration par parties ou de changement de variable dans ce cadre. Même si on peut en trouver dans la littérature, la conviction de l'auteur est qu'il vaut mieux revenir systématiquement aux définitions avec des intégrales sur des segments puis en passant à la limite avec vigilance, plutôt que de s'encombrer d'énoncés avec des hypothèses mystérieuses.

5.3 Exercices

Exercice 5.22. 1. Soit (f_n) une suite de fonctions continues sur D. Si la série de terme général $(f_n)_{n\in\mathbb{N}}$ converge uniformément, alors la fonction $\sum_{n\in\mathbb{N}} f_n$ est continue.

2. Soit (f_n) une suite de fonctions continues sur $[a, b]$. Si la série de terme général $(f_n)_{n\in\mathbb{N}}$ converge uniformément, alors

$$\int_a^b \sum_{n\in\mathbb{N}} f_n(t)dt = \sum_{n\in\mathbb{N}} \int_a^b f_n(t)dt.$$

3. Soit (f_n) une suite de fonctions C^1 sur un intervalle I. Si
   — il existe $a \in I$ tel que la série de terme général $(f_n(a))_{n\in\mathbb{N}}$ converge vers un réel S,
   — la série de terme général $(f_n')_{n\in\mathbb{N}}$ converge uniformément, et on note $G = \sum_{n\in\mathbb{N}} f_n'$,

   alors la série de terme général $(f_n)_{n\in\mathbb{N}}$ converge simplement vers la fonction

$$F(x) = S + \int_a^x G(t)dt$$

   qui est C^1. En particulier

$$\forall x \in D, \quad \left(\sum_{n\in\mathbb{N}} f_n\right)'(x) = \sum_{n\in\mathbb{N}} f_n'(x).$$

Exercice 5.23. Il peut être commode de constater que

$$C_m^0([a,b]) = C^0([a,b]) + Esc([a,b]).$$

Exercice 5.24. Montrer que la croissance de l'intégrale de Riemann implique la propriété classique

$$\forall f \in C_m^0([a,b]), \quad \left| \int_a^b f(t)dt \right| \leq \int_a^b |f(t)|dt.$$

Concluons par quelques exercices classiques sur les fonctions réglées :

Exercice 5.25.
1. Exhiber une fonction réglée qui n'est pas continue par morceaux.[132]
2. Montrer qu'une fonction $f : [a,b] \to \mathbb{R}$ est réglée si et seulement si elle a une limite réelle à droite et à gauche en tout point de $[a,b]$.
3. Déduire de la question précédente qu'une fonction réglée a un nombre dénombrable de discontinuités.
4. Montrer qu'une fonction monotone sur $[a,b]$ est réglée.

Exercice 5.26.
1. Montrer que la fonction $\mathbb{1}_{\mathbb{Q}}$ n'est pas Riemann intégrable sur $[0,1]$.
2. Montrer que les fonctions réglées sont Riemann intégrables.
3. Montrer qu'une fonction bornée[133] est Riemann intégrable si et seulement si il existe deux suites $(f_n)_{n\in\mathbb{N}}$ et $(\varphi_n)_{n\in\mathbb{N}}$ de fonctions en escaliers telles que[134] :

$$\forall n \in \mathbb{N}, \ \ |f - f_n| \leq g_n, \qquad \text{et} \qquad \lim_{n\to+\infty} \int_a^b g_n(t)dt = 0$$

et dans ce cas

$$\int_a^b f(t)dt = \lim_{n\to+\infty} \int_a^b f_n(t)dt.$$

4. **On peut en fait caractériser comme pour les fonctions réglées le caractères Riemann-intégrable d'une fonction en fonction de son ensemble de discontinuité : montrer qu'une fonction bornée est Riemann-intégrable si et seulement si la mesure de Lebesgue de l'ensemble de ses discontinuités et nulle. Voir [Gou98, Problème 18].

Exercice 5.27. Etant donnée une fonction f continue par morceaux (*voire Riemann-intégrable sur $[a,b]$), on peut toujours définir la fonction (5.3), qu'on note F. Exhibez une fonction f telle que F n'est pas dérivable sur $[a,b]$. Montrer que F est néanmoins continue et même lipschitzienne sur $[a,b]$.[135] Voir également [God03, page 61] pour une preuve de la dérivabilité à droite et à gauche de F dans le cas où f est réglée.

Exercice 5.28. [Continuité et dérivabilité des intégrales à paramètre] Voir par exemple [Gou98, Page 157] pour la première question[136], voir aussi [Gou08b, Corollaire 1 page 158]. Soit I un intervalle de $\mathbb{R}$, $[a,b]$ un segment de $\mathbb{R}$, et $f : [a,b] \times I \to \mathbb{R}$. (5.4).

1. On suppose f continue sur $I \times [a,b]$[137]. Montrer que la fonction φ définie par (5.4) est bien définie et continue sur I.

132. Sans utiliser les questions qui suivent !
133. En fait le caractère borné était nécessaire pour la Définition 5.10, mais pour la définition équivalente qui est donnée ici, ça n'est pas nécessaire de supposer f bornée ; par contre, on peut montrer que si f satisfait cette définition alternative, alors elle est nécessairement bornée, voir [Mar09, Corollaire 10.16].
134. Comparez à la Définition-proposition 5.5
135. Comparez avec le cas où on suppose f intégrable au sens de Lebesgue ; dans ce cas on peut montrer que F est continue, mais elle n'est pas lipschitzienne a priori.
136. Pour la seconde question, je dirais qu'il manque une hypothèse de continuité sur f ; cela a été corrigé dans la réédition, mais les résultats ont changé en s'adaptant au programme de CPGE qui ont vu venir le théorème de convergence dominée, voir la note 118.
137. On sort ici du cadre des fonctions d'une variable réelle ; le lecteur qui ne connaît pas la continuité d'une fonction définie sur un sous-ensemble de $\mathbb{R}^2$ est amené à étudier le début du Chapitre III.

2. On suppose f continue sur $I \times [a,b]$, et dérivable par rapport à x et telle que $\frac{\partial f}{\partial x}$ est continue sur $I \times [a,b]$. Montrer que φ est de classe C^1, et que

$$\forall x \in I, \quad \varphi'(x) := \int_a^b \frac{\partial f}{\partial x}(t,x)dt.$$

Notez que le lecteur peut se contenter des énoncés du Chapitre IV dont les précédents seront des corollaires (l'hypothèse de domination vient du fait que f ou $\frac{\partial f}{\partial x}$ est uniformément bornée sur $K \times [a,b]$ pour tout compact K de I. Il faut néanmoins avoir l'habitude de reconnaitre qu'une telle situation est simple.

Exercice 5.29. Montrer que si f est de classe C^1, alors

$$\forall n \in \mathbb{N}^*, \quad \left| \int_a^b f(x)dx - R_n(f) \right| \leq \frac{\|f'\|_\infty (b-a)^2}{2n},$$

où R_n vaut pour R_n^g ou R_n^d. Comme l'exposant de n au dénominateur est 1, on dit que la méthode a un ordre de convergence 1 [138].

Exercice 5.30. Etudiez l'intégrabilité de la fonction $x \in]0,+\infty[\mapsto \frac{1}{x^\alpha}$

1. sur $[1,+\infty[$,
2. sur $]0,1]$,
3. sur $]0,+\infty[$,

en fonction de $\alpha \in \mathbb{R}$.

Précisons plutôt une différence notable avec le cas des séries :

Exercice 5.31. Exhiber une fonction $f : [0,+\infty[\to \mathbb{R}$ qui est continue et positive, telle que l'intégrale $\int_0^{+\infty} f(t)dt$ est convergente, mais qui ne converge pas vers 0 en $+\infty$.

---

6 Sujets de travail et de développement

Ici, on rentre un peu plus dans ce que j'appelle le folklore de l'agrégation. Il s'agit de propositions de directions de travail, se basant sur les outils vu dans ce chapitre, et qui peuvent constituer de bonnes façons de tester les connaissances de l'étudiant sur ces outils, ainsi qu'une base de travail pour les épreuves orales. Rappelons si besoin que ce ne sont que des suggestions, et que le lecteur cherchant plus d'originalité (et donc ambitieux) est chaleureusement invité à trouver ses propres directions de travail, tant qu'il n'oublie pas que je l'invite au principe d'humilité détaillé en Introduction de ce cours. Le lecteur moins ambitieux ne doit pas pour autant se jeter tête baissée en pensant qu'il doit maîtriser tous les sujets proposés ici. Il peut commencer par choisir un des sujets qui lui semble le plus abordable et/ou le plus séduisant, et revenir plus tard dans l'année, lorsqu'il aura acquis plus d'aisance et de recul, sur d'autres sujets qui lui semblaient trop difficiles en premier lieu.

Ces directions peuvent sembler plus adaptées à l'épreuve orale, et même s'il est vrai qu'elles constituent des directions de travail riches pour cette épreuve, elles constituent également une mise en contexte et une mise en œuvre des notions de ce chapitre, ce qu'est exactement (à une échelle plus large) une épreuve écrite ; si le lecteur essaie d'aborder ces paragraphes comme des exercices plutôt que comme une théorie qu'il va ingurgiter, alors cela peut constituer un bon entraînement pour l'écrit. Il n'est pas rare, d'ailleurs, qu'une ou plusieurs parties d'un écrit d'agreg abordent un tel sujet "classique" (ce qui ne veut pas dire qu'on doit essayer de se préparer à l'épreuve écrite en couvrant l'"ensemble" de ces sujets ; on ne l'a pas encore dit, mais la préparation aux épreuves ne peut être vue comme le travail acharné d'un perroquet qui chercherait à "tout" apprendre ; cette stratégie est vouée à l'échec).

138. Attention à l'expression "méthode d'ordre 1" ici, qui a plusieurs sens, voir ici page 60.

6.1 Suites récurrentes réelles $u_{n+1} = f(u_n)$

Dans ce paragraphe, on donne quelques éléments d'étude des suites définies par

$$\begin{cases} u_0 \in I \\ \forall n \in \mathbb{N}, u_{n+1} = f(u_n) \end{cases} \tag{6.1}$$

où I est un intervalle de $\mathbb{R}$, et $f : I \to \mathbb{R}$ est telle que $f(I) \subset I$.

Cette dernière hypothèse sur la stabilité de l'ensemble I par f est fondamentale, c'est ce qui permet de voir que (6.1) définit bien, par récurrence, une suite $(u_n)_{n\in\mathbb{N}}$. En pratique, le choix d'un tel intervalle I peut ne pas être anodin : si par exemple f n'est pas a priori définie sur un ensemble stable I, la première chose à faire consiste à trouver J telle que $u_0 \in J$ et $f(J) \subset J$ [139]. Aussi, affiner le choix de l'intervalle de définition pourra servir à mieux étudier la suite, par exemple en appliquant la proposition 6.1.

On reviendra plus tard dans ce cours sur les suites définies par une telle récurrence : d'une part de façon abstraite en remplaçant I par un espace métrique complet (voir notamment le théorème 1.81 du Chapitre III), et également pour discuter les récurrences d'ordre $p \geq 2$, c'est-à-dire de la forme

$$\begin{cases} u_0, u_1, \cdots, u_{p-1} \in I \\ \forall n \in \mathbb{N}, u_{n+p} = f(u_n, u_{n+1}, \cdots, u_{n+p-1}) \end{cases} \tag{6.2}$$

où $f : I^n \to I$.

6.1.1 Remarques basées sur la monotonie

On regroupe dans ce premier énoncé quelques résultats élémentaires. Ce sont les premiers éléments à tenter de mettre en œuvre pour comprendre une suite récurrente réelle. Notamment, comme on l'a remarqué en introduction de ce paragraphe, le choix de l'intervalle I peut influencer l'application de ces résultats.

Proposition 6.1. *Soit $f : I \to \mathbb{R}$ telle que $f(I) \subset I$, et $(u_n)_{n\in\mathbb{N}}$ définie par* (6.1).

1. *Si f est croissante, alors $(u_n)_{n\in\mathbb{N}}$ est monotone, à savoir croissante si $u_1 - u_0 \geq 0$, et décroissante si $u_1 - u_0 \leq 0$.*
2. *Si f est décroissante, alors $(u_{2n})_{n\in\mathbb{N}}$ et $(u_{2n+1})_{n\in\mathbb{N}}$ sont monotones et de sens de variation opposés.*
3. *Si $\forall x \in I, f(x) \geq x$, alors $(u_n)_{n\in\mathbb{N}}$ est croissante.*
   Si $\forall x \in I, f(x) \leq x$, alors $(u_n)_{n\in\mathbb{N}}$ est décroissante.
4. *Si $(u_n)_{n\in\mathbb{N}}$ converge vers $\ell \in I$ [140], et si f est continue sur I [141], alors ℓ est un point fixe de f, c'est-à-dire $f(\ell) = \ell$.*

Remarque 6.2. Une fois n'est pas coutume, je ne pense pas qu'il faille apprendre tout cela par cœur. Il convient plutôt de savoir que de tels résultats existent, de les connaître approximativement, et de savoir les retrouver et les appliquer à la situation à laquelle vous êtes confrontés.

Il convient également, et c'est le plus important, de savoir donner une représentation graphique d'une suite récurrente, et de se convaincre des résultats précédents par cette représentation graphique [142].

Remarque 6.3. Si vous prouvez la monotonie de $(u_n)_{n\in\mathbb{N}}$, vous pouvez en déduire qu'elle a une limite réelle ou infinie. Si cette limite est susceptible d'être dans I (voir la note 140) et que f a le bon goût d'être continue, alors vous savez que vous devez chercher cette limite parmi les points fixes de f.

139. Il peut arriver qu'on doivent étudier quelques premiers termes (disons n_0) à la main puis trouver J stable par f tel que $u_{n_0} \in J$.
140. Attention, il peut arriver que $(u_n)_{n\in\mathbb{N}}$ converge vers une limite qui n'est plus dans l'ensemble de définition de f.
141. La continuité en ℓ suffit, mais bien souvent, c'est ℓ qu'on cherche, d'où le fait qu'on a mis l'hypothèse sur tout l'intervalle I.
142. Pour l'heure, ce document n'a pas de dessin ; c'est un défaut non négligeable, qui sera corrigé quand le temps me le permettra.

Dans le second point de la proposition 6.1 où vous concluez que $(u_{2n})_{n\in\mathbb{N}}$ et $(u_{2n+1})_{n\in\mathbb{N}}$, vous savez qu'elles ont des limites ℓ et ℓ'. Attention, ℓ et ℓ' seront des points fixes de $f\circ f$ (si elle est continue), par forcément des points fixes de f. Il peut aussi tout à fait arriver que $\ell \neq \ell'$. Regardez par exemple la suite

$$u_0 = 42, \quad \forall n \in \mathbb{N}, u_{n+1} = -u_n.$$

Démonstration.

1. Le premier point se fait par une récurrence élémentaire, tout reposant sur l'initialisation.
2. Pour le second point, on remarque que $f\circ f$ est croissante, et comme

$$\forall n \in \mathbb{N}, u_{2(n+1)} = f\circ f(u_{2n}), \quad u_{2(n+1)+1} = f\circ f(u_{2n+1}),$$

   on en déduit que les deux suites sont monotones, et que leurs sens de monotonie sont liés aux signes de $u_2 - u_0$ et $u_3 - u_1$ respectivement. Mais si par exemple $u_2 \geq u_0$, alors par décroissance de f, $u_3 = f(u_2) \leq f(u_0) = u_1$, ce qui explique que $(u_{2n})_{n\in\mathbb{N}}$ et $(u_{2n+1})_{n\in\mathbb{N}}$ ont des sens de monotonie opposés.
3. Cet énoncé est très simple, il est là pour mémoire. En effet, si $\forall x \in I, f(x) \geq x$ par exemple, alors

$$\forall n \in \mathbb{N}, \quad u_{n+1} = f(u_n) \geq u_n$$

   d'où la monotonie de $(u_n)_{n\in\mathbb{N}}$ [143].
4. Pour le dernier point, sachant que $(u_n)_{n\in\mathbb{N}}$ converge, on sait que $(u_{n+1})_{n\in\mathbb{N}}$ converge vers la même limite, et donc par continuité de f (en cette limite ℓ), on peut passer à la limite dans l'expression

$$\forall n \in \mathbb{N}, u_{n+1} = f(u_n)$$

   d'où le résultat.

□

6.1.2 *Points fixes attractifs, répulsifs

On a vu l'importance des points fixes de f à la proposition 6.1. On peut classifier un peu plus précisément ces points fixes avec l'énoncé suivant :

Proposition 6.4. *Soit $f : I \to \mathbb{R}$, et ℓ un point fixe de f qui est à l'intérieur de I. On suppose f dérivable en ℓ.*

1. *Si $|f'(\ell)| < 1$, alors il existe $\alpha > 0$ tel que si $u_0 \in]\ell-\alpha, \ell+\alpha[$, alors la suite définie par (6.1) est bien définie et converge vers ℓ. On dit que ℓ est un point fixe attractif.*
2. *Si $|f'(\ell)| > 1$, alors il existe $\alpha > 0$ tel que si $u_0 \in]\ell-\alpha, \ell+\alpha[\setminus\{\alpha\}$, alors la suite définie par (6.1) sort de l'intervalle $]\ell-\alpha, \ell+\alpha[$. On dit que ℓ est un point fixe répulsif*[144] [145].

Remarque 6.5. Dans le premier point, on va montrer que la convergence est au moins géométrique. Vous trouverez cet énoncé dans [Rou03, Exercice 48], avec quelques éléments supplémentaires sur la vitesse de convergence des suites. Par contre, il suppose et utilise le caractère continu de f', ce que nous ne faisons pas ici [146].

Remarque 6.6. Le cas où ℓ est un point fixe tel que $|f'(\ell)| = 1$ est plus délicat à analyser, il s'appelle un point fixe parabolique. Il peut y avoir convergence vers ℓ, mais alors la convergence est plus lente que dans le cas attractif.

---

143. J'ai trouvé sur internet des preuves par récurrence de ce fait. Ce n'est pas faux, mais quand même particulièrement cocasse, et un peu effrayant, à mon avis.
144. Notez qu'on a exclu $u_0 = \ell$, puisqu'alors $\forall n \in \mathbb{N}, u_n = \ell$.
145. On n'a rien dit sur le caractère défini ou non de la suite $(u_n)_{n\in\mathbb{N}}$. Les termes sont bien définis tant qu'ils restent dans $]\ell-\alpha, \ell+\alpha[$, on ne sait pas ce qu'il en est une fois que la suite sort de l'intervalle.
146. Ça ne veut pas dire qu'il faut négliger les preuves dans ce cas, il est encore une fois instructif de s'essayer aux deux preuves et de mesurer les différences.

Démonstration. 1. Supposons $|f'(\ell)| < 1$. Comme ℓ est à l'intérieur de I, il existe $\beta > 0$ tel que $]\ell-\beta, \ell+\beta[\subset I$. De plus par continuité de la fonction valeur absolue,

$$\lim_{x\to\ell}\left|\frac{f(x)-f(\ell)}{x-\ell}\right| = |f'(\ell)|,$$

donc en posant $k = \frac{|f'(\ell)|+1}{2} \in]|f'(\ell)|, 1[$, il existe $\alpha \in]0, \beta[$ tel que

$$\forall x \in]\ell-\alpha, \ell+\alpha[\setminus\{\alpha\}, \quad \left|\frac{f(x)-f(\ell)}{x-\ell}\right| \leq k.$$

D'une part, $]\ell-\alpha, \ell+\alpha[$ est stable par f car

$$\forall x \in]\ell-\alpha, \ell+\alpha[, |f(x)-\ell| = |f(x)-f(\ell)| \leq k|x-\ell| \leq k\alpha < \alpha.$$

D'autre part, si $u_0 \in]\ell-\alpha, \ell+\alpha[$ et $(u_n)_{n\in\mathbb{N}}$ est définie par (6.1), alors

$$\forall n \in \mathbb{N}, \quad |u_{n+1}-\ell| = |f(u_n)-f(\ell)| \leq k|u_n-\ell|$$

ce qui donne par récurrence immédiate :

$$\forall n \in \mathbb{N}, \quad |u_n-\ell| \leq k^n|u_0-\ell|,$$

d'où la convergence de $(u_n)_{n\in\mathbb{N}}$ vers ℓ.

2. Supposons $|f'(\ell)| > 1$. On pose $k = \frac{|f'(\ell)|+1}{2} \in]1, |f'(\ell)|[$ et comme au point précédent, on montre l'existence de $\alpha > 0$ tel que $]\ell-\alpha, \ell+\alpha[\subset I$ et

$$\forall x \in]\ell-\alpha, \ell+\alpha[, \quad |f(x)-f(\ell)| \geq k|x-\ell|.$$

Ainsi, si on suppose que pour tout $n \in \mathbb{N}, u_n \in]\ell-\alpha, \ell+\alpha[$, alors comme ci-dessus, on montre

$$\forall n \in \mathbb{N}, \quad \alpha > |u_n-\ell| \geq k^n|u_0-\ell|,$$

ce qui implique $u_0 = \ell$, ce qu'on a exclu dans l'énoncé.

□

6.1.3 **Chaos

Dans ce paragraphe, je souhaite vous convaincre que, contrairement à ce que laissent présager les éléments simples qu'on a vu précédemment, l'étude d'une suite récurrente de la forme (6.1), qui s'écrit d'ailleurs aussi $(f^n(x_0))_{n\in\mathbb{N}}$ où $f^n = f\circ\ldots\circ f$ est la composée $n$ième de f, n'est en général pas simple du tout. C'est en fait ce qu'on appelle un système dynamique à temps discret. [147]

Théorème de Charkovski [148]

Commençons par une définition :

Définition 6.7. Soit $f : I \to \mathbb{R}$ où I est un intervalle et $f(I) \subset I$. On dit que $x \in I$ est un point de période $k \in \mathbb{N}^*$ (pour f) si

$$f^k(x) = x, \quad \text{et} \quad \forall i \in [\![1, k-1]\!], f^i(x) \neq x.$$

Autrement dit, x est un point fixe de f^k où k est minimal. On dit que $(x, f(x), \ldots, f^{k-1}(x))$ est un cycle d'ordre k.

147. On abordera les systèmes dynamiques à temps continu dans le Chapitre X ; mais contrairement à ce qu'on pourrait penser, en général l'analyse des cas discrets est bien plus délicates que celle des cas continus.

148. Ce mathématicien est ukrainien, et les orthographes qu'on peut trouver sont multiples ; on peut trouver Sarkovskii, Charkovsky, Sharkovsky ou encore Šarkov'skiĭ

Les points fixes de f sont les points de période 1. Si u_0 est un tel point, la suite $(f^n(u_0))_{n\in\mathbb{N}}$ est constante. Si par contre u_0 est un point fixe d'ordre $k \geq 2$, alors $(f^n(u_0))_{n\in\mathbb{N}}$ est une suite périodique de période k.

Théorème 6.8 (Version faible du théorème de Charkovski)**.** *Soit $f : I \to \mathbb{R}$ où I est un intervalle et $f(I) \subset I$. On suppose f continue. Si f admet un point de période 3, alors f admet un point de période n pour tout $n \in \mathbb{N}^*$.*

La démonstration est élémentaire mais difficile, et repose uniquement sur le théorème des valeurs intermédiaires. Elle peut faire l'objet d'un développement (un peu délicat) et on trouvera une preuve dans [FGN07b, page 92].

Remarque 6.9. La version forte du théorème (qui date de 1964) va beaucoup plus loin. Elle définit un ordre (qu'on note $\lhd$) sur l'ensemble des entiers naturels non nuls, et affirme que pour tout $k \in \mathbb{N}^*$, si f admet un point de période k, alors pour tout $n \in \mathbb{N}^*$ tel que $k \lhd n$, f admet un point de période n.

L'ordre s'écrit ainsi :

$$\begin{array}{ccccccccccccc}
3 & \lhd & 5 & \lhd & 7 & \lhd & 9 & \lhd & \cdots & \lhd & (2n+1) & \lhd & \ldots \\
3.2 & \lhd & 5.2 & \lhd & 7.2 & \lhd & 9.2 & \lhd & \cdots & \lhd & (2n+1).2 & \lhd & \ldots \\
3.2^2 & \lhd & 5.2^2 & \lhd & 7.2^2 & \lhd & 9.2^2 & \lhd & \cdots & \lhd & (2n+1).2^2 & \lhd & \ldots \\
 & \vdots & \vdots & \vdots & & & & & & & & & \\
\ldots & \lhd & 2^n & \lhd & \ldots & \lhd & 2^3 & \lhd & 2^2 & \lhd & 2 & \lhd & 1
\end{array}$$

On constate que 3 est le “plus petit” entier pour l'ordre de Charkovski, ce qui redonne bien la version faible énoncée ci-dessus.

Exemple de la suite logistique : $f(x) = \lambda x(1-x)$ [149].

On considère désormais la suite récurrente définie par :

$$\begin{cases} u_0 \in [0,1] \\ \forall n \in \mathbb{N}, u_{n+1} = \lambda.u_n(1-u_n) \end{cases} \tag{6.3}$$

où λ est un réel strictement positif. En fait, on prendra $\lambda \in]0,4]$ car pour que l'ensemble $[0,1]$ soit stable par la fonction $f : x \mapsto \lambda x(1-x)$, il est nécessaire que $\lambda \leq 4$.

On va donner seulement quelques éléments permettant de comprendre cette suite. On renvoie à ce document de D. Perrin pour une analyse plus détaillée (mais accessible). On peut remarquer que

— si $\mu < 1$, alors f a un unique point fixe, à savoir 0, qui est attractif,

— si $\mu \in]1,3[$, alors f a exactement deux points fixes, à savoir 0 et $\tau := 1 - \frac{1}{\mu}$, le premier étant répulsif et le second attractif,

— si $\mu \in]3,4[$ alors f a les mêmes points fixes qu'au cas précédent, mais ils sont tous les deux répulsifs.

Notez qu'on a exclu les cas intermédiaires ($\mu = 1$ ou $\mu = 3$) mais qu'on pourrait aussi les analyser.

On peut par exemple montrer (voir [Dev89, Proposition 5.3] pour le second point, le premier étant plus simple) :

Théorème 6.10. *On suppose $u_0 \in]0,1[$* [150] *et on note $(u_n)_{n\in\mathbb{N}}$ la suite définie par (6.3).*

— *si $\mu \leq 1$, alors*

$$u_n \xrightarrow[n\to+\infty]{} 0.$$

149. Ce modèle est dû à P.F. Verhulst qui vers 1840 le propose pour représenter l'évolution d'une population ; c'est une amélioration du modèle de T.R. Malthus (fin 18e) qui était une suite géométrique.

150. On a exclu $u_0 \in \{0,1\}$ car 0 et 1 sont les deux antécédents de 0 par f, et donc les suites partant de 0 et 1 stationnent à 0 en au plus une itération.

— *si* $\mu \in]1,3[$, *alors*

$$u_n \xrightarrow[n\to+\infty]{} \tau = 1 - \frac{1}{\mu}.$$

Pour $\mu > 3$, les choses se compliquent : on doit s'intéresser aux points fixes des itérées de f. On peut par exemple montrer :

— si $\mu \in]3, 1+\sqrt{6}[$ et $u_0 \notin \{0, \tau, \frac{1}{\mu}, 1\}$, alors $(u_{2n})_{n\in\mathbb{N}}$ et $(u_{2n+1})_{n\in\mathbb{N}}$ convergent vers les points fixes de f^2.

— lorsque μ parcourt l'intervalle $]1+\sqrt{6}, \mu_\infty \simeq 3,5699456[$, on voit apparaître des points de période $4, 8, \ldots 2^k \ldots$, autrement dit les suites $(f^n(x))_{n\in\mathbb{N}}$ vont osciller entre 4,8... valeurs.

— pour $\mu \in [1+\sqrt{8}, 4]$, f admet un cycle d'ordre 3 (et donc de tout ordre d'après le théorème de Charkovski).

— pour $\mu = 4$, la dynamique est chaotique [151], au sens qu'il existe $u_0 \in]0,1[$ tel que l'ensemble $(f^n(u_0))_{n\in\mathbb{N}}$ est dense dans $[0,1]$.

Remarque 6.11. On peut également parler de caractère attractif ou répulsif d'un cycle $\{x_1, x_2, \ldots x_k\}$ d'ordre k. On montre d'ailleurs que le cycle est attractif si

$$\left|\prod_{i=1}^{n} f'(x_i)\right| < 1,$$

ce qui découle du fait que

$$(f^k)'(x_1) = \prod_{i=1}^{n} f'(x_i).$$

Remarque 6.12. Citons deux développements possibles en lien avec cette suite logistique : [FGN07b, Exercice 2.20] remplace la famille $f(x) = \mu x(1-x)$ par $g(x) = 1 - \lambda x^2$. Comme expliqué dans [Per, Paragraphe 7.2], ces familles sont conjuguées. Il semble [152] qu'ils étudient une plage de paramètre λ telle qu'il y a un cycle d'ordre 2, ce qui corrspond à la plage $\mu \in]1,3[$. Aussi [FGN07b, Exercice 2.21] rentre dans une autre classe de conjugaison et semble correspondre au cas $\mu = 4$. Ces classes de conjugaison semblent avoir des calculs un peu plus commode que la suite logistique initiale.

6.2 Fonctions réelles monotones et convexes

On propose d'étudier un peu plus en détail ces deux classes de fonctions, qui sont d'ailleurs au cœur de la leçon 229.

6.2.1 Fonctions monotones

Points de discontinuité : Commençons par regarder la continuité des fonctions monotones :

Proposition 6.13. *Soit $f :]a,b[\to \mathbb{R}$ une fonction monotone, avec $-\infty \leq a < b \leq +\infty$. Alors l'ensemble des points de discontinuité de f est dénombrable.*

Démonstration. Supposons par exemple f croissante. On sait que pour tout $x \in]a,b[$, f admet une limite à gauche et à droite en x, qu'on note $f(x^-)$ et $f(x^+)$. Ces limites sont réelles, puisqu'on peut écrire

$$f\left(\frac{x+a}{2}\right) \leq f(x^-) \leq f(x) \leq f(x^+) \leq f\left(\frac{x+b}{2}\right).$$

On note $D = \{x \in]a,b[, f(x^+) > f(x^-)\}$, qui est l'ensemble des points de discontinuité de f. Pour chaque $x \in D$, par densité de $\mathbb{Q}$ dans $\mathbb{R}$, on peut choisir un point $y_x \in]f(x^-), f(x^+)[\cap \mathbb{Q}$. Par monotonie de f, on voit que l'application

$$\phi : \begin{array}{ccc} D & \longrightarrow & \mathbb{Q} \\ x & \longmapsto & y_x \end{array}$$

151. Attention, on donne ici une définition simplifiée de chaos ; on se l'est permis car il n'y a pas, à mon avis, de définition universellement acceptée. On renvoit à [Dev89] pour plus d'information.

152. À vérifier.

est injective, et donc D est dénombrable [153]. □

Remarque 6.14. On peut montrer un résultat plus général [154], à savoir que l'ensemble des points de discontinuité de première espèce d'une fonction réelle (ceux pour lesquels il y a une limite à droite et à gauche qui sont distinctes) est dénombrable. Voir [Gou98, Exercice 7 page 36] où l'énoncé est donné lorsque l'espace d'arrivée est un espace métrique (ce qui ne change rien, vous pouvez remplacer par $\mathbb{R}$ sans perdre l'intérêt de l'énoncé) [155].

***Théorème de Helly :** Le résultat suivant permet de montrer très facilement un résultat de compacité via la monotonie des fonctions considérées. On trouvera la preuve dans [Gou98, page 235] ou dans [FGN09] : il peut faire l'objet d'un développement.

Théorème 6.15. *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions définies et croissantes sur I un intervalle ouvert de $\mathbb{R}$, et telle que*

$$\forall x \in I, (f_n(x))_{n\in\mathbb{N}} \text{ est bornée dans } \mathbb{R}$$

Alors il existe une extraction φ et une fonction $f : I \to \mathbb{R}$ telle que

$$f_n \xrightarrow[n\to+\infty]{} f \quad \text{simplement sur } I.$$

On peut se servir de ce résultat en théorie de la mesure et des probabilités, via la notion de fonction de répartition.

****Dérivabilité presque partout.** Citons pour mémoire le résultat suivant, très délicat :

Théorème 6.16. *Soit $f :]a,b[\to \mathbb{R}$ une fonction monotone, avec $-\infty \leq a < b \leq +\infty$, alors f est dérivable presque partout sur $]a,b[$.*

Ce résultat est dû à Lebesgue (1904), on peut en trouver une preuve dans [Rud98, page 193]. Attention par contre, même si la fonction f' peut s'avérer intégrable au sens de Lebesgue, il n'est pas vrai en général que

$$\forall (x,y) \in]a,b[^2,\ f(y) - f(x) = \int_x^y f'(t)dt$$

même si on suppose f continue.

6.2.2 Fonctions convexes

Dans tout le paragraphe, I sera un intervalle de $\mathbb{R}$.

Définitions et caractérisation :

Définition 6.17. Soit I un intervalle de $\mathbb{R}$ [156].

- Une fonction $f : I \to \mathbb{R}$ est dite convexe si

$$\forall (x,y) \in I^2,\ \ \forall \lambda \in [0,1],\ \ f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \tag{6.4}$$

- Elle sera dite strictement convexe si

$$\forall (x,y) \in I^2 \text{ tel que } x \neq y,\ \ \forall \lambda \in]0,1[,\ \ f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y).$$

- elle sera dite concave (resp. strictement concave) si $-f$ est convexe (resp. strictement convexe).

153. On renvoit au Chapitre IV pour les notions de base de la dénombrabilité.
154. Il s'agit du théorème de Froda (1929), même si d'après Wikipedia le résultat avait été obtenu par les époux Young sous des formes plus générales vers 1910.
155. L'énoncé est un peu différent, car l'exercice suppose que tous les points de discontinuité sont de première espèce ; mais on peut sans presque rien changer adapter la démonstration pour obtenir l'énoncé proposé.
156. Plus généralement, on verra qu'une fonction convexe doit avoir un ensemble de définition qui est convexe ; voyez pourquoi en regardant la définition !

Remarque 6.18. Comme souvent, on vous suggère d'avoir une représentation graphique en tête d'une telle formule : le segment qui rejoint les points $(x, f(x))$ et $(y, f(y))$ est décrit par le membre de droite de (6.4), et se trouve donc au dessus du graphe de f entre ces deux points.

Commeçons pas un premier lien avec les fonctions croissantes :

Proposition 6.19. *Une application $f : I \to \mathbb{R}$ est convexe si et seulement si*

$$\forall x_0 \in I, \text{ l'application } \Delta_{x_0}^f : x \in I \setminus \{x_0\} \longmapsto \frac{f(x)-f(x_0)}{x-x_0} \quad \text{est croissante.}$$

Démonstration. $\Rightarrow$ Soit $x < y$ dans $I \setminus \{x_0\}$. Supposons d'abord $x_0 < x < y$. Il est classique qu'il existe $\lambda \in]0,1[$ tel que $x = \lambda x_0 + (1-\lambda)y$, et d'ailleurs $\lambda = \frac{y-x}{y-x_0}$.

On se demande si

$$\frac{f(x)-f(x_0)}{x-x_0} \leq \frac{f(y)-f(x_0)}{y-x_0}$$

qui se réécrit

$$f(x) \leq \frac{x-x_0}{y-x_0}\big[f(y)-f(x_0)\big] + f(x_0) = \frac{x-x_0}{y-x_0} f(y) + \frac{y-x}{y-x_0} f(x_0) = (1-\lambda)f(y) + \lambda f(x_0).$$

Ceci découle bien de (6.4). Les deux autres cas $x < x_0 < y$ et $x < y < x_0$ se traitent de façon similaire, en écrivant le terme du milieu comme combinaison convexe des termes à l'extérieur [157].

$\Leftarrow$ Soit $x < y$, et $\lambda \in]0,1[$; on pose $z = \lambda x + (1-\lambda)y$. Comme $x < z < y$, on peut écrire :

$$\Delta_x^f(z) \leq \Delta_x^f(y), \quad \text{ce qui donne après calcul} \quad f(z) \leq \lambda f(x) + (1-\lambda) f(y).$$

□

Lemme 6.20 (Lemme des 3 pentes [158]). *Soit $f : I \to \mathbb{R}$ convexe. Alors pour tout $x < z < y$ dans I, on a*

$$\frac{f(z)-f(x)}{z-x} \leq \frac{f(y)-f(x)}{y-x} \leq \frac{f(y)-f(z)}{y-z}. \tag{6.5}$$

Démonstration. Cela découle de la proposition précédente, car

$$\frac{f(z)-f(x)}{z-x} = \Delta_x^f(z) \leq \Delta_x^f(y) = \frac{f(y)-f(x)}{y-x} = \Delta_y^f(x) \leq \Delta_y^f(z) = \frac{f(y)-f(z)}{y-z}.$$

□

On donne désormais des caractérisations très pratiques de la convexité pour les fonctions suffisamment dérivables :

Théorème 6.21 (Caractérisations de la convexité). *Soit $f : I \to \mathbb{R}$.*

— *On suppose f dérivable sur I. Alors*

$$f \text{ convexe sur } I \;\Leftrightarrow\; f' \text{ est croissante sur } I \;\Leftrightarrow\; \Big[\forall (x,y) \in I,\; f(y) \geq f(x) + (y-x) f'(x)\Big].$$

La dernière formule signifie que le graphe de f est situé au-dessus de ses tangentes.

— *On suppose f deux fois dérivable sur I. Alors*

$$f \text{ convexe sur } I \;\Leftrightarrow\; f'' \geq 0 \quad \text{sur } I.$$

157. Attention quand même aux signes des termes dans la manipulation des inégalités.
158. Une fois n'est pas coutume, je conseille fortement de ne pas retenir une telle formule par cœur. On a vite fait de se mélanger ; il suffit de savoir qu'une telle formule existe, et de la retrouver par un dessin.

Démonstration. Pour la première partie, supposons f dérivable.
1. $\Rightarrow$ 2. : supposons f convexe, et $x < y$ dans I. Dans (6.5), on fait tendre z vers x^+ dans l'iégalité de gauche, ce qui donne $f'(x) \leq \frac{f(y)-f(x)}{y-x}$. De même, en faisant tendre z vers y^- dans l'inégalité de droite de (6.5), on obtient $\frac{f(y)-f(x)}{y-x} \leq f'(y)$, d'où le résultat.
2. $\Rightarrow$ 3. on suppose f' croissante. Soit $(x, y) \in I$, et supposons $x < y$. Alors par l'égalité des accroissements finis, il existe $c \in]x, y[$ tel que $\frac{f(y)-f(x)}{y-x} = f'(c_x)$, mais alors par croissance de f', $f'(c_x) \geq f'(x)$ ce qui donne bien $f(y) \geq f(x) + (y-x)f'(x)$. Le cas $x > y$ est similaire.
3. $\Rightarrow$ 1. On suppose $\forall (x, y) \in I,\ f(y) \geq f(x) + (y-x)f'(x)$. On fixe y et on prend le sup en x. Cela donne

$$f(y) = \sup_{x \in I} \left\{ f(x) + (y-x)f'(x) \right\},$$

car on obtient d'abord l'inégalité $\leq$, mais on se rend compte qu'il y a égalité lorsque $x = y$. Ainsi f s'écrit comme la borne supérieure de fonction affine donc convexe, donc par l'exercice 6.31, f est convexe.
Le second point combine le premier avec le Théorème 3.45. $\square$

Remarque 6.22. Revoyez cette preuve en supposant f de classe C^1, et constatez que certains passages peuvent avoir une preuve alternative (plus simple ?) via l'égalité $f(y) - f(x) = \int_x^y f'(t)dt$.

***Régularité des fonctions convexes en dimension 1 :** Nous revenons sur la question de la régularité des fonctions convexes, et notamment on donne une version plus forte du théorème 6.21.

Théorème 6.23. *Soit $f : I \to \mathbb{R}$ convexe avec I ouvert. Alors*

— *f est continue sur I.*

— *f admet des dérivées à gauche et à droite en tout point de I, et*

$$\forall (x, y) \in I \text{ tels que } x < y, \quad f'_g(x) \leq f'_d(x) \leq f'_g(y).$$

— *Pour tout $(x, y) \in I^2$ et tout $\alpha \in [f'_g(x), f'_d(x)]$, on a*

$$f(y) \geq f(x) + \alpha(y - x).$$

Remarque 6.24. Nous reviendrons sur la question de la continuité des fonctions convexes dans un cadre plus général. Notamment, la propriété reste vraie si l'espace de départ est un ouvert convexe de $\mathbb{R}^n$; mais la preuve sera différente, notamment parce qu'on ne pourra plus passer par la dérivation à droite et à gauche. On montrera en fait que f est lipschitzienne. On évoquera aussi que ceci devient faut si on remplace $\mathbb{R}^n$ par un espace de dimension infinie.

Remarque 6.25. On déduit du théorème précédent que si f est convexe, alors elle est dérivable en dehors d'un nombre dénombrable de points (points pour lesquels $f'_g(x) \neq f'_d(x)$).

Démonstration. Soit donc $x \in I$. Il existe $\varepsilon > 0$ tel que $[x - \varepsilon, x + \varepsilon] \subset I$. Alors on peut écrire, d'après le Lemme 6.20

$$\forall h \in]0, \varepsilon[, \quad \frac{f(x) - f(x-\varepsilon)}{\varepsilon} \leq \Delta_x^f(x+h) = \frac{f(x+h) - f(x)}{h} \leq \frac{f(x+\varepsilon) - f(x)}{\varepsilon}. \tag{6.6}$$

Le fait que $h \in]0, \varepsilon[\mapsto \Delta_x^f(h)$ soit borné implique que $\lim_{h \to 0^+} \left[f(x+h) - f(x) \right] = 0$ et donc f est continue à droite en x[159]. De plus par la monotonie de Δ_x^f vue à la proposition 6.19, Δ_x^f a une limite à droite en x, cette dernière étant réelle du fait de (6.6), ce qui donne la dérivabilité à droite de f en x.

On procède de même pour montrer la dérivabilité à gauche. Il reste à écrire, encore une fois d'après le Lemme 6.20 que

$$\forall h \in]0, \varepsilon[, \quad \frac{f(x) - f(x-h)}{h} \leq \frac{f(x+h) - f(x)}{h}$$

159. On aurait aussi pu appliquer la Remarque 3.35, qui dit que la continuité va découler du caractère dérivable à gauche et à droite, que l'on montre juste après.

ce qui donne à la limite $\varepsilon \to 0$ $f'_g(x) \leq f'_d(x)$.

Pour montrer que si $x < y$, $f'_d(x) \leq f'_g(y)$, on procède exactement comme dans la preuve de $1 \Rightarrow 2$ du théorème 6.21, et on constate qu'on obtient en fait

$$f'_d(x) \leq \frac{f(y) - f(x)}{y - x} \leq f'_g(y).$$

Cette dernière inégalité (partie gauche) donne, si $\alpha \in [f'_g(x), f'_d(x)]$,

$$\forall x < y, \quad f(y) \geq f(x) + f'_d(x)(y - x) \geq f(x) + \alpha(y - x)$$

et la partie droite donne, après avoir échangé les rôles de x et y :

$$\forall x > y, \quad f(y) \geq f(x) + f'_g(x)(y - x) \geq f(x) + \alpha(y - x)$$

ce qui conclut la preuve. $\square$

6.3 Théorème de "prolongement" C^1

Exercice 6.26. 1. Soit I un intervalle, et $a \in I$.

(a) Soit $f : I \to \mathbb{R}$ continue sur I et dérivable sur $I \setminus \{a\}$. On suppose de plus qu'il existe $\ell \in \mathbb{R}$ tel que

$$\lim_{x \to a} f'(x) = \ell.$$

Montrer que f est dérivable en a [160], et que $f'(a) = \ell$.

(b) Soit $f : I \to \mathbb{R}$ continue sur I, de classe C^k sur $I \setminus \{a\}$ pour un certain $k \in \mathbb{N}^*$. On suppose que $f^{(j)}$ admet une limite réelle en a pour tout $j \in [\![1, k]\!]$.

Montrer que f est de classe C^k sur I.

2. (a) Soit $a \in \mathbb{R}$, et $b \in \mathbb{R} \cup \{+\infty\}$ tel que $a < b$. On suppose $f :]a, b[\to \mathbb{R}$ dérivable, et telle qu'il existe $\ell \in \mathbb{R}$ tel que

$$\lim_{x \to a} f'(x) = \ell.$$

Montrer que f se prolonge par continuité en a, et que ce prolongement est dérivable sur $[a, b[$. [161]

(b) Montrer qu'il existe une fonction $f : \mathbb{R}^* \to \mathbb{R}$ dérivable et telle que f' a une limite réelle en 0, mais qui n'est pas prolongeable par continuité en 0.

6.4 *Un théorème de Glaeser

Le résultat suivant (qu'on a trouvé dans [GT98]) donne des conditions pour que la racine carrée d'une fonction positive soit de classe C^1. On constate que la question n'est pas anodine par exemple en considérant la fonction $x \mapsto x^2$ qui est positive et C^∞ sur $\mathbb{R}$, mais sa racine carrée n'est pas dérivable en 0.

Théorème 6.27. *Soit I un intervalle ouvert, et $f : I \to \mathbb{R}$ de classe C^2 et positive. On pose $F = f^{-1}(\{0\})$. Alors*

$$\sqrt{f} \text{ est de classe } C^1 \quad \Leftrightarrow \quad \forall t \in F, \quad f''(t) = 0.$$

160. Attention, dans la littérature on lit parfois les expressions "Théorème de prolongement de la dérivée", ou "Théorème de prolongement C^1". Ces expressions prêtent à confusion, car rigoureusement, dans ces premières questions, on ne prolonge aucune fonction. La fonction f est initialement définie sur I, donc il n'y a rien à prolonger. Quant à f', elle est définie en a si et seulement si f est dérivable en a. Ici on montre qu'effectivement f est dérivable en a, ce qui ne correspond pas à un prolongement de f', simplement au fait qu'elle est effectivement bien définie en a.

161. Ici par contre, on prolonge bien f. La difficulté réside en fait à montrer que f a bien une limite en a.

Démonstration. Soit $t_0 \in F$. Par l'absurde, si $f'(t_0) \neq 0$, on a $f(t_0+h) \sim_{h\to 0} f'(t_0)h$ change de signe au voisinage de 0, ce qui contredit l'hypothèse de signe faite sur f, donc $f'(t_0) = 0$. Par suite,

$$\frac{f(t_0+h)}{h^2} \longrightarrow \frac{f''(t_0)}{2}$$

et donc $f''(t_0) \geq 0$.

Posons $g = \sqrt{f}$. De plus

$$g(t_0+h) = |h|\sqrt{\frac{f''(t_0)}{2} + o_{h\to 0}(1)},$$

et donc g est dérivable en t_0 si et seulement si f'' s'annule sur F.

Supposons $\forall t \in F,\ f''(t) = 0$: on sait donc que g est dérivable sur I, il reste à montrer que g est de classe C^1 sur I. Comme la fonction racine carrée est C^∞ en dehors de 0, par composition g est de classe C^2 sur l'ouvert $I \setminus F$. Soit donc $t_0 \in F$. Pour $\alpha > 0$, on pose $I(\alpha) = [t_0 - \alpha, t_0 + \alpha]$. Comme I est ouvert, il existe α_0 tel que $I(\alpha_0) \subset I$. Soit $\alpha \leq \alpha_0$. Par le théorème de Taylor-Lagrange, pour $(t, t+h) \in I(\alpha)$, il existe $\xi \in I(\alpha)$ tel que

$$0 \leq f(t+h) = f(t) + hf'(t) + \frac{h^2}{2}f''(\xi) \leq f(t) + hf'(t) + \frac{h^2}{2}\|f''\|_{\infty,I(\alpha)}$$

On pose $M(\alpha) = \|f''\|_{\infty,I(\alpha)}$. Si $M(\alpha) = 0$ alors $f = 0$ sur $I(\alpha)$ et donc g est C^1 sur un voisinage de t_0. On suppose donc $M(\alpha) > 0$.

Pour $t \in I(\alpha/2)$, on constate que le trinôme

$$h \mapsto f(t) + hf'(t) + \frac{h^2}{2}M(\alpha) \tag{6.7}$$

est positif sur $[-\alpha/2, \alpha/2]$, et son minimum est atteint en $h_0 = -\frac{f'(t)}{M(\alpha)}$.

Mais comme $f'(t_0) = 0$, par l'inégalité des accroissements finis on a aussi $|f'(t)| \leq M(\alpha)|t - t_0|$, et donc $h_0 \leq |t - t_0| \leq \alpha/2$, et donc le trinôme (6.7) est positif sur $\mathbb{R}$, son discriminant est donc négatif. Ainsi

$$f'(t)^2 \leq 2M(\alpha)f(t).$$

La dérivée de g sur $I \setminus F$ est $\frac{f'}{2\sqrt{f}}$, et 0 sur F. Mais on vient de montrer

$$\forall t \in I(\alpha/2),\quad |g'(t)| \leq \sqrt{\frac{M(\alpha)}{2}}$$

et comme f est de classe C^2, $M(\alpha) \xrightarrow[\alpha\to 0]{} 0$, on conclut que g' est continue en t_0, et donc sur I. □

6.5 Autres suggestions

1. Les théorèmes de Dini : Voir [FGN09, page 155], [Nou06, page 109], [Gou08b, page 228]
2. Equirépartition
3. Méthode de Newton (version uni-dimensionnelle) : Voir [Rou03, Exercice 49]
4. *Fonctions à variations bornées.
5. **Construction des fonctions de Cantor et de Weierstrass
6. **Intégrale de Riemann-Stieltjes

6.6 Exercices

Exercice 6.28. Etant donnés $(a, b) \in \mathbb{R}^2$, et $(u_0, v_0) \in \mathbb{R}^2$, les suites

$$\forall n \in \mathbb{N}, \quad u_{n+1} = u_n + b, \quad v_{n+1} = a.v_n$$

sont appelées arithmétique et géométrique, respectivement.

1. Calculez les termes généraux des suites $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$.
2. Etant donné $w_0 \in \mathbb{R}$, on définit $\forall n \in \mathbb{N}, \quad w_{n+1} = a.w_n + b$ (qu'on appelle parfois suite arithmético-géométrique). Montrez qu'il existe un réel α tel que la suite $(w_n - \alpha)_{n\in\mathbb{N}}$ est géométrique, et en déduire une expression du terme général de $(w_n)_{n\in\mathbb{N}}$.

Exercice 6.29. Soit $f : I \to I$. On a défini le caractère chaotique de la dynamique (6.1) associée à f par le fait qu'il existe $u_0 \in I$ tel que $(f^n(u_0))_{n\in\mathbb{N}}$ est dense dans I.

1. Montrer alors que cela reste vrai pour une partie dense de $u_0 \in I$.
2. On suppose f continue. On dit que f est transitive si

$$\forall (U, V) \text{ deux ouverts non vides de } I,\ \exists p \in \mathbb{N} \text{ tel que } f^p(U) \cap V \neq \emptyset.$$

Montrer que

$$\text{la dynamique associée à } f \text{ est chaotique } \Leftrightarrow f \text{ est transitive.}$$

Exercice 6.30. Le cas $\mu = 4$ est finalement assez simple car on peut dans ce cas calculer le terme général de la suite logistique :

1. Montrer que si $x \in [0, 1]$ alors il existe un unique $\theta \in [0, \frac{\pi}{2}]$ tel que $x = \sin^2(\theta)$.
2. Montrer que pour tout $n \in \mathbb{N}$, $f^n(x) = \sin^2(2^n\theta)$.
3. En déduire que la dynamique est chaotique pour $\mu = 4$ (voir [Per, Théorème 4.13]).

Exercice 6.31. Il peut être commode de considérer des fonctions convexes à valeurs dans $\mathbb{R} \cup \{+\infty\}$; cela ne change rien aux définitions. Mais ça permet d'énoncer des résultats du type : montrer qu'un sup de fonctions convexes est convexe (c'est-à-dire si $(f_i)_{i\in I}$ est une famille de fonctions convexes, alors la fonction $\sup_{i\in I} f_i(x)$ qui est définie à valeurs dans $\mathbb{R} \cup \{+\infty\}$ est aussi convexe). On invite le lecteur à prouver ce résultat.

Exercice 6.32. Soit $f : I \to \mathbb{R}$ convexe. Montrer que pour tout $n \in \mathbb{N}^*$, tout $(x_i)_{i\in[\![1,n]\!]} \in I^n$ et tout $(\lambda_i)_{i\in[\![1,n]\!]} \in [0, 1]^n$ tels que $\sum_{i=1}^{n} \lambda_i = 1$,

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i).$$

Exercice 6.33. 1. Montrer que la fonction ln est concave sur $]0, +\infty[$.

2. En déduire l'inégalité arithmético-géométrique : pour tout $n \in \mathbb{N}^*$ et tout $(x_i)_{i\in[\![1,n]\!]} \in \mathbb{R}_+^n$,

$$\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} \leq \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right).$$

Exercice 6.34 (Inégalité de Hölder et de Minkowski, cas $\mathbb{R}^n$)**.** Soit $p > 1$.

1. Montrer qu'il existe un unique $p' > 1$ tel que $\dfrac{1}{p} + \dfrac{1}{p'} = 1$. On l'appelle exposant conjugué de p.
2. Montrer que si $(x, y) \in \mathbb{R}_+$, alors

$$x^{\frac{1}{p}} y^{\frac{1}{p'}} \leq \frac{x}{p} + \frac{y}{p'}.$$

3. (Inégalité de Hölder) En déduire que si $n \in \mathbb{N}^*$, $(x_i)_{i\in[\![1,n]\!]} \in \mathbb{R}^n$ et $(y_i)_{i\in[\![1,n]\!]} \in \mathbb{R}^n$, alors [162]

$$\sum_{i=1}^{n} |x_i y_i| \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |y_i|^{p'}\right)^{\frac{1}{p'}}.$$

Remarquez que le cas $p = 2$ est l'inégalité de Cauchy-Schwarz.

4. (Inégalité de Minkowski) Montrer que [163]

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{\frac{1}{p}}.$$

Remarquez que l'inégalité reste vraie si $p = 1$, et elle est alors très simple.

Exercice 6.35. Il n'est pas anodin qu'on ait choisi un intervalle ouvert. Exhibez en effet une fonction convexe sur $[a, b]$ mais qui n'est pas continue en a et/ou en b.

Exercice 6.36 (*)**.** Soit I un intervalle ouvert de $\mathbb{R}$. On note

$$\mathrm{Aff}(I) := \{\text{fonctions affines sur } I\}.$$

1. Montrer qu'une fonction $\varphi : I \to \mathbb{R}$ est convexe si et seulement si

$$\forall x \in I, \quad \varphi(x) = \sup\Big\{h(x), h \in \mathrm{Aff}(I), h \leq \varphi\Big\}.$$

2. (Inégalité de Jensen) Soit $\varphi : I \to \mathbb{R}$ convexe, μ une mesure de probabilité, et $f \in L^1(I, \mu)$. Montrer

$$\varphi\left(\int_I f d\mu\right) \leq \int_I \varphi \circ f d\mu.$$

---

7 Extraits d'annales

Exercice 7.1 (2020, Partie IV)**.** On définit sur $]0, +\infty[$ la fonction ρ par

$$\rho(x) = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor$$

où $\lfloor \cdot \rfloor$ dénote la fonction partie entière.

1. Soit n un entier strictement positif. Déterminer l'expression de ρ sur l'intervalle $\left]\frac{1}{n+1}, \frac{1}{n}\right]$. Préciser en particulier la valeur de $\rho(1/n)$. Déterminer également ρ sur $]1, +\infty[$.
2. Représenter la fonction ρ sur l'intervalle $[1/4, 3]$.
3. Déterminer le domaine de continuité de ρ sur $]0, +\infty[$, montrer que ρ est bornée et déterminer l'image par ρ de l'intervalle $]0, +\infty[$.

---

162. Indication : on pourra appliquer l'inégalité précédente à

$$x = \frac{|x_i|^p}{\sum_{i=1}^n |x_j|^p}, \quad \text{et} \quad y = \frac{|y_i|^{p'}}{\sum_{i=1}^n |y_j|^{p'}},$$

puis sommer en i.

163. On pourra écrire

$$|x_i + y_i|^p \leq |x_i|.|x_i + y_i|^{p-1} + |y_i|.|x_i + y_i|^{p-1}$$

et appliquer l'inégalité de Hölder à chaque morceau du membre de droite.

Exercice 7.2 (2022, Partie I)**.** En distinguant les cas $0 < \alpha < 1$, $\alpha = 1$, $1 < \alpha < 2$, $\alpha = 2$ et $\alpha > 2$, tracer l'allure du graphe de la fonction définie sur $]0, 1[$ par $t \mapsto (-\ln(t))^{\alpha-1}$.

Exercice 7.3 (2022, Partie II)**.** Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction **croissante** satisfaisant la propriété

$$\forall x, y \in \mathbb{R}, \quad f(x + y) = f(x) + f(y).$$

Démontrer qu'il existe un réel $\alpha \geq 0$ tel que $f(x) = \alpha x$, pour tout réel x.

Exercice 7.4 (2013, Partie I)**.** On considère $(u_n)_{n\in\mathbb{N}}$ une suite décroissante de réels strictement positifs telle que la série de terme général u_n converge, avec $u_0 = 1$.

1. Montrer que la suite (nu_n) converge vers 0.

Soit s un entier positif non nul, on note $E_s = \left\{n \in \mathbb{N} \mid u_n \geq \dfrac{1}{s}\right\}$.

2. Montrer que E_s est fini, que son cardinal K_s croît vers l'infini et que $E_s = \{0, \ldots, K_s - 1\}$.
3. Etablir que $\dfrac{K_s - 1}{2s} \leq \dfrac{1}{K_s} \displaystyle\sum_{n\in E_s} nu_n$.
4. Conclure que $\displaystyle\lim_{s\to+\infty} \frac{K_s}{s} = 0$.

Exercice 7.5 (Docteur 2020)**.** Soit $(u_n)_{n\in\mathbb{N}}$ une suite de nombres réels. On pose pour tout $n \in \mathbb{N}$,

$$v_n = \frac{u_0 + 2u_1 + \cdots + 2^n u_n}{1 + 2 + \cdots + 2^n}.$$

1. Montrer que si $(u_n)_{n\in\mathbb{N}}$ converge vers $\ell \in \mathbb{R}$, alors $(v_n)_{n\in\mathbb{N}}$ converge aussi vers ℓ.

   (Indication : le candidat pourra, s'il le juge utile, commencer par traiter le cas $\ell = 0$).
2. Dans cette question, on suppose que $u_n \underset{n\to+\infty}{\sim} n$. Déterminer un équivalent simple de v_n lorsque n tend vers $+\infty$.
3. Montrer que si la série de terme général u_n est absolument convergente, alors la série de terme général v_n l'est aussi.
4. Dans cette question, on considère les suites de fonctions $(U_n)_{n\in\mathbb{N}}$ et $(V_n)_{n\in\mathbb{N}}$ définies sur $]-1, 1[$ par

   $$U_n(x) = x^n \qquad \text{et} \qquad V_n(x) = \frac{U_0(x) + 2U_1(x) + \cdots + 2^n U_n(x)}{1 + 2 + \cdots + 2^n}.$$

   La série de fonctions $\sum V_n$ converge-t-elle simplement sur $]-1, 1[$? Uniformément sur $]-1, 1[$? Uniformément sur tout segment $[-a, a]$ tel que $0 < a < 1$?

Exercice 7.6 (2021, Partie I)**.** Soient $n \in \mathbb{N}^*$ et $h_n : x \in \mathbb{R} \mapsto x + 1/n$.

1. Montrer que la suite $(h_n)_{n\in\mathbb{N}^*}$ converge uniformément sur $\mathbb{R}$ vers une fonction h que l'on précisera.
2. Montrer que la suite $(h_n^2)_{n\in\mathbb{N}^*}$ converge simplement sur $\mathbb{R}$ vers h^2, mais que la convergence n'est pas uniforme sur $\mathbb{R}$.
3. Soient $(f_n)_{n\in\mathbb{N}}$ et $(g_n)_{n\in\mathbb{N}}$ deux suites de $C_b^0(\mathbb{R}, \mathbb{C})$ (les fonctions continues et bornées sur $\mathbb{R}$) qui convergent uniformément sur $\mathbb{R}$, vers f et g respectivement. Montrer que la suite $(f_n g_n)_{n\in\mathbb{N}}$ converge uniformément vers fg sur $\mathbb{R}$.

Exercice 7.7 (Docteur 2018)**.**

1. Soit I un intervalle de $\mathbb{R}$, et soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions continues de I sur $\mathbb{R}$. On suppose que la suite (f_n) converge uniformément vers une fonction $f : I \to \mathbb{R}$. Montrer que f est continue sur I.

2. Donner un exemple d'intervalle I et de suite $(f_n)_{n\in\mathbb{N}}$ de fonctions continues sur I qui converge simplement vers une fonction f, avec f non-continue sur I.
3. Soient I un intervalle de $\mathbb{R}$, $g : I \to \mathbb{R}_+$ une fonction intégrable sur tout compact de I et $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions de $C^1(I, \mathbb{R})$ telle que $|f_n'(t)| \leq g(t)$ pour tout $n \in \mathbb{N}$ et tout $t \in I$. On suppose de plus que la suite (f_n) converge simplement vers une fonction f. Montrer que f est continue sur I.

Exercice 7.8 (Docteur 2017, Problème Partie I)**.** Soient α un élément de $]1, +\infty[$, f une fonction continue de $\mathbb{R}$ dans $\mathbb{C}$. On suppose qu'il existe $C \in \mathbb{R}_+^*$ tel que

$$\forall x \in \mathbb{R}, \quad |f(x)| \leq \frac{C}{(1+|x|)^\alpha}.$$

Montrer qu'en posant

$$\forall x \in \mathbb{R}, \quad F(x) = \sum_{k=-\infty}^{+\infty} f(x+k),$$

on définit une fonction continue et 1-périodique de $\mathbb{R}$ dans $\mathbb{C}$.

Exercice 7.9 (2022, Partie I)**.** On rappelle que la série de terme général $(n^{-\sigma})_{n\in\mathbb{N}^*}$ est convergente, pour tout réel $\sigma > 1$. On introduit alors la fonction ζ définie pour $\sigma > 1$ par

$$\zeta(\sigma) = \sum_{n=1}^{\infty} n^{-\sigma}.$$

1. Montrer que ζ est à valeurs dans $\mathbb{R}_+^*$, de classe C^1 sur $]1, +\infty[$. Montrer que $-\zeta'(\sigma) = \sum_{n=1}^{\infty} n^{-\sigma} \log n$, pour tout réel $\sigma > 1$.
2. Par comparaison série-intégrale, montrer que $\lim_{\sigma\to 0^+} \sigma\zeta(1+\sigma) = 1$.

Exercice 7.10 (2019, Partie III)**.** Soit $R > 0$ fixé. Pour tout $r \in [0, R[$ et tout $t \in \mathbb{R}$, on pose

$$P_R(r,t) = \sum_{n=-\infty}^{\infty} \left(\frac{r}{R}\right)^{|n|} e^{int}.$$

1. En utilisant un théorème de régularité sous le signe somme, justifier que $t \mapsto P_R(r,t)$ est de classe C^∞ sur $\mathbb{R}$.
2. Soit $r \in [0, R[$ et $t \in \mathbb{R}$.
   (a) Montrer que $P_R(r,t) = 1 + 2\sum_{n=1}^{+\infty} \left(\frac{r}{R}\right)^n \cos(nt) = \mathrm{Re}\left(\frac{R + re^{it}}{R - re^{it}}\right) = \frac{R^2 - r^2}{R^2 - 2rR\cos(t) + r^2}$.
   (b) En déduire que $t \mapsto P_R(r,t)$ est $2\pi-$périodique, uniformément continue, positive et paire sur $\mathbb{R}$.

Exercice 7.11 (Docteur 2021, Problème Partie I)**.** Soit $\varphi : \mathbb{R} \to \mathbb{R}$ définie par :

$$\varphi : x \mapsto \sum_{j=0}^{\infty} \frac{1}{2^j} \sin(2^{j+1}\pi x).$$

1. Justifier que φ est bien définie et continue.
2. Montrer que φ n'est pas dérivable *[On pourra étudier* $2^k\varphi(2^{-k})$ *pour* $k \in \mathbb{N}$*]*

Chapitre II

Analyse à une variable complexe

L'analyse complexe au niveau de l'agrégation, se découpe en deux mondes : les séries entières, dont vous trouverez toutes les informations dans les livres de L2 et de CPGE, et les fonctions holomorphes, qui sont clairement un sujet du niveau L3. On trouve de nombreux livres sur les fonctions holomorphes, cependant une difficulté majeure pour la préparation de ce chapitre à l'agrégation, est que ces livres vont souvent un peu trop loin par rapport à ce qui est attendu au concours [1]. Il y a en effet de nombreuses façons de présenter les outils et les théorèmes majeurs sur les fonctions holomorphes, et certains livres préfèrent débuter par des préliminaires topologiques assez poussés, ce qui permet d'avoir des outils performants pour la suite. Mais certains de ces outils sont particulièrement délicats et/ou hors programme, et on peut comprendre qu'ils découragent les étudiants. On peut citer [AM04], qui est très fourni et très bon, mais qui débute par 60 pages sur les intégrales curvilignes et les formes différentielles, et qui s'en sert pour présenter la formule de Cauchy, alors que ce point de vue n'est pas vraiment celui du programme de l'agrégation [2]. Le classique [Car61] utilise également le language des formes différentielles.

J'ai beaucoup utilisé ici [Tau06], qui à mon avis a l'avantage d'aller dans un ordre de difficulté croissant et qui convient bien à l'agrégation si on veut s'initier en douceur aux fonctions holomorphes. C'est ce qu'on essaiera de faire ici également. Les plus téméraires apprécieront le style de [Rud98] sur ces questions. Comme tout ce qui concerne le programme de L3, on peut également citer [Mar09]. Enfin, j'ai aussi beaucoup utilisé le polycopié de V. Minerbe [Min], très clair. Evidemment, tout ceci n'a rien d'exhaustif, on laisse chaque étudiant trouver sa source favorite. On m'a également conseillé [QQ17].

Notons qu'il est très fréquent de rencontrer aux épreuves écrites de l'agrégation des questions, voire des parties dédiées aux séries entières et/ou aux fonctions holomorphes. Il est donc indispensable de connaître le “strict” minimum sur ces sujets. Pour l'oral, deux leçons seulement sont au cœur du présent chapitre (**243-245**), néanmoins les notions abordées ici peuvent illustrer la quasi-totalité des leçons du programme (en allant de la théorie des séries à l'analyse fonctionnelle, en passant par l'étude de fonctions classiques ou les probabilités...), et donc l'investissement des étudiants sur ce chapitre sera assurément payant.

1 Corps des nombres complexes

1.1 Définition

Les éléments qui suivent sont tirés de [God98, II.1] : les nombres complexes ont été inventés par les italiens à la fin du XVI^e siècle car ils ont découvert des formules qui, bien qu'elles fassent apparaître des racines carrées de nombres potentiellement négatifs, fournissaient effectivement des racines réelles

1. C'est peut-être d'ailleurs une des raisons qui amènent les étudiants à redouter (un peu trop) ce chapitre, d'après mon expérience.

2. Le livre reste une excellente source pour l'agrégation, énormément de sujets sont traités, incluant des éléments d'analyse fonctionnelle ou de fonctions spéciales qui sont tout à fait dans l'esprit du concours ; plusieurs sujets sont traités avec différentes approches, dont certaines bien sûr ne nécessitent pas un investissement particulier sur les préliminaires évoqués. Il faut seulement prendre un peu de temps à se familiariser avec le livre, et savoir ce qu'on souhaite en tirer. Chaque étudiant se fera son opinion.

de polynômes du 3ième degré. Ils ont donc considéré les nombres de la forme $a+b\sqrt{-1}$ où $(a,b)\in\mathbb{R}^2$ et en utilisant les règles de calcul habituelles. Euler introduit plus tard la notation i pour ce fameux $\sqrt{-1}$.

Si on note $\mathbb{C}$ l'ensemble des éléments ainsi obtenus, on obtient la caractérisation suivante :

— $\mathbb{C}$ est un corps,

— $\mathbb{R}$ est un sous-corps de $\mathbb{C}$,

— $i^2=-1$,

— tout $z\in\mathbb{C}$ peut s'écrire $z=x+iy$ avec $(x,y)\in\mathbb{R}^2$.

C'est seulement en 1835 que W.R. Hamilton donne une construction rigoureuse du corps $\mathbb{C}$ des nombres complexes. Comparativement à la difficulté de construire $\mathbb{R}$ évoquée au Chapitre I, la construction de $\mathbb{C}$ (une fois qu'on a construit $\mathbb{R}$ évidemment) est très simple :

Définition-Proposition 1.1. On peut définir $\mathbb{C}$ comme l'ensemble des couples $(a,b)\in\mathbb{R}^2$ muni des lois

$$\forall (a,b),(c,d)\in\mathbb{C}^2,\quad (a,b)+(c,d)=(a+c,b+d),\quad (a,b).(c,d)=(ac-bd,ad+bc).$$

Alors $\mathbb{C}$ est bien un corps. On peut définir $i=(0,1)$ et constater que $i^2=(-1,0)$. On peut aussi assimiler $\mathbb{R}$ à un sous-corps de $\mathbb{C}$ via l'injection [3] $x\in\mathbb{R}\mapsto(x,0)\in\mathbb{C}$, et observer que

$$\forall (a,b)\in\mathbb{C},\quad (a,b)=(a,0)+(0,b)=(a,0)+i.(b,0),\quad \text{qu'on écrira plus simplement}\quad a+ib.$$

Démonstration. Le seul élément pas complètement évident consiste à trouver l'inverse d'un élément $a+ib$ où $(a,b)\in\mathbb{R}^2\setminus\{(0,0)\}$. Il s'agit de $\frac{a}{a^2+b^2}-\frac{b}{a^2+b^2}i$. □

On voit donc directement apparaître le lien étroit entre le corps des complexes et la géométrie planaire, puisque tout nombre complexe $a+ib$ peut être représenté par le point $(a,b)\in\mathbb{R}^2$. On définit également

$$\forall z=a+ib\in\mathbb{C},\quad \mathrm{Re}(z)=a,\quad \mathrm{Im}(z)=b,\quad |z|=\sqrt{a^2+b^2},\quad \overline{z}=a-ib$$

les parties réelles et imaginaires, le module et le conjugué de z. On vérifie facilement que

$$\forall (z,z')\in\mathbb{C}^2,\quad |z|^2=z.\overline{z},\quad |z.z'|=|z|.|z'|,\quad \overline{z.z'}=\overline{z}.\overline{z'}\ldots$$

La liste de formule se poursuit, on ne la détaille pas de façon exhaustive en considérant que le lecteur est familier avec l'usage des nombres complexes. Rappelons quand même

$$\forall (z,z')\in\mathbb{C}^2,\quad |z+z'|^2=|z|^2+2\mathrm{Re}(z.\overline{z'})+|z'|^2$$

qui a son importance du point de vue de la géométrie et de la théorie des espaces de Hilbert, ainsi que *les* inégalités triangulaires

$$\forall (z,z')\in\mathbb{C}^2,\quad \Big||z|-|z'|\Big|\le|z+z'|\le|z|+|z'|.$$

Remarque 1.2. Le module sur $\mathbb{C}$ étend la notion de valeur absolue qui était définie sur $\mathbb{R}$. C'est également ce qui donne à $\mathbb{C}$ une structure de $\mathbb{R}$-espace vectoriel normé.

On verra que la construction de l'exponentielle complexe au paragraphe 2.4 permet d'écrire tout complexe sous forme trigonométrique/exponentielle, avec les avantages auxquels vous êtes déjà certainement familiers.

3. Rigoureusement, la notation $\mathbb{R}\subset\mathbb{C}$ est en fait un abus, fréquent en mathématiques, qui cache en fait une injection parfois dite "canonique", même si ce mot n'est pas très bien déterminé en général.

1.2 Reprise du Chapitre I dans $\mathbb{C}$

C'est un exercice particulièrement important de savoir reconnaître quand une propriété est réellement restreinte aux nombres réels, et quand elle peut s'étendre sans inquiétude aux nombres complexes. Il y a plusieurs différences notables entre les corps $\mathbb{R}$ et $\mathbb{C}$: d'un point de vue algébrique par exemple, le second est algébriquement clos [4], alors que le premier ne l'est pas. Mais d'un point de vue analytique, la principale différence réside dans le fait que $\mathbb{C}$ n'est pas ordonné, ou pour être plus précis, qu'il n'y a pas de relation d'ordre totale sur $\mathbb{C}$ qui soit compatible avec la structure de corps de $\mathbb{C}$.

Donc, si vous souhaitez énerver un correcteur ou les membres du jury, il vous suffira de faire un raisonnement avec le signe $\leq$ entre deux nombres complexes [5] ; résultat garanti.

Plus sérieusement, nous faisons une liste non exhaustive des énoncés/paragraphes du Chapitre I et de leurs éventuelles adaptations. Nous ne détaillons pas grand chose, d'une part parce qu'il vaut mieux que le lecteur y mette sa propre réflexion, d'autre part parce que l'on traitera dans les Chapitre III et VII des cadres plus généraux qui apporteront également des réponses.

- Pour les ensembles : bien sûr, il n'y a plus de notion de majorant ou de borne supérieure pour un ensemble $A \subset \mathbb{C}$. Néanmoins, on peut dire d'un ensemble $A \subset \mathbb{C}$ qu'il est borné si

$$\exists M \in \mathbb{R}, \quad \forall x \in A, \quad |z| \leq M,$$

  voir aussi l'exercice 2.15 du Chapitre I. Plus généralement, de nombreuses notions et énoncés du Chapitre I resterons valables dans le cadre de $\mathbb{C}$ en remplaçant la valeur absolue par le module.
- En ce qui concerne les suites $(u_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$:
  - On peut définir la convergence dans $\mathbb{C}$ exactement comme on l'a fait dans $\mathbb{R}$. De même pour les suites de Cauchy et alors $\mathbb{C}$ est encore un espace complet, voir le paragraphe 1.4 du Chapitre III pour plus de détails.
  - Par contre, il n'y a plus de sens à parler de suites qui tendent vers $\pm\infty$, mais il n'est pas interdit de considérer la suite réelle $(|u_n|)_{n\in\mathbb{N}}$.
  - Les relations de comparaison/notations de Landau $o, O, \sim$ restent valables car tout a été défini avec $|\cdot|$.
  - Il va sans dire que tout ce qui concerne les suites monotones ou le signe $\leq$, comme le théorème 2.8 ou la proposition 2.10 au Chapitre I, n'ont pas d'équivalent pour des suites à valeurs complexes.
  - Il n'y a plus de notion de $\liminf$ ou $\limsup$ pour des suites à valeurs complexes.
  - Le cas du théorème de Bolzano-Weierstrass est intéressant : il reste valable, mais nécessite un peu de réflexion. On reviendra sur ce point au paragraphe 1.5 du Chapitre III.
- Pour l'étude des séries dont le terme général est à valeurs dans $\mathbb{C}$,
  - le fait que le terme général converge vers 0 reste une condition nécessaire de convergence de la série,
  - le fait que les séries absolument convergentes sont convergentes (Proposition 2.39) reste valable. La seconde preuve qui repose sur la complétude est inchangée, elle repose simplement sur la complétude de $\mathbb{C}$. La première preuve peut également être adaptée, en décomposant

$$u_n = \mathrm{Re}(u_n)_+ - \mathrm{Re}(u_n)_+ + i.(\mathrm{Im}(u_n)_+ - \mathrm{Im}(u_n)_-).$$

- En ce qui concerne la théorie des fonctions, on peut vouloir prendre un ensemble de départ et/ou d'arrivée qui est un sous-ensemble de $\mathbb{C}$ plutôt qu'un sous-ensemble de $\mathbb{R}$:

4. Ce dont on donnera une preuve dans ce chapitre, voir le théorème 3.37.
5. Ça marchera aussi si ce sont des éléments d'un $\mathbb{R}$-espace vectoriel de dimension supérieure ou égale à 2, mais vu que les nombres complexes font partie du programme de terminale, l'effet sera probablement décuplé dans cette situation particulière.

  - La définition de la continuité ne pose pas de problème, que ce soit l'espace de départ ou d'arrivée qui soit changé. Mais dans le cas où c'est l'espace de départ qui change, il faudra remplacer $]a-\alpha, a+\alpha[$ par
$$D(a,\alpha) := \Big\{z \in \mathbb{C}, |z-a| < \alpha\Big\} \tag{1.1}$$
qui est un disque de centre a et de rayon α.
  - Les résultats du paragraphe 3.2.2 au Chapitre I n'ont pas d'adaptation évidente. On verra une version générale du théorème des valeurs intermédiaires quand on abordera la connexité au paragraphe 1.6 du Chapitre III.
  - La notion de continuité uniforme et le théorème de Heine ne posent pas de problème quand on travaille dans $\mathbb{C}$ plutôt que $\mathbb{R}$. L'adaptation des notions et des preuves se résume à considérer (1.1) à la place des intervalles $]a-\alpha, a+\alpha[$. Voir aussi le paragraphe 1.3 au Chapitre III pour une version plus générale.
  - Le cas de la dérivabilité est à prendre avec attention. Quand l'espace de départ est inchangé mais l'espace d'arrivée devient $\mathbb{C}$, la définition et les règles de calcul fonctionnent sans problème, mais la preuve du théorème de Rolle n'est plus valable [6] (car repose sur la notion d'extremum de f), et il se trouve que son énoncé, ainsi que ses corollaires, sont faux [7], à l'exception du corollaire 3.43 qui sera abordé au Chapitre VII. Quand l'espace de départ devient un sous-ensemble de $\mathbb{C}$, on peut définir une notion de dérivabilité avec la même formule (3.8) qu'au Chapitre I, mais l'analyse des fonctions "dérivables" sera radicalement différente. C'est en fait le rôle principal du paragraphe 3 ci-après que de comprendre cette notion. Remarquons que si on regarde $\mathbb{C}$ comme le $\mathbb{R}$-espace vectoriel $\mathbb{R}^2$ (c'est-à-dire en ne prenant pas en compte la division qui donne à $\mathbb{C}$ sa structure de corps) et qu'on veut comprendre une notion de "dérivée" qui s'adapte aux fonctions définies sur un sous-ensemble de $\mathbb{R}^2$, alors on peut également définir la notion de différentiabilité (différente dans ce cas de la notion de dérivabilité), ce qui sera abordé au Chapitre VII, voir aussi le paragraphe 3.2.

- En ce qui concerne la notion de suites et de séries de fonctions (définies sur un sous-ensemble de $\mathbb{C}$ au lieu de $\mathbb{R}$ et/ou à valeurs dans $\mathbb{C}$), tout reste à peu près la même chose, exception faite de la proposition 4.13 quand l'espace de départ est dans $\mathbb{C}$. En effet, on a déjà dit que la notion de dérivation se comportait très différemment, on aura donc l'occasion de revenir sur ce point, voir le notamment le Théorème 3.47.
- Pour la théorie de l'intégration, à nouveau les choses sont différentes si on change l'espace de départ ou d'arrivée.
  - si on veut intégrer des fonctions $f : I \to \mathbb{C}$ au sens de Riemann, comme on l'a d'ailleurs évoqué à la remarque 5.9, il n'y a pas de difficulté. On peut soit reprendre la construction de l'intégrale de Riemann dans ce cadre (qui repose sur le théorème de Heine dont on a déjà dit qu'il restait valable), soit voir $\mathbb{C}$ comme un $\mathbb{R}$-espace vectoriel de dimension finie, et définir
$$\int_I f(x)dx = \int_I \mathrm{Re}(f(x))dx + i\int_I \mathrm{Im}(f(x))dx,$$
et utiliser (plutôt que généraliser) la construction que nous avons faite pour des fonctions à valeurs réelles. Attention tout de même, dans le cas où I n'est pas un segment, il faut

6. Attention ; ce que je suis en train d'expliquer est une source d'erreur fréquente et grave, je détaille donc un peu. Je veux dire ici que la preuve ne semble pas pouvoir fonctionner, car fondamentalement on utilise que l'espace d'arrivée est ordonné. Cela ne peut en aucun cas constituer une démontration de la non-véracité de l'énoncé. Cela peut éventuellement vous suggérer que l'énoncé est faux, mais vous ne pourrez échapper à produire un contre-exemple pour être convaincu. Faites donc la différence entre "l'énoncé n'est plus valable" et "la preuve n'est plus valable" qui sont deux choses fondamentalement différentes. Si le jury vous demande "est-ce que "ça" marche encore si on enlève cette hypothèse ?" demandez-vous si votre preuve persiste et/ou s'il veut savoir si votre résultat reste vrai. Si ça n'est pas clair, n'hésitez pas à lui poser la question. L'erreur fréquente consiste à se concentrer sur sa démonstration (qui ne fonctionne peut-être plus mais à ne pas comprendre qu'on nous demande de fournir un contre-exemple.

7. Trouvez un contre-exemple !

demander que les intégrales de $\mathrm{Re}(f)$ et $\mathrm{Im}(f)$ soient convergentes pour avoir le droit d'écrire une telle définition.

  - si on veut intégrer des fonctions $f : U \to \mathbb{C}$ où U est un sous-ensemble de $\mathbb{C}$, alors on est dans le cadre des intégrales multiples, et $\mathbb{C}$ est vu comme $\mathbb{R}^2$, sa structure de corps n'apportant rien de spécial. On abordera cela au Chapitre IV.

- La notion de suites récurrentes $\forall n \in \mathbb{N}, u_{n+1} = f(u_n)$ où $f : U \to U$ et $U \subset \mathbb{C}$ reste tout à fait pertinente, mais on n'a plus de notion de monotonie, ce qui rend l'étude plus difficile. C'est d'ailleurs par ce biais que l'on rencontre bon nombre d'ensembles fractals, comme l'ensemble de Julia ou de Mandelbrot. On n'abordera pas ces notions dans ce document.
- La notion de convexité d'une fonction nécessite que l'ensemble d'arrivée soit $\mathbb{R}$, et que l'ensemble de départ soit convexe. On peut donc étudier les fonctions convexes $f : U \to \mathbb{R}$ où $U \subset \mathbb{C}$ est un ensemble convexe, mais la structure de corps n'apporte rien ; on renvoit donc au paragraphe sur les fonctions convexes définies sur un convexe de $\mathbb{R}^n$ avec $n \geq 1$.

Conclusion : finalement, la plupart des outils du Chapitre I s'adaptent au cadre de $\mathbb{C}$ et plus généralement de $\mathbb{R}^n$ avec $n \in \mathbb{N}^*$. Mais il faut faire attention au théorème des valeurs intermédiaires et à l'égalité des accroissements finis. Enfin on reviendra sur la notion de dérivation basée sur la division complexe, qui est très différente de la dérivation "réelle".

1.3 Préliminaire topologique

Comme au paragraphe 1.2, on donne quelques éléments de topologie dans le cas de l'espace $\mathbb{C}$. Ça se complique par rapport au cas de $\mathbb{R}$ principalement parce que dans $\mathbb{R}$, les ensembles connexes et convexes sont les mêmes (les intervalles) ; ici les choses sont plus riches. Si $U \subset \mathbb{C}$:

- on définit un point adhérent à un ensemble et l'adhérence d'un ensemble exactement comme dans le cas de $\mathbb{R}$. Si un ensemble est égal à son adhérence, on dit qu'il est fermé (on devrait dire "fermé dans $\mathbb{C}$" : voir la note 9).
- pour la notion de point intérieur, on doit remplacer les intervalles par des disques : on dit qu'un point $z \in U$ est intérieur à U s'il existe $\alpha > 0$ tel que $D(z, \alpha) \subset U$ (voir (1.1) pour la définition de $D(z, \alpha)$). On dit que U est ouvert si tous ses points sont intérieurs à U, sinon on note $\mathring{U}$ l'ensemble des points intérieurs à U.
- on appelle voisinage de z un ensemble qui contient $D(z, \alpha)$ pour un certain $\alpha > 0$.
- on définit $\partial U = \overline{U} \setminus \mathring{U}$ la frontière de U.
- on dit que U est compact si de toute suite d'éléments de U, on peut extraire une sous-suite qui converge dans U. On verra au paragraphe 1.5 du Chapitre III que comme $\mathbb{C}$ est un espace vectoriel de dimension finie, un ensemble est compact si et seulement si il est fermé et borné.
- on dit que $z \in \mathbb{C}$ est un point d'accumulation de U si pour tout $\varepsilon > 0$, il existe $z' \in U$ tel que $|z - z'| < \varepsilon$ et $z' \neq z$. De façon équivalente, cela signifie qu'il existe une suite $(z_p)_{p \in \mathbb{N}}$ d'éléments de $U \setminus \{z\}$ qui converge vers z [8].
- l'ensemble U est dit convexe si

$$\forall (z_1, z_2) \in U^2, \ \ \forall t \in [0, 1], \ \ (1 - t)z_1 + tz_2 \in U.$$

- on peut toujours considérer la notion de segment $[z_1, z_2]$ si $(z_1, z_2) \in \mathbb{C}^2$, mais on doit la définir sans utiliser d'ordre comme on a pu le faire dans $\mathbb{R}$: il s'agit de

$$[z_1, z_2] = \big\{(1 - t)z_1 + tz_2, \ \ t \in [0, 1]\big\},$$

ce qui correspond à l'ensemble des combinaisons convexes de z_1 et z_2. D'ailleurs, la définition précédente devient : U est convexe si pour tout $(z_1, z_2) \in U^2$, $[z_1, z_2] \subset U$.

8. Un point d'accumulation de U est donc un point adhérent de U, mais la réciproque est fausse. Plus précisément, les points adhérents à U sont soit des points d'accumulation de U, soit des points isolés de U ($z_0 \in \mathbb{C}$ est un point isolé de U s'il existe $r_0 > 0$ tel que $D(z_0, r_0) \cap U = \{z_0\}$).

— Un ensemble U est dit connexe si pour tout $A \subset U$ tel que A est ouvert et fermé dans U [9] , A est vide ou $A = U$.

— Un ensemble U est dit connexe par arcs si pour tout $(z, z') \in U^2$, il existe un chemin continu entre z et z' qui est tracé dans U, c'est-à-dire qu'il existe $\gamma : [0,1] \to \mathbb{C}$ continue et telle que $\gamma([0,1]) \subset U$, $\gamma(0) = z$ et $\gamma(1) = z'$.

La notion de connexité est un peu difficile à manipuler, et les détails seront traités au Chapitre III. Par contre, elle est intuitivement très simple : un sous-ensemble de $\mathbb{C}$ est connexe s'il est en un seul morceau. Le lecteur ne doit donc pas être effrayé par cette notion dans la lecture de ce chapitre. On admet aussi pour le moment deux résultats standards sur la connexité, qui vont permettre de contourner la plupart des difficultés liées à cette notion, et permettront au lecteur de prouver qu'un ensemble est connexe :

Proposition 1.3. *Soit $U \subset \mathbb{C}$. Alors*

1. *Si U est connexe par arcs, alors U est connexe.*
2. *Si U est ouvert et connexe, alors U est connexe par arcs, et même connexe par arcs de classe C^1 par morceaux*[10]

Voir le paragraphe 1.6 au Chapitre III pour une démonstration.

Distance entre un point et une partie, distance entre deux ensembles : Si U est un sous-ensemble de $\mathbb{C}$ et $z \in \mathbb{C}$, on pose

$$d(z, U) = \inf\{|z - z'|, \ \ z' \in U\}.$$

Enfin, si U_1 et U_2 sont deux ensembles non vides, on définit

$$d(U_1, U_2) = \inf \Big\{ |z_1 - z_2|, \ z_1 \in U_1, z_2 \in U_2 \Big\}$$

la distance entre ces deux ensembles (faites un dessin !).

On utilisera les propriétés suivantes (voir l'exercice 1.6) :

1. pour tout $z \in \mathbb{C}$, $d(z, U) = 0 \Leftrightarrow z \in \overline{U}$,
2. si U_1 et U_2 sont disjoints, que U_1 est compact et U_2 est fermé, alors $d(U_1, U_2) > 0$.

Conclusion : toutes les notions et propriétés abordées ici sont valables dans les $\mathbb{R}$-espaces vectoriels de dimension finie, et $\mathbb{C}$ est un cas particulier.

1.4 Exercices

Exercice 1.4. On peut donner d'autres constructions du corps des nombres complexes, qui à mon avis sont pertinentes vis-à-vis du programme d'algèbre de l'agrégation :

1. Considérer l'anneau quotient $\mathbb{R}[X]/(X^2 + 1)$ et montrer qu'il s'agit d'un corps isomorphe à $\mathbb{C}$.

9. Attention, ici se cache une subtilité de topologie induite, car précédemment on a donné les définitions d'un fermé et d'un ouvert de $\mathbb{C}$, alors qu'ici on a besoin de définir fermé et ouvert de U :

— A est dit fermé de U si pour toute suite $(z_n)_{n \in \mathbb{N}} \in A^{\mathbb{N}}$ qui converge **dans** U, alors la limite de $(z_n)_{n \in \mathbb{N}}$ est en fait dans A ;

— A est dit ouvert de U si pour tout $z \in A$, il existe $\alpha > 0$ tel que $D(z, \alpha) \cap U \subset A$. Notez que c'est l'intersection avec U qui doit être dans A, et non le disque entier.

Pour l'heure, ces subtilités ne sont pas primordiales, on utilisera cette notion principalement dans la preuve du théorème 2.36.

10. Ce résultat "amélioré n'est pas plus difficile à prouver (voir la proposition 2.22 au Chapitre III), et nous l'utiliserons quand nous utiliserons l'intégration sur de chemins C^1 par morceaux.

2. Considérer l'ensemble des matrices de similitudes

$$\left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad (a,b) \in \mathbb{R}^2 \right\},$$

que l'on munit des opérations usuelles sur les matrices. Constatez que cet ensemble est stable par produit, et définit un corps isomorphe à $\mathbb{C}$.

Exercice 1.5. Soit $z_0 \in \mathbb{C}$, $r > 0$.

1. Montrer que $D(z_0, r)$ est ouvert.
2. Montrer que l'adhérence de $D(z_0, r)$ est

$$\left\{ z \in \mathbb{C}, \quad |z - z_0| \leq r \right\}.$$

Pour cette raison, on notera $\overline{D}(z_0, r)$ cet ensemble.

Exercice 1.6. 1. Soit $U \subset \mathbb{C}$, et $z \in \mathbb{C}$. Montrer que $d(z, U) = 0$ si et seulement si $z \in \overline{U}$.

2. Soit U_1, U_2 des sous-ensembles non vides de $\mathbb{C}$.
   (a) Montrez qu'il existe $(z_n)_{n\in\mathbb{N}}$ suite de U_1 et z'_n suite de U_2 tels que [11]

$$|z_n - z'_n| \xrightarrow[n\to+\infty]{} d(U_1, U_2).$$

   (b) On suppose U_1 compact et U_2 fermé, et $U_1 \cap U_2 = \emptyset$. Montrez que $d(U_1, U_2) > 0$.

---

2 Séries entières

Dans ce paragraphe, on s'intéresse aux séries de fonctions de la forme

$$z \longmapsto \sum_{n\in\mathbb{N}} a_n z^n$$

où $z \in \mathbb{C}$ et $(a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$, qui sont des outils très utiles et interviennent dans de nombreux champs mathématiques, voir le paragraphe 4.2.

2.1 Définition et rayon de convergence

Définition 2.1. Une série entière est une série de fonctions dont le terme général $(f_n)_{n\in\mathbb{N}}$ est de la forme $\forall z \in \mathbb{C},\ f_n(z) = a_n z^n$, où $(a_n)_{n\in\mathbb{N}}$ est une suite donnée de nombres complexes.

Remarque 2.2. Il est fréquent de lire dans la littérature les notations $\sum a_n z^n$ ou $\sum_{n\in\mathbb{N}} a_n z^n$ pour désigner une série entière. Je n'aime pas trop cet abus, d'une part parce qu'on écrit une série sans avoir justifié sa convergence, d'autre part parce qu'on n'écrit pas où vit z, et même qu'on ne spécifie pas qu'il s'agit d'une fonction de z. J'éviterai donc d'utiliser cette notation, en parlant de la série entière associée à la suite $(a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$. Si la série entière a pour rayon R (voir ci-dessous pour une définition), on pourra bien sûr écrire

$$z \in D(0, R) \longmapsto \sum_{n\in\mathbb{N}} a_n z^n$$

qui cette fois est bien définie [12].

11. Voir aussi l'exercice 2.18 du Chapitre I.
12. Du fait du caractère très répandu des abus évoqués ci-dessus, il ne serait pas raisonnable de vous déconseiller avec trop d'ardeur de les utiliser, mais ne perdez jamais de vue qu'il s'agit d'un abus et qu'il faut manipuler ces notations avec parcimonie et clairvoyance.

Remarque 2.3. On a considéré les fonctions $(f_n)_{n\in\mathbb{N}}$ définies sur $\mathbb{C}$, mais on peut tout à fait les définir sur $\mathbb{R}$. Dans le cas où $(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, on peut donc étudier $x \mapsto \sum_{n\in\mathbb{N}} a_n x^n$ comme une fonction réelle définie sur un sous-ensemble de $\mathbb{R}$. Néanmoins, on va voir qu'il est plus naturel et plus général d'étudier ces fonctions comme définies sur un sous-ensemble de $\mathbb{C}$ et à valeurs dans $\mathbb{C}$. Il ne sera pas difficile de restreindre a posteriori les fonctions à un sous-ensemble de $\mathbb{R}$ si besoin.

Remarque 2.4. On peut être amené à étudier des séries de fonctions dont le terme général s'écrit $\forall z \in \mathbb{C},\ f_n(z) = a_n(z-z_0)^n$ où $z_0 \in \mathbb{C}$ est fixé, qu'on peut appeler par exemple une série entière centrée en z_0. On se focalise sur le cas $z_0 = 0$, mais il sera facile d'adapter les énoncés aux autres cas, quitte à faire un changement de variable sur l'espace de départ. Ce sera particulièrement important au paragraphe 2.6.

On n'a pas encore parlé de l'ensemble des z pour lesquels la série sera convergente, et qui définira l'ensemble de définition de la série entière. Le lemme suivant permet de montrer qu'il existe un rayon $R \in [0,+\infty]$ tel qu'il y aura convergence à l'intérieur du disque de rayon R, et divergence à l'extérieur, le cas du bord du disque étant incertain et à étudier dans chaque cas :

Lemme 2.5 (Lemme d'Abel)**.** *Soit $(a_n)_{n\in\mathbb{N}}$, et $r_0 \in \mathbb{R}_+$ tel que la suite $(a_n r_0^n)_{n\in\mathbb{N}}$ est bornée. Alors pour tout $z \in \mathbb{C}$ tel que $|z| < r_0$, la série de terme général $(a_n z^n)_{n\in\mathbb{N}}$ est absolument convergente.*

Démonstration. Si $r_0 = 0$, il n'y a rien à montrer ; supposons donc $r_0 > 0$. Par hypothèse, il existe $M \in \mathbb{R}$ tel que $\forall n \in \mathbb{N}, |a_n r_0^n| \leq M$. Alors

$$\forall n \in \mathbb{N},\quad |a_n z^n| \leq M \left|\frac{z}{r_0}\right|^n$$

qui est une série convergente car $\left|\frac{z}{r_0}\right| \in [0,1[$. □

Définition-Proposition 2.6. Soit $(a_n) \in \mathbb{C}^{\mathbb{N}}$. On pose

$$R_a := \sup\Big\{ r \in \mathbb{R}_+, (a_n r^n)_{n\in\mathbb{N}} \text{ est bornée } \Big\},$$

qui existe bien dans $[0,+\infty]$ et est unique. On l'appelle rayon de convergence de la série entière associée à la suite $(a_n)_{n\in\mathbb{N}}$. On a alors

- pour tout $z \in \mathbb{C}$ tel que $|z| < R_a$, la série de terme général $(a_n z^n)_{n\in\mathbb{N}}$ est absolument convergente,
- pour tout $z \in \mathbb{C}$ tel que $|z| > R_a$, le terme général $(a_n z^n)_{n\in\mathbb{N}}$ n'est pas borné, et donc la série associée diverge grossièrement,
- pour tout $r < R_a$, la série entière $z \mapsto \sum_{n\in\mathbb{N}} a_n z^n$ est normalement convergente sur le disque fermé $\big\{z \in \mathbb{C},\ |z| \leq r\big\}$.

Démonstration.
- Soit $z \in \mathbb{C}$ tel que $|z| < R_a$.[13] Par définition, $|z|$ n'est pas un majorant de l'ensemble $\Big\{ r \in \mathbb{R}_+, (a_n r^n)_{n\in\mathbb{N}} \text{ est bornée } \Big\}$, donc il existe r_0 tel que $|z| < r_0 \leq R_a$ et tel que $(a_n r_0^n)_{n\in\mathbb{N}}$ est bornée. Alors par le lemme d'Abel 2.5, la série de terme général $(a_n z^n)_{n\in\mathbb{N}^*}$ est absolument convergente.
- Si maintenant $z \in \mathbb{C}$ est tel que $|z| > R_a$, le fait que $(a_n z^n)_{n\in\mathbb{N}}$ n'est pas bornée vient de la définition de la borne supérieure.
- Soit $r < R_a$. Alors par le premier point, la série de terme général $(a_n r^n)_{n\in\mathbb{N}}$ est absolument convergente. Or $\forall z \in \overline{D}(0,r),\quad |a_n z^n| \leq |a_n| r^n$, donc

$$\sum_{n\in\mathbb{N}} \sup_{z\in\overline{D}(0,r)} |a_n z^n| < +\infty.$$

  Il y a convergence normale, donc uniforme (par la Proposition 4.17 du Chapitre I) sur $\overline{D}(0,r)$. □

13. Attention, il faut se convaincre que cette partie de l'énoncé ne découle pas trivialement du lemme d'Abel 2.5 ; en effet la définition de R_a ne permet pas de dire que $(a_n R_a^n)_{n\in\mathbb{N}}$ est bornée, car une borne supérieure n'appartient pas forcément à l'ensemble dont elle est borne supérieure.

Ainsi l'ensemble de définition d'une série entière Δ est tel que

$$D(0, R_a) \subset \Delta \subset \overline{D}(0, R_a).$$

En les points de $\partial D(0, R_a)$, il faut une étude plus approfondie et spécifique à chaque situation pour étudier la convergence. Il est possible qu'il y ait convergence en tout point, ou en aucun point de ce cercle de convergence, et il y a des situations intermédiaires, voir aussi le paragraphe 4.1.

Remarque 2.7. Pour le dernier point de la définition/proposition 2.6, on peut également dire qu'il y a convergence normale (donc uniforme) de la série sur tout compact inclus dans $D(0, R_a)$. En effet, si K est un tel compact, on peut montrer l'existence de $r < R_a$ tel que $K \subset \overline{D}(0, r)$ [14].

2.2 Calcul de rayons de convergence

Commençons par dire qu'il n'est pas si rare que la solution la plus simple pour calculer le rayon de convergence, consiste à revenir à sa définition via le lemme 2.5 [15].

On peut aussi réutiliser les règles de d'Alembert et de Cauchy que nous avons vues au paragraphe 2.2.2 du Chapitre I au cas des séries de terme général $(a_n z^n)_{n\in\mathbb{N}}$. On obtient :

Proposition 2.8 (Règles de d'Alembert et de Cauchy)**.** *Soit $(a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$, et $\ell \in [0, +\infty]$. Si l'une des deux conditions suivantes est vérifiée :*

— *la suite $(a_n)_{n\in\mathbb{N}}$ ne s'annule pas à partir d'un certain rang et $\left|\dfrac{a_{n+1}}{a_n}\right| \xrightarrow[n\to+\infty]{} \ell$,*

— *$|a_n|^{\frac{1}{n}} \xrightarrow[n\to+\infty]{} \ell$,*

alors le rayon de convergence de la série entière de terme général $(a_n z^n)_{n\in\mathbb{N}}$ est $\frac{1}{\ell}$. [16]

Démonstration. Appliquez les exercices 2.50 et 2.51. Vous pouvez également reproduire la preuve dans ce contexte ; je vous suggère de faire les deux ! □

Regardons maintenant les opérations algébriques sur les séries entières :

Proposition 2.9. *Soient $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ deux suites complexes, et R_a, R_b les rayons de convergence des séries entières associées. Alors*

— *le rayon de convergence R_c de la série entière associée à $c_n = a_n + b_n$ est tel que :*

$$R_c \geq \min\{R_a, R_b\}, \qquad \textit{et} \qquad \left[\, R_c = \min\{R_a, R_b\} \ \textit{si} \ R_a \neq R_b \,\right]$$

et de plus on a

$$\forall z \in D(0, \min\{R_a, R_b\}), \ \sum_{n\in\mathbb{N}} (a_n + b_n) z^n = \sum_{n\in\mathbb{N}} a_n z^n + \sum_{n\in\mathbb{N}} b_n z^n. \tag{2.1}$$

— *le rayon de convergence R_d de la série entière associée à $d_n = \sum_{k=0}^{n} a_k b_{n-k}$ est tel que :*

$$R_d \geq \min\{R_a, R_b\},$$

et de plus on a

$$\forall z \in D(0, \min\{R_a, R_b\}), \ \sum_{n\in\mathbb{N}} \left(\sum_{k=0}^{n} a_k b_{n-k}\right) z^n = \left(\sum_{n\in\mathbb{N}} a_n z^n\right)\left(\sum_{n\in\mathbb{N}} b_n z^n\right).$$

Démonstration. — Etant donné $z \in D(0, \min\{R_a, R_b\})$, les séries de termes généraux $(a_n z^n)_{n\in\mathbb{N}}$ et $(b_n z^n)_{n\in\mathbb{N}}$ sont absolument convergentes. Il en est donc de même pour $((a_n + b_n) z^n)_{n\in\mathbb{N}}$, ce qui montre $R_c \geq \min\{R_a, R_b\}$. De plus (2.1) est vraie d'après les opérations élémentaires sur les limites de suites.

14. Prouvez-le ! Voir aussi l'exercice 1.6.
15. Il arrive souvent que les étudiants "foncent" sur la règle de d'Alembert ; cette dernière peut être très commode dans certaines situations, et très peu dans d'autres.
16. Il est entendu ici que $\frac{1}{0} = +\infty$ et $\frac{1}{+\infty} = 0$.

— Supposons $R_a \neq R_b$, par exemple $R_a < R_b$. Alors pour z tel que $R_a < |z| < R_b$, la série de terme général $(a_n z^n)_{n\in\mathbb{N}}$ diverge et celle de terme général $(b_n z^n)_{n\in\mathbb{N}}$ converge, donc la série de terme général $((a_n + b_n)z^n)_{n\in\mathbb{N}}$ est également divergente, d'où $R_c = R_a$.

— Quant au produit, on applique la proposition 2.45 du Chapitre I aux suites $(a_n z^n)_{n\in\mathbb{N}}$ et $(b_n z^n)_{n\in\mathbb{N}}$ quand $|z| < \min\{R_a, R_b\}$.

□

Remarque 2.10. On ne traite pas l'inverse, le quotient, ni la composée de séries entières. En effet, ces opérations sont difficiles à étudier par le calcul ; on peut trouver le cas de l'inverse dans [Gou08b, page 251], et les cas de l'inverse et de la composition dans [AF89]. Néanmoins, le paragraphe 3 permettra de traiter ces cas une fois qu'on aura vu que le caractère $\mathbb{C}$-dérivable sur un ouvert U est équivalent au caractère développable en série entière (sur ce même ouvert), voir le Théorème 3.28 : il est en effet facile de voir sous quelles conditions un quotient ou une composée de fonctions $\mathbb{C}$-dérivables sont $\mathbb{C}$-dérivables, voir le paragraphe 3.3 et aussi [Pom94, 25.1.4].

2.3 Propriétés

2.3.1 Dérivation

Définition-Proposition 2.11. Soit $(a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$, et $R_a \in [0, +\infty]$ le rayon de la série entière associée. On appelle série entière dérivée la série entière de terme général $((n+1)a_{n+1}z^n)_{n\in\mathbb{N}}$; elle a le même rayon de convergence que celle associée à $(a_n)_{n\in\mathbb{N}}$.

Démonstration. Notons R_b le rayon de la série entière associée à $((n+1)a_n z^n)_{n\in\mathbb{N}}$. Comme $\forall n \in \mathbb{N}, |a_{n+1}| \leq (n+1)|a_{n+1}|$, $R_b \leq R_a$.

Pour montrer l'inégalité inverse, on constate d'une part que le cas $R_a = 0$ est trivial, et donc on suppose $R_a > 0$ et on se donne $z \in D(0, R_a)$. Il existe r' tel que $|z| < r' < R_a$ [17]. Alors

$$\forall n \in \mathbb{N}, \quad |(n+1)a_{n+1}z^n| = |a_{n+1}|(r')^n(n+1)\left(\frac{|z|}{r'}\right)^n$$

avec $(|a_{n+1}|(r')^n)_{n\in\mathbb{N}}$ bornée (d'après le lemme d'Abel 2.5), et $\left((n+1)\left(\frac{|z|}{r'}\right)^n\right)_{n\in\mathbb{N}}$ qui est le terme général d'une série convergente [18]. Donc par le lemme 2.5, $R_b \geq R_a$, ce qui permet de conclure. □

Remarque 2.12. On peut itérer et définir la série dérivée p-ième pour $p \in \mathbb{N}^*$, dont le terme général est

$$\forall n \in \mathbb{N}, \quad (n+p)\dots(n+2)(n+1)a_{n+p}z^n = \frac{(n+p)!}{n!}a_{n+p}z^n.$$

Par récurrence élémentaire, le rayon de la série entière associée est à nouveau R_a.

Pour le moment on a employé le mot "dérivée" sans le justifier et sans dériver aucune fonction ; nous allons corriger cela dans l'énoncé suivant. Cela anticipe sur le paragraphe 3 ; le lecteur qui n'est pas encore familier peut regarder la remarque 2.14.

Proposition 2.13. *Soit $(a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$, et $R_a \in [0, +\infty]$ le rayon de la série entière associée. Alors la fonction*

$$f : z \in D(0, R_a) \longmapsto \sum_{n\in\mathbb{N}} a_n z^n \tag{2.2}$$

est continue, $\mathbb{C}$-dérivable [19] sur $D(0, R_a)$ et même indéfiniment $\mathbb{C}$-dérivable [20] sur $D(0, R_a)$. De plus

$$\forall z \in D(0, R_a), \ f'(z) = \sum_{n\in\mathbb{N}} (n+1)a_{n+1}z^n = \sum_{n\in\mathbb{N}^*} n.a_n z^{n-1}. \tag{2.3}$$

17. Si $R_a < +\infty$ on peut prendre $r' = \frac{R_a + |z|}{2}$, et si $R_a = \infty$ on peut prendre $r' = |z| + 42$.
18. Par exemple parce que $(n+1)\left(\frac{|z|}{r'}\right)^n = o\left(\frac{1}{n^2}\right)$.
19. Voir la définition 3.1.
20. On évite de dire C^∞, car il n'est pas clair si cela vient de la terminologie du calcul différentiel (voir Chapitre VII) ou de l'holomorphie (voir paragraphe 3). Cela étant, il n'est pas faux de dire C^∞ puisque dans le cas présent, ce sera vrai dans les deux sens.

Remarque 2.14. On peut également, notamment si on n'a pas envie de parler de $\mathbb{C}$-dérivabilité, regarder la restriction de la fonction (2.2) à l'intervalle réel $]-R_a, R_a[$. On conclut que la fonction $x \in]-R_a, R_a[\mapsto \sum_{n\in\mathbb{N}} a_n x^n$ est de classe C^∞. D'ailleurs, on peut le faire en en appliquant la proposition 4.13 du Chapitre I (voir aussi l'exercice 5.22 du Chapitre I).

Remarque 2.15. Etant donnée une série entière associée à la suite $(a_n)_{n\in\mathbb{N}}$, on peut également définir la série primitive, dont le terme général est $\left(\frac{a_n}{n+1} z^{n+1}\right)_{n\in\mathbb{N}}$. Le rayon est à nouveau R_a, et si on définit f par (2.2) et F par $\forall z \in D(0, R_a),\ \ F(z) = \sum_{n\in\mathbb{N}} \frac{a_n}{n+1} z^{n+1}$, alors

$$F \text{ est } \mathbb{C}\text{-dérivable sur } D(0, R_a) \text{ et } \quad F' = f \ \text{ sur } D(0, R_a),$$

d'après la proposition 2.13.

Démonstration. Comme la série de fonctions définissant f convergence normalement sur chaque disque $\overline{D}(0, r)$ pour $r < R_a$, f est continuer sur $\bigcup_{r<R_a} \overline{D}(0, r) = D(0, R_a)$.

Si $R_a = 0$, il n'y a rien à démontrer, on suppose donc $R_a > 0$. Soit $z_0 \in D(0, R_a)$. Alors [21]

$$\forall z \in D(0, R_a) \setminus \{z_0\},\quad \frac{f(z) - f(z_0)}{z - z_0} = \sum_{n\geq 1} a_n \sum_{k=0}^{n-1} z_0^k z^{n-1-k}. \tag{2.4}$$

On pose $\forall n \in \mathbb{N}^*, \forall z \in D(0, R_a), g_n(z) = a_n \sum_{k=0}^{n-1} z_0^k z^{n-1-k}$. Il existe r tel que $|z_0| < r < R_a$ (voir la note 17), on a

$$\forall n \in \mathbb{N},\quad \|g_n\|_{\infty, \overline{D}(0,r)} \leq n|a_n| r^{n-1}$$

donc la convergence du terme de droite dans (2.4) est uniforme sur $\overline{D}(0, r)$. D'après l'exercice 4.21 du Chapitre I, on peut donc écrire l'interversion :

$$\lim_{z\to z_0} \sum_{n\geq 1} a_n \sum_{k=0}^{n-1} z_0^k z^{n-1-k} = \sum_{n\geq 1} a_n \lim_{z\to z_0} \sum_{k=0}^{n-1} z_0^k z^{n-1-k} = \sum_{n\geq 1} n.a_n z_0^{n-1},$$

d'où la $\mathbb{C}$-dérivabilité de f et le fait que $f'(z_0) = \sum_{n\geq 0} (n+1) a_{n+1} z_0^n$.

Pour le caractère indéfiniment $\mathbb{C}$-dérivable de f, on peut itérer la propriété précédente, puisqu'on obtient f' sous la forme d'une série entière de rayon R_a. [22] □

Corollaire 2.16. *Soit $(a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$, et $R_a \in [0, +\infty]$ le rayon de la série entière associée. On suppose $R_a > 0$. Alors*

$$\forall n \in \mathbb{N},\quad a_n = \frac{f^{(n)}(0)}{n!}$$

où f est définie par (2.2).

Démonstration. Il suffit d'itérer (2.3) et de prendre la valeur en $z = 0$. □

Remarque 2.17. Ce corollaire implique notamment une forme d'unicité des coefficients pour une série entière. En effet, si $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ sont les termes généraux de deux séries entières dont les rayons R_a, R_b sont *strictement positifs* et tels que

$$\forall z \in D(0, \min\{R_a, R_b\}),\quad \sum_{n\in\mathbb{N}} a_n z^n = \sum_{n\in\mathbb{N}} b_n z^n$$

alors par le corollaire 2.16 précédent on déduit que pour tout $n \in \mathbb{N}$, $a_n = b_n$. [23]

Remarquons qu'on aurait eu le même résultat dans le cadre réel en remplaçant $D(0, \min\{R_a, R_b\})$ par son intersection avec $\mathbb{R}$. On va voir en fait un résultat plus fort au paragraphe suivant.

---

21. On utilise l'identité classique

$$\forall n \in \mathbb{N}, \forall (a, b) \in \mathbb{C},\quad a^n - b^n = (a - b) \sum_{k=0}^{n-1} a^k b^{n-1-k}.$$

22. Bien sûr, on verra au paragraphe 3 (Théorème 3.28) qu'une fonction $\mathbb{C}$-dérivable est automatiquement indéfiniment $\mathbb{C}$-dérivable, mais ce résultat est difficile et nous n'en avons pas besoin ici.

23. On peut montrer ce principe d'unicité sans le corollaire 2.16 : en effet si on suppose $\forall z \in D(0, \varepsilon), \sum_{n\in\mathbb{N}} a_n z^n = \sum_{n\in\mathbb{N}} b_n z^n$, alors l'évaluation en $z = 0$ donne $a_0 = b_0$. En retirant a_0 de chaque côté puis en divisant par z, on obtient $\forall z \in D(0, \varepsilon) \setminus \{0\}, \sum_{n\in\mathbb{N}} a_{n+1} z^n = \sum_{n\in\mathbb{N}} b_{n+1} z^n$. L'égalité reste valable en 0 par continuité des séries entières. L'évaluation en 0 donne justement $a_1 = b_1$. En itérant ce procédé, on prouve que $\forall n \in \mathbb{N},\ a_n = b_n$.

2.3.2 Principe des zéros isolés "faible"

On peut généraliser la remarque 2.17 avec le principe des zéros isolés. On donne ici un énoncé qui se focalise autour du point 0 (d'où le mot "faible"), on reviendra sur ce principe dans le cadre plus général des fonctions développables en séries entières, voir le théorème 2.37.

Proposition 2.18 (Principe des zéros isolés autour de 0)**.** *Soit $(a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$ telle que la série entière associée ait un rayon $R_a > 0$. On pose $\forall z \in D(0, R_a), f(z) = \sum_{n\in\mathbb{N}} a_n z^n$.*

S'il existe une suite de complexes $(z_p)_{p\in\mathbb{N}}$ d'éléments de $D(0, R_a)$ telle que

$$\forall p \in \mathbb{N}, \Big[\ z_p \neq 0, \quad f(z_p) = 0\Big], \qquad \textit{et} \qquad z_p \xrightarrow[p\to+\infty]{} 0$$

alors $\forall n \in \mathbb{N}, a_n = 0$, c'est-à-dire $f \equiv 0$.

Démonstration. Par l'absurde, supposons qu'il existe $n_0 \in \mathbb{N}$ tel que $a_{n_0} \neq 0$. On choisit n_0 minimal pour cette propriété. Alors on peut écrire

$$\forall z \in D(0, R_a), \quad f(z) = z^{n_0} g(z), \qquad \text{où} \qquad g(z) = \sum_{n\in\mathbb{N}} a_{n+n_0} z^n.$$

Par hypothèse, $g(z_p) = 0$ pour tout $p \in \mathbb{N}$ (on a utilisé que $z_p \neq 0$). Mais alors par continuité de g en 0, $g(0) = 0 = a_{n_0}$, ce qui constitue une contradiction. □

2.4 Fonction exponentielle

On va maintenant définir et étudier la fonction exponentielle, définie sur $\mathbb{C}$. À vrai dire, la construction que nous proposons ici est probablement plus simple que celles proposées au paragraphe 3.4.2 du Chapitre I qui étaient restreintes à $\mathbb{R}$; elle est cependant un peu plus élaborée [24] car elle utilise des résultats sur les séries, ce qui l'exclut des programmes du lycée, mais par contre l'insère parfaitement dans le programme de l'agrégation. Il est à noter qu'avec la méthode proposée ici, on définit directement l'exponentielle sur $\mathbb{C}$, ce qui veut dire qu'on n'a pas besoin de l'exponentielle réelle pour le faire [25]. On déduira de cette construction une définition de sin et cos sur $\mathbb{C}$ et donc sur $\mathbb{R}$.

2.4.1 Définitions et propriétés

Définition-Proposition 2.19. Pour tout $z \in \mathbb{C}$, on définit

$$\exp(z) = e^z := \sum_{n\in\mathbb{N}} \frac{z^n}{n!},$$

ce qui est possible car cette série entière a pour rayon de convergence $+\infty$. On appellera $\exp : \mathbb{C} \to \mathbb{C}$ la fonction exponentielle.

Démonstration. On applique la règle de d'Alembert : en effet

$$\forall n \in \mathbb{N}, \quad \frac{\frac{1}{(n+1)!}}{\frac{1}{n!}} = \frac{1}{n+1}.$$

□

24. Pour moi les mots "élaboré" et "simple" ne sont pas antagonistes, ni incompatibles.
25. On pourrait définir l'exponentielle complexe à partir de l'exponentielle réelle, par exemple en définissant

$$\forall (a, b) \in \mathbb{R}^2, \quad \exp(a + ib) := e^a.(\cos(b) + i\sin(b))$$

mais cela nécessite d'avoir construit les fonctions cos et sin sur $\mathbb{R}$ (en plus de $\exp : \mathbb{R} \to \mathbb{R}$), ce qui n'est pas non plus évident.

Ainsi comme toute série entière, la fonction exponentielle est indéfiniment $\mathbb{C}$-dérivable, et on constate avec le paragraphe 2.3.1 que

$$\forall z \in \mathbb{C}, \quad \exp'(z) = \sum_{n\in\mathbb{N}} (n+1)\frac{1}{(n+1)!}z^n = \exp(z).$$

On retrouve ainsi très facilement ce qui avait motivé la construction du thérorème 3.65 du Chapitre I. On va maintenant voir que exp satisfait à l'équation fonctionnelle qui avait motivé le théorème 3.66 du Chapitre I :

Proposition 2.20. *On a la propriété*

$$\forall (z,z') \in \mathbb{C}^2, \quad \exp(z+z') = \exp(z).\exp(z').$$

Démonstration. Cela découle de la proposition 2.45 du Chapitre I, en constatant que

$$\forall (z,z') \in \mathbb{C}^2, \quad \forall n \in \mathbb{N}, \quad \sum_{k=0}^{n} \frac{z^k}{k!}\frac{(z')^{n-k}}{(n-k)!} = \frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k} z^k (z')^{n-k} = \frac{(z+z')^n}{n!}$$

d'après la formule du binôme de Newton. □

Proposition 2.21. *1. La fonction* exp *ne s'annule pas, et satisfait :*

$$\forall z \in \mathbb{C}, \quad \frac{1}{\exp(z)} = \exp(-z).$$

2. Pour $z \in \mathbb{C}$, on a

$$\overline{\exp(z)} = \exp(\overline{z}), \qquad \textit{et } |\exp(z)| = \exp(\mathrm{Re}(z)).$$

3. Enfin on a

$$\forall z \in \mathbb{C}, \quad |\exp(z)| = 1 \Leftrightarrow z \in i\mathbb{R}.$$

Démonstration. Il s'agit de calculs simples en utilisant la définition et la Proposition 2.20 □

Théorème 2.22. *La fonction*

$$\exp : (\mathbb{C},+) \longrightarrow (\mathbb{C}^*,\times)$$

est un morphisme de groupe continu, sujectif mais non injectif.

Démonstration. On propose plusieurs démonstrations pour la surjectivité :

Preuve 1 : *Voir par exemple [Pom94, 40.1.4], et aussi [AM04, page 75]. Cette preuve utilise un résultat du Chapitre VII (et en partie le théorème 3.3 donné plus loin) ; elle est classique, se base sur les propriétés de morphisme de l'exponentielle, et peut être adaptée à l'exponentielle matricielle (avec un peu plus de travail, voir [Zav13, Problème 9.II] qui fait un développement classique et qui se place dans des leçons d'analyse et d'algèbre).

Posons $H = \exp(\mathbb{C})$. Il s'agit d'un sous-groupe de $(\mathbb{C}^*,\times)$. On va utiliser un argument de connexité. Pour montrer que H est ouvert, on commence par voir que exp est de classe C^1 et $d\exp(0)h = h$ pour tout $h \in \mathbb{R}^2$ (on peut le voir “à la main” ou appliquer le théorème 3.3 pour le lien entre dérivée et différentielle), en particulier $d\exp(0)$ est inversible, donc par le théorème d'inversion locale (théorème 3.7 du Chapitre VII), il existe U et V deux voisinages de 0 et 1 respectivement tels que $\exp(U) = V$. Ainsi V est un voisinage de 1 qui est dans H, et par structure de groupe, si $z \in H$, $z.V = \{z.z', z' \in V\}$ est un voisinage ouvert de z qui est aussi dans H (car si $z = \exp(a)$, $z' = \exp(b)$ alors $z.z' = \exp(a+b)$). En conclusion, H est bien ouvert dans $\mathbb{C}^*$.

On a $\mathbb{C}^* = \bigcup_{z\in\mathbb{C}^*} z.G$, donc $\mathbb{C}^* \setminus H = \bigcup_{z\in\mathbb{C}^*, z\notin H} z.H$ [26] et donc $\mathbb{C}^* \setminus H$ est une union d'ouverts (si H est ouvert, on vérifie facilement que $z.H$ aussi), et est donc ouvert, ce qui signifie que H est fermé dans $\mathbb{C}^*$. Par connexité de $\mathbb{C}^*$ et le fait que H est non vide, on conclut bien que $H = \mathbb{C}^*$. [27]

26. Si vous aimez le language de la théorie des groupes, on a fait agir H sur $G = \mathbb{C}^*$ par translation (à gauche ou à droite vu que $\mathbb{C}^*$ est commutatif), ce qui donne une partition de G par ses orbites sous cette action.

27. **On vient de prouver un lemme classique de théorie des groupes topologiques : si G est un groupe topologique et H un sous-groupe de G. Alors

— si H est ouvert, alors il est aussi fermé,

— mieux, si H contient un voisinage du neutre de G, alors H est ouvert et fermé.

Preuve 2 : On suit [Tau06, Théorème 3.7.4]. Soit $\zeta \in \mathbb{C} \setminus \mathbb{R}_-$. On pose $\gamma : t \in [0,1] \mapsto 1 - t + t\zeta$ qui est C^∞ et ne passe pas par 0, et

$$g : \begin{array}{ccc} [0,1] & \longrightarrow & \mathbb{C} \\ t & \longmapsto & \displaystyle\int_0^t \frac{\gamma'(s)}{\gamma(s)} ds, \end{array} \qquad \text{et} \qquad h := \gamma \exp(-g).$$

On a g et h de classe C^1, et par calcul on obtient $h' = 0$, et donc $h(0) = h(1)$ [28], ce qui donne $\zeta = \exp(g(1))$, d'où $\mathbb{C} \setminus \mathbb{R}_- \subset \mathrm{Im}(\exp)$. [29]

Pour conclure, il reste à voir que les réels strictement négatifs sont dans l'image de l'exponentielle. Pour cela, on utilise ce qui précède, qui montre que $i \in \mathbb{C} \setminus \mathbb{R}_-$ a un antécédent par exp, qu'on note $\xi \in \mathbb{C}$. Alors $\exp(2\xi) = i^2 = -1$, et par suite si $x \in \mathbb{R}_- \setminus \{0\}$, on note α un antécédent par exp de $|x| \in \mathbb{C} \setminus \mathbb{R}_-$, et on écrit $x = -|x| = -\exp(\alpha) = \exp(\alpha + 2\xi)$, ce qui conclut.

Preuve 3 : En fait, cette preuve consiste à prouver d'abord le théorème 2.23 ci-après, ainsi que l'étude de la fonction exp restreinte à $\mathbb{R}$ puis à en déduire la surjectivité de exp : en effet, si $z \in \mathbb{C}^*$, alors $z = |z|\frac{z}{|z|}$. Par le théorème 2.23, il existe θ tel que $\frac{z}{|z|} = e^{i\theta}$, et donc $z = \exp(\ln(|z|) + i\theta)$.

Pour prouver la non-injectivité, nous allons utiliser la surjectivité ! En effet par surjectivité, il existe $\xi \in \mathbb{C}$ tel que $e^\xi = i$, et alors $e^{4\xi} = 1 = e^0$. □

2.4.2 Définition de π et des fonctions trigonométriques

Nous allons voir qu'en conséquence de la construction précédente, on peut définir le nombre π et construire les fonctions trigonométriques. On pose $\mathcal{U} = \{z \in \mathbb{C},\ \ |z| = 1\}$.

Théorème 2.23. *L'application*

$$\begin{array}{ccc} (\mathbb{R}, +) & \longrightarrow & (\mathcal{U}, \times) \\ t & \longmapsto & e^{it} \end{array}$$

est un morphisme de groupe, continu, surjectif et non injectif. Son noyau est de la forme $a\mathbb{Z}$ avec $a \in]0, +\infty[$, et a est le plus petit réel t tel que $e^{it} = 1$. On pose

$$\pi = \frac{a}{2}.$$

Démonstration. On suppose ici le théorème 2.22 déjà démontré (donc par la preuve 1 ou 2). On donnera à l'exercice 2.50 une possible démonstration du théorème 2.23 qui n'utilise pas le théorème 2.22, ce qui permet d'utiliser la preuve 3 de ce dernier.

Notons $\forall t \in \mathbb{R}, \varphi(t) = e^{it}$, qui est bien à valeurs dans $\mathcal{U}$ par la Proposition 2.21. Le fait que φ est un morphisme continu est une simple restriction à $\mathbb{R}$ de propriétés valables sur $\mathbb{C}$ (la multiplication par i dans l'exponentielle ne changeant pas ces propriétés). La surjectivité de φ est conséquence de la surjectivité de $\exp : \mathbb{C} \to \mathbb{C}^*$ et de la question 3 de la Proposition 2.21. Avec la notation de la preuve du théorème 2.22, on a $e^{4\xi} = 1$ et avec le point 3 de la Proposition 2.21, $4\xi \in i\mathbb{R}$. Donc φ est non injective. Son noyau est un sous-groupe non trivial de $\mathbb{R}$. D'après la proposition 1.3 du Chapitre I, les sous-groupes de $\mathbb{R}$ sont denses ou de la forme $a\mathbb{Z}$ pour $a \in \mathbb{R}_+^*$. On exclut la première éventualité car $\mathrm{Ker}(\varphi)$ est fermé comme image réciproque de $\{1\}$ par φ qui est continu. Ainsi il existe bien un unique $a \in]0, +\infty[$ tel que $\mathrm{Ker}(\varphi) = a\mathbb{Z}$, et a est bien le plus petit réel strictement positif dont l'image est 1. □

Corollaire 2.24. *Le noyau de* $\exp : \mathbb{C} \to \mathbb{C}^*$ *est* $2i\pi\mathbb{Z}$*, et*

$$\forall (z, z') \in \mathbb{C}^2,\ \exp(z) = \exp(z') \ \ \Leftrightarrow \ \ z - z' \in 2i\pi\mathbb{Z}.$$

28. Ici on utilise un résultat du premier chapitre mais pour une fonction $h : I \to \mathbb{C}$ où I est un intervalle de $\mathbb{C}$. Le résultat est encore valable car il suffit de dire que les fonctions coordonnées de h (c'est-à-dire $\mathrm{Re}(h)$ et $\mathrm{Im}(h)$ ont une dérivée nulle donc sont constantes.

29. On pourrait trouver cette preuve un peu mystérieuse ; mais en fait, elle anticipe sur le paragraphe 3.4.1 et la vision de la fonction Logarithme principal définie par l'intégration d'une fonction d'une variable complexe sur un chemin (qui est ici un segment). En effet, le terme $g(1)$ n'est autre que $\mathrm{Log}(\zeta)$ tel qu'on le construit dans l'exercice 3.61.

Corollaire 2.25 (Forme exponentielle d'un complexe). *Soit $z \in \mathbb{C}$. Alors il existe $\rho \in \mathbb{R}_+$ et $\theta \in \mathbb{R}$ tel que*

$$z = \rho e^{i\theta}.$$

Le nombre ρ est unique et est donné par $|z|$. Si $z \neq 0$, alors θ est unique modulo 2π [30] *; on peut donc choisir $\theta \in]-\pi, \pi]$ et on note $\mathrm{Arg}(z) = \theta$ qu'on appelle argument principal de z.*

On peut également déduire de ce qui précède une construction des fonctions sin et cos :

Définition 2.26. Pour $t \in \mathbb{R}$, on pose

$$\cos(t) := \mathrm{Re}(e^{it}) = \frac{e^{it} + e^{-it}}{2}, \quad \sin(t) := \mathrm{Im}(e^{it}) = \frac{e^{it} - e^{-it}}{2i}. \tag{2.5}$$

En conséquence on a les formules :

$$\forall t \in \mathbb{R}, \quad e^{it} = \cos(t) + i.\sin(t), \quad \cos(t) = \sum_{n\in\mathbb{N}} (-1)^n \frac{t^{2n}}{(2n)!}, \quad \sin(t) = \sum_{n\in\mathbb{N}} (-1)^n \frac{t^{2n+1}}{(2n+1)!}. \tag{2.6}$$

Remarque 2.27. On peut en fait définir, pour $z \in \mathbb{C}$:

$$\cos(z) := \frac{e^{iz} + e^{-iz}}{2}, \quad \sin(z) := \frac{e^{iz} - e^{-iz}}{2i},$$

et les formules (2.6) reste valable. Mais attention, on ne peut alors plus écrire $\cos(z) := \mathrm{Re}(e^{iz})$ ou $\sin(z) := \mathrm{Im}(e^{iz})$.

On peut également écrire $\cos(z) = \cosh(iz)$ et $\sin(z) = \frac{\sinh(iz)}{i}$.

Enfin, on peut montrer

$$\cos^2 + \sin^2 = 1.$$

2.5 Première construction du Logarithme principal

On sort un peu de la question des séries entières, et on utilise le paragraphe précédent pour construire un logarithme complexe [31]. Comme la fonction exp n'est pas bijective, on ne peut pas définir le logarithme comme son inverse. Néanmoins, comme on l'a fait au paragraphe 3.4.3 du Chapitre I pour les fonctions arccos, arcsin, on va restreindre la fonction exp à un ouvert sur lequel elle sera injective. Nous présentons ici un choix possible de tel ouvert, mais gardons en tête que ce choix est arbitraire et que d'autres sont possibles. On parlera de Logarithme principal pour le choix que nous faisons ici.

Définition-Proposition 2.28. La fonction exponentielle restreinte à la bande $O = \{(x, y) \in \mathbb{R}^2, y \in]-\pi, \pi[\}$ est injective, et d'image $\mathbb{C} \setminus \mathbb{R}_-$. Ainsi, tout $z \in \mathbb{C} \setminus \mathbb{R}_-$ admet un unique antécédent par exp dans O, que l'on note $\mathrm{Log}(z)$ et qu'on appelle Logarithme principal de z. On définit ainsi

$$\mathrm{Log} : \mathbb{C} \setminus \mathbb{R}_- \longrightarrow O \subset \mathbb{C}$$

telle que pour tout $z \in \mathbb{C} \setminus \mathbb{R}_-, \exp(\mathrm{Log}(z)) = z$.

Démonstration. C'est une conséquence de l'exercice 2.24. □

Voyons qu'on peut écrire cette fonction logarithme à l'aide du log réel [32] et de l'argument d'un nombre complexe (exercice 2.25) :

Proposition 2.29. *Si $z \in \mathbb{C} \setminus \mathbb{R}_-$, alors*

$$Log(z) = \ln(|z|) + i\mathrm{Arg}(z)$$

où $Arg(z)$ est l'unique argument de z dans $]-\pi, \pi[$.

30. C'est-à-dire que si θ' convient également, alors $\theta - \theta' \in 2\pi\mathbb{Z}$.
31. Notez l'emploi de l'article "un" ; il y a effectivement plusieurs logarithmes sur $\mathbb{C}$, voir le Parahraphe 4.4.
32. On peut par exemple définir le logarithme néparien comme la fonction réciproque de exp : $\mathbb{R} \to \mathbb{R}_+^*$ dont on vérifie facilement qu'elle est continue et bijective.

Démonstration. On écrit $\mathrm{Log}(z) = x + iy$ avec $x \in \mathbb{R}$ et $y \in]-\pi, \pi[$. Alors $\exp(x+iy) = e^x.e^{iy} = z$ d'où $e^x = |z|$ et $e^{iy} = e^{iArg(z)}$, d'où le résultat d'après le théorème 2.23. □

Remarque 2.30 (Une définition des fonctions puissances complexes)**.** Étant donné $\alpha \in \mathbb{C}$, on peut définir

$$\forall z \in \mathbb{C} \setminus \mathbb{R}_-, \ \ z^\alpha = e^{\alpha \mathrm{Log}(z)}.$$

Il s'agit d'une fonction $\mathbb{C}$-dérivable (comme composée de fonctions $\mathbb{C}$-dérivables, voir le paragraphe 3.3). Si $\alpha \in \mathbb{N}$ on retrouve l'objet habituel. Sinon, il faut faire attention que cette fonction dépend du choix de la fonction Log, dont on a déjà dit qu'il n'était pas unique. On parle ici de fonction “puissance α”. On peut aussi montrer

$$\forall z \in D(0,1), \quad (1+z)^\alpha = \sum_{n\in\mathbb{N}} \frac{\prod_{i=0}^{n-1}(\alpha - i)}{n!} z^n.$$

Voir aussi le paragraphe 4.4.

2.6 Fonctions développables en série entière

Jusqu'à présent, on est parti d'une série entière, et on a vu que cela pouvait définir une fonction sur un disque de $\mathbb{C}$. On inverse le point de vue ici, en prenant une fonction f définie sur un ouvert de $\mathbb{C}$, en supposant qu'elle peut être développée en série entière dans un disque autour de tout point de cet ouvert. Nous allons étudier des conditions sur f pour qu'une telle condition puisse être satisfaite, mais cela sera drastiquement renforcé dans le paragraphe 3.

2.6.1 Définition et exemples

Définition 2.31. Soit U un ouvert de $\mathbb{C}$. On dit qu'une fonction $f : U \to \mathbb{C}$ est développable en série entière en $z_0 \in U$ si

$$\exists r \in]0, d(z_0, U^c)[, \ \exists (a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}, \ \forall z \in D(z_0, r), \text{la série } \sum_{n\in\mathbb{N}} a_n (z-z_0)^n \text{ est convergente}$$
$$\text{et} \quad f(z) = \sum_{n\in\mathbb{N}} a_n (z-z_0)^n.$$

On dit que f est développable en série entière sur U, ou analytique sur U, si elle est développable en série entière en tout point de U.

Dans ce contexte la remarque 2.4 devient particulièrement utile.

Remarque 2.32. On peut également donner la même définition pour une fonction $f : I \to \mathbb{R}$ où I est un intervalle ouvert de $\mathbb{R}$. Dans ce cas on peut remplace $D(z_0, r)$ par $]z_0 - r, z_0 + r[$ où $z_0 \in I$.

Commençons par montrer qu'une série entière définit bien une fonction analytique. Attention, cela n'est pas évident ; une série entière définit naturellement une fonction qui est définie sur le disque $D(0, R)$ et qui est développable autour de son point origine (ici 0), mais on va vérifier qu'elle est aussi développable autour de tous les points du disque $D(0, R)$.

Proposition 2.33. *Soit $(a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$ telle que la série entière associée a un rayon $R_a > 0$. On pose $\forall z \in D(0, R_a), f(z) = \sum_{n\in\mathbb{N}} a_n z^n$. Alors la fonction f est analytique sur $D(0, R_a)$.*

Remarque 2.34. On sera un peu plus précis dans la démonstration. En effet, on montrera que le rayon de convergence du développement en série entière de f autour d'un point $z_0 \in D(0, R_a)$ sera supérieur ou égal à $R_a - |z_0|$, autrement dit que le disque le plus grand de centre z_0 qui soit inclus dans $D(0, R_a)$ est bien un disque de convergence. On verra que ce phénomène est général, voir la remarque 3.31.

Démonstration. Soit $z_0 \in D(0, R_a)$. Nous allons montrer que

$$\forall z \in D(z_0, R_a - |z_0|), \quad \text{la série de terme général } \left(\frac{f^{(n)}(z_0)}{n!}(z-z_0)^n\right)_{n\in\mathbb{N}} \quad \text{est convergente et}$$

$$f(z) = \sum_{n\in\mathbb{N}} \frac{f^{(n)}(z_0)}{n!}(z-z_0)^n.$$

D'après la remarque 2.12, on a

$$f^{(n)}(z_0) = \sum_{k\in\mathbb{N}} \frac{(n+k)!}{k!} a_{n+k} z_0^k,$$

d'où, pour $z \in D(z_0, R_a - |z_0|)$,

$$\begin{aligned}
\sum_{n\in\mathbb{N}} \left|\frac{f^{(n)}(z_0)}{n!}(z-z_0)^n\right| &\leq \sum_{n\in\mathbb{N}}\sum_{k\in\mathbb{N}} \binom{n+k}{n} |a_{k+n}||z_0|^k |z-z_0|^n \quad \text{Changement d'indice : } p = n+k \\
&\leq \sum_{n\in\mathbb{N}}\sum_{p=n}^{+\infty} \binom{p}{n} |a_p||z_0|^{p-n}|z-z_0|^n \quad \text{On applique le théorème de Fubini 2.46} \\
&\leq \sum_{p\in\mathbb{N}} |a_p| \sum_{n=0}^{p} \binom{p}{n} |z_0|^{p-n}|z-z_0|^n = \sum_{p\in\mathbb{N}} |a_p| \Big(|z_0| + |z-z_0|\Big)^p
\end{aligned}$$

Mais on sait que $|z_0| + |z-z_0| < R_a$, et donc $\sum_{p\in\mathbb{N}} |a_p| \Big(|z_0| + |z-z_0|\Big)^p$ est convergente, et donc la série de terme général $\left(\frac{f^{(n)}(z_0)}{n!}(z-z_0)^n\right)_{n\in\mathbb{N}}$ est absolument convergente, donc convergente. De plus, le même calcul que précédemment sans les modules[33] donne :

$$\sum_{n\in\mathbb{N}} \frac{f^{(n)}(z_0)}{n!}(z-z_0)^n = \sum_{p\in\mathbb{N}} a_p z^p = f(z)$$

ce qui conclut la démonstration. □

2.6.2 Propriétés

La Proposition 2.13 implique déjà facilement que

Proposition 2.35. *Soit U un ouvert de $\mathbb{C}$. Si $f : U \to \mathbb{C}$ est analytique, alors f est indéfiniment $\mathbb{C}$-dérivable sur U, et si $z_0 \in U$, alors les coefficients $(a_n)_{n\in\mathbb{N}}$ qui apparaissent dans le développement en série entière de f autour de z_0 satisfait*

$$\forall n \in \mathbb{N}, \quad a_n = \frac{f^{(n)}(z_0)}{n!}.$$

Voici maintenant deux propriétés qui témoignent de la rigidité des fonctions analytiques.

Théorème 2.36 (Théorème du prolongement analytique)**.** *Soit U un ouvert connexe de $\mathbb{C}$, $z_0 \in U$ et $f : U \to \mathbb{C}$ analytique. Alors*

$$\Big[\forall n \in \mathbb{N}, \quad f^{(n)}(z_0) = 0\Big] \iff f \equiv 0$$

33. Cette façon de rédiger (commencer par un calcul avec les valeurs absolues/modules à l'intérieur, pour pouvoir être en mesure d'appliquer un théorème et reprendre le calcul sans modules) est très fréquente, notamment quand on utilise des théorèmes de type Fubini ou d'interversion de limites ou limite/intégrales ; on l'utilisera également au Chapitre IV. Si vous faites bien le premier calcul, le second est en général très similaire et peut être raccourci. Les étudiants font souvent le contraire, ils font bien le second calcul, mais ne font pas du tout le premier, ce qui fait qu'ils n'ont pas justifié leur calcul, ce qui est fortement pénalisé.

Démonstration. Seul le sens $\Rightarrow$ est non trivial.[34] D'après le développement de f autour de z_0, il existe $r_0 > 0$ tel que $D(z_0, r_0) \subset U$ et f s'annule sur $D(z_0, r_0)$. On considère alors

$$V = \Big\{z \in U, \ \ \exists r > 0, \ D(z, r) \subset U \ \text{ et } \ f \text{ s'annule sur } D(z, r)\Big\}.$$

L'ensemble V est non vide car $z_0 \in V$ (en effet le développement en série entière de f au voisinage de z_0 donne que f est nulle sur ce voisinage), et il est clairement ouvert. Montrons qu'il est également fermé : soit $(y_p)_{p\in\mathbb{N}}$ une suite d'éléments de V qui converge vers $y_\infty \in U$[35]. Alors pour chaque $p \in \mathbb{N}$, comme f s'annule sur un voisinage de y_p, on a $f^{(n)}(y_p) = 0$ pour tout $n \in \mathbb{N}$. Pour tout $n \in \mathbb{N}$, par continuité de $f^{(n)}$, on obtient $f^{(n)}(y_\infty) = 0$, et donc en développant en série entière autour de y_∞ on obtient que $y_\infty \in V$, ce qui montre que V est un fermé de U. Par connexité de U, on obtient $V = U$, ce qui conclut la preuve. $\square$

Théorème 2.37 (Théorème des zéros isolés)**.** *Soit U un ouvert connexe de $\mathbb{C}$ et $f : U \to \mathbb{C}$ analytique. Si l'ensemble $\{z \in U, f(z) = 0\}$ possède un point d'accumulation*[36] *dans U, alors $f \equiv 0$.*

Remarque 2.38. L'énoncé du théorème 2.37 peut être donné de façon équivalente en disant que si $f : U \to \mathbb{C}$ est analytique et non identiquement nulle (avec U connexe), alors l'ensemble les points de U où f s'annule sont isolés, c'est-à-dire que si $z_0 \in U$ est tel que $f(z_0) = 0$, alors il existe $\varepsilon > 0$[37] tel que f ne s'annule pas sur $D(z_0, \varepsilon) \setminus \{z_0\}$. On peut montrer que cela signifie que pour chaque compact $K \subset U$, f s'annule un nombre fini de fois dans K. C'est une version faible de ce que satisfont les polynômes, qui ont eux un nombre fini de zéros.

Remarque 2.39. Dans l'énoncé du théorème des zéros isolés, attention à ceux qui disent seulement $\{z \in U, f(z) = 0\}$ n'a pas de point d'accumulation. Il peut arriver en effet que l'ensemble $\{z \in U, f(z) = 0\}$ ait des points d'accumulation dans $\mathbb{C}$. Par exemple la fonction $f : z \in \mathbb{C}^* \mapsto \sin\left(\frac{1}{z}\right)$ est analytique sur $\mathbb{C}^*$[38], mais elle s'annule sur les points de la forme $\frac{1}{k\pi}$ où $k \in \mathbb{Z}^*$. Il y a donc une infinité de zéros de cette fonction autour de 0, mais précisément 0 n'est pas dans le domaine de définition de f.

Démonstration. Supposons que $(z_p)_{p\in\mathbb{N}}$ et $z_\infty \in U$ sont tels que pour tout $p \in \mathbb{N}$, $f(z_p) = 0$, $z_p \neq z_\infty$, et $z_p \xrightarrow[n\to+\infty]{} z_\infty$[39]. Supposons par l'absurde qu'il existe $n_0 \in \mathbb{N}$ tel que $f^{(n_0)}(z_\infty) \neq 0$: on choisit n_0 minimal pour cette propriété. Alors en écrivant le développement de f autour de z_∞, on obtient qu'il existe $r > 0$ tel que

$$\forall z \in D(z_0, r), \ \ f(z) = (z - z_\infty)^{n_0} g(z), \quad \text{où} \quad g(z) = \sum_{n \geq n_0} \frac{f^{(n)}(z_\infty)}{n!}(z - z_\infty)^{n-n_0}$$

et pour tout $p \in \mathbb{N}$, comme $z_p \neq z_\infty$, on obtient $g(z_p) = 0$. Par continuité de g, on récupère en faisant $p \to +\infty$ que $g(z_\infty) = 0 = \frac{f^{(n_0)}(z_\infty)}{n_0!}$, ce qui est une contradiction.

Ainsi $\forall n \in \mathbb{N}$, $f^{(n)}(z_\infty) = 0$ donc par le théorème 2.36, $f \equiv 0$. $\square$

Remarque 2.40. On peut adapter ces deux principes à deux fonctions $f : U \to \mathbb{C}$ et $g : U \to \mathbb{C}$ définies et analytiques sur U (supposé connexe) en considérant la fonction $f - g$. En particulier, si l'ensemble $\{z \in U, f(z) = g(z)\}$ possède un point d'accumulation dans U, alors $f \equiv g$. Ainsi si l'une des propriétés suivantes est vérifiée :

1. f et g sont égales sur un (petit) disque $D(z_0, \alpha) \subset U$ avec $\alpha > 0$
2. f et g sont égales sur un segment $[z_0, z_1] \subset U$ avec $z_0 \neq z_1$

34. La preuve qui suit est typique du raisonnement par connexité.
35. Attention à bien supposer que la limite est dans U ; on ne peut pas montrer que V est un fermé de $\mathbb{C}$, et pour cause in fine $V = U$ n'est pas fermé dans $\mathbb{C}$. Voir aussi la note 9.
36. Voir le paragraphe 1.3 pour une définition d'un point d'accumulation.
37. Ce $\varepsilon > 0$ peut dépendre de z_0.
38. Ceci n'est pas évident a priori ; c'est clair si on sait que les fonctions holomorphes sont analytiques (voir le Théorème 3.28, mais sans cela il faudrait étudier la composition de fonctions analytiques (comme on l'a fait pour le produit à la Proposition 2.9), ce qui n'est pas simple (mais possible) via le calcul, voir la remarque 2.10.
39. Ici on pourrait appliquer directement la proposition 2.18 autour du point z_∞. On reproduit la preuve par commodité.

3. f et g sont égales en une infinité de points qui sont tous dans un même compact de U

alors f et g sont égales sur tout U.

Remarque 2.41. Une conséquence est que si $f : I \to \mathbb{C}$ est analytique sur un intervalle I dans $\mathbb{R}$, et si U est un ouvert connexe de $\mathbb{C}$ qui contient I, alors f admet au plus un prolongement analytique. Par exemple, la fonction exponentielle qu'on a définie au paragraphe 2.4 est l'unique fonction analytique qui prolonge $\exp : \mathbb{R} \to \mathbb{R}$ à $\mathbb{C}$.

2.6.3 Différences entre fonctions C^∞ et fonctions analytiques dans le cas réel

Dans ce paragraphe, on se place dans le cadre réel, c'est-à-dire que les fonctions sont définies sur I un intervalle ouvert de $\mathbb{R}$. On a vu que les fonctions développables en série entière étaient C^∞. Réciproquement, si $f : I \to \mathbb{R}$ est C^∞, on voudrait savoir si f est analytique.

On va voir que la réponse est non en général, et il y a essentiellement deux types d'obstacles :

1. Si $f : I \to \mathbb{R}$ est C^∞, on appelle série de Taylor de f en $x \in I$ la série de terme général $\left(\frac{f^{(n)}(x)}{n!}\right)_{n\in\mathbb{N}}$. La première question est de savoir si cette série entière a un rayon de convergence non nul. On verra des situations où ce n'est pas le cas (auquel cas f ne peut pas être développable en série entière en x),
2. et même si le rayon de convergence de la série de Taylor de f en $x \in I$ est de rayon non nul, nous allons voir qu'il n'est pas assuré que la fonction f coïncide avec cette série sur un voisinage de x.

Commençons par le premier obstacle : le résultat suivant exprime qu'on peut obtenir n'importe quelle série entière comme série de Taylor d'une fonction C^∞. Il suffit donc pour construire un contre-exemple de fonction C^∞ non développable en série entière d'appliquer le théorème de Borel à une suite $(a_n)_{n\in\mathbb{N}}$ telle que le rayon de convergence de la série entière associée est nul, par exemple avec $\forall n \in \mathbb{N},\ a_n = n!$. A l'Exercice 2.61 on voit un autre exemple plus explicite d'une situation similaire.

Théorème 2.42 (*Théorème de Borel [40]). *Soit $(a_n) \in \mathbb{C}^\mathbb{N}$. Alors il existe $f : \mathbb{R} \to \mathbb{C}$ qui est C^∞ et telle que*

$$\forall n \in \mathbb{N}, \quad \frac{f^{(n)}(0)}{n!} = a_n.$$

On invite le lecteur à aller voir [Gou08b, page 280] ou [Rou03, page 359] ou [QZ13, page 296] pour une démonstration, qui représente un développement d'assez bon niveau.

Quant à un exemple de la seconde situation, il s'agit de l'exemple classique

$$\forall x \in \mathbb{R}, \quad f(x) = \begin{cases} e^{-\frac{1}{x}} \text{ si } x > 0 \\ 0 \text{ sinon} \end{cases},$$

qui est C^∞ et dont la série de Taylor en 0 est la série nulle, voir l'Exercice 2.60, et donc f n'est pas développable en série entière en 0.

Remarque 2.43. Les exemples ci-dessus fournissent des fonctions C^∞ qui ne sont pas développables en série entière en un point de leur ensemble de définition. On propose à l'exercice **??** une fonction qui est C^∞ mais qui n'est développable en série entière en aucun point de son ensemble de définition.

On peut maintenant donner le résultat suivant, qui donne une condition nécessaire et suffisante sur les fonctions C^∞ pour qu'elles soient développables en série entière :

Théorème 2.44. *Soit $f : I \to \mathbb{R}$ de classe C^∞ sur un intervalle ouvert I. Alors f est analytique si et seulement si*

$$\forall [a,b] \subset I, \quad \exists M \in \mathbb{R}, \exists \alpha > 0, \quad \forall n \in \mathbb{N}, \quad \|f^{(n)}\|_{\infty,[a,b]} \leq M n! \alpha^n.$$

Voir par exemple [QZ13, page 294].

40. Il semble que ce résultat a été prouvé par Peano en 1884 et par Borel en 1895.

2.7 Exercices

Exercice 2.45. Calculer le rayon de convergence de la série entière $z \mapsto \sum_{n\in\mathbb{N}} nz^{2n}$.

Exercice 2.46 (Formule de Hadamard)**.** Étant donné $(a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$, montrer que le rayon R_a de la série entière associée satisfait :

$$\frac{1}{R_a} = \limsup_{n\to+\infty} \left(|a_n|^{\frac{1}{n}}\right).$$

Exercice 2.47. Soit $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ deux suites complexes, et R_a, R_b les rayons des séries entières associées. Montrer

— si $a_n = O(b_n)$, alors $R_a \geq R_b$,

— si $a_n = o(b_n)$, alors $R_a \geq R_b$,

— si $a_n \sim b_n$, alors $R_a = R_b$.

Exercice 2.48. On utilise les notations de la Proposition 2.9. Montrer qu'on n'a pas forcément $R_d = \min\{R_a, R_b\}$, même si $R_a \neq R_b$.

Exercice 2.49. On pose $H_0 = 0$, et $\forall n \in \mathbb{N}^*, H_n = \sum_{k=1}^{n} \frac{1}{k}$, somme partielle de la série harmonique. Calculer le rayon de la série entière associée à $(H_n)_{n\in\mathbb{N}}$, et calculer explicitement $\sum_{n\in\mathbb{N}} H_n z^n$ quand z est dans le disque de convergence.

Exercice 2.50. Voir aussi [Min, page 23-24]. On peut étudier les fonctions cos et sin définies par (2.5).

1. Montrer que cos et sin sont C^∞ et $\sin' = \cos$, $\cos' = -\sin$.
2. Montrer qu'il existe un plus petit réel strictement positif t_0 tel que $\cos(t_0) = 0$. On peut définir $t_0 = \frac{\pi}{2}$.
3. Montrer que cos et sin sont 2π-périodiques et étudier leur parité.
4. Etudier les variations de cos et sin.
5. Soit $(x, y) \in \mathbb{R}^2$ tel que $x^2 + y^2 = 1$. Montrer qu'il existe $t \in [-\pi, \pi]$ tel que $x = \cos(t)$ et $y = \sin(t)$ et en déduire la surjectivité de $\exp : i\mathbb{R} \to \mathcal{U}$.

Exercice 2.51 (Log est continue)**.** Soit $z \in \mathbb{C} \setminus \mathbb{R}_-$.

1. Montrer que
$$\mathrm{Arg}(z) = \begin{cases} \arccos\left(\mathrm{Re}\,\frac{z}{|z|}\right) & \text{si } \mathrm{Im}(z) \geq 0 \\ -\arccos\left(\mathrm{Re}\,\frac{z}{|z|}\right) & \text{sinon.} \end{cases}$$
2. En déduire que les fonctions Arg et Log sont continues sur $\mathbb{C} \setminus \mathbb{R}_-$.

Exercice 2.52 (Attention !)**.** On pose $j = e^{\frac{2i\pi}{3}}$. Calculer $\mathrm{Log}(j)$ et $\mathrm{Log}(j^2)$, et constatez qu'on a

$$\mathrm{Log}(j^2) \neq 2\mathrm{Log}(j).$$

On n'a donc plus l'équation fonctionnelle satisfaite par Log, mais on garde l'expression de sa dérivée.

Exercice 2.53 (Dérivée de Log)**.**

1. Montrer[41] que Log est $\mathbb{C}$-dérivable sur $\mathbb{C} \setminus \mathbb{R}_-$, et que
$$\forall z \in \mathbb{C} \setminus \mathbb{R}_-, \ \ \mathrm{Log}'(z) = \frac{1}{z}.$$
2. Montrer que la série entière de terme général $\left((-1)^{n-1}\frac{z^n}{n}\right)_{n\in\mathbb{N}^*}$ a pour rayon de convergence 1, et calculer sa série dérivée.
3. En déduire
$$\forall z \in D(0,1), \ \ \mathrm{Log}(1+z) = \sum_{n\in\mathbb{N}^*} (-1)^{n-1}\frac{z^n}{n}.$$

41. On pourra se convaincre que la preuve de la proposition 3.39 du Chapitre I s'adapte sans changement. On utilisera en particulier la continuité de Log prouvée à l'exercice 2.51.

Exercice 2.54 (Il n'y a pas de fonction logarithme qui soit continue sur $\mathbb{C}^*$)**.** Supposons $L : \mathbb{C}^* \to \mathbb{C}$ continue telle que $\forall z \in \mathbb{C}^*, \exp(L(z)) = z$.

1. Montrer qu'il existe $k \in \mathbb{Z}$ tel que $\forall z \in \mathbb{C} \setminus \mathbb{R}_-, \quad L(z) = \mathrm{Log}(z) + 2ik\pi$.
2. En déduire une contradiction.

Exercice 2.55. Montrer que la fonction $z \in \mathbb{C}^* \mapsto \frac{1}{z}$ est analytique sur $\mathbb{C}^*$.

Exercice 2.56. Montrer que si $f : U \to \mathbb{C}$ est définie sur un ensemble $U \subset \mathbb{C}$, alors l'ensemble des points d'analyticité de f, c'est-à-dire l'ensemble des $z \in U$ autour desquels f admet un développement en série entière, est un ouvert.

Exercice 2.57. Donnez une preuve un tout petit peu alternative en montrant que l'ensemble

$$W = \{z \in U, \quad \forall n \in \mathbb{N}, f^{(n)}(z) = 0\}$$

est ouvert, fermé et non vide dans U.

Exercice 2.58 (Multiplicité d'une racine)**.** Soit $f : U \to \mathbb{C}$ analytique sur l'ouvert connexe U, et $z_0 \in U$ tel que $f(z_0) = 0$. On suppose que f n'est pas identiquement nulle. Montrer qu'il existe $p \in \mathbb{N}^*$ tel que

$$f(z) = (z - z_0)^p h(z)$$

au voisinage de z_0, et où h est analytique et ne s'annule pas au voisinage de z_0.

On dira que z_0 est une racine de multiplicité p, par analogie avec la situation des polynômes.

Exercice 2.59. Montrer que si U est connexe, et $f, g : U \to \mathbb{C}$ sont analytiques, alors

$$fg = 0 \Rightarrow \big[f = 0 \text{ ou } g = 0\big].$$

On dit que l'anneau des fonctions analytiques est intègre.

L'exercice suivant exhibe un contre-exemple classique :

Exercice 2.60 (Fonction C^∞ qui n'est pas égale à sa série de Taylor)**.** On pose

$$\forall x \in \mathbb{R}, \quad f(x) = \begin{cases} e^{-\frac{1}{x}} \text{ si } x > 0 \\ 0 \text{ sinon} \end{cases}$$

Montrer que f est de classe C^∞ sur $\mathbb{R}$, que la série de Taylor en 0 est de rayon infini, mais que la fonction f n'est pas développable en série entière en 0.

Exercice 2.61 (*Fonction C^∞ dont la série de Taylor a un rayon de convergence nul)**.** Cet exemple est tiré de [Gou08b, page 241]. On pose

$$\forall x \in \mathbb{R}, \quad f(x) = \int_0^{+\infty} \frac{e^{-t}}{1 + x^2 t} dt$$

1. On pose $\forall x \in \mathbb{R}_+, \varphi(x) = f(\sqrt{x})$. Montrer que φ est C^∞ sur $\mathbb{R}_+$ et calculer $\varphi^{(n)}$ pour tout $n \in \mathbb{N}$.
2. En déduire que f est C^∞ et que sa série de Taylor en 0 est de rayon de convergence nul.

Exercice 2.62 (*Une fonction C^∞ nulle part analytique)**.** On pose

$$\forall x \in \mathbb{R}, \quad f(x) = \sum_{n \in \mathbb{N}} e^{-\sqrt{2^n}} \cos(2^n x).$$

1. Montrer que f est C^∞.
2. Montrer que si $x = \frac{\pi p}{2^q}$ pour $(p, q) \in \mathbb{N}^2$, alors

$$\forall k \in \mathbb{N}, \quad f^{(2^k)}(x) \geq e^{-\sqrt{k}} k^k + O(q^k)$$

et en déduire le rayon de convergence de la série de Taylor de f en x.

3. Conclure que f n'est développable en série entière autour d'aucun point de $\mathbb{R}$.

Exercice 2.63 (*Théorème de Bernstein sur les séries entières)**.** On trouve la première question dans [QZ13, page 295], et la seconde dans [Gou08b, page 250] et [Pom94, 11.3.2]. Soit I un intervalle ouvert, et $f : I \to \mathbb{R}$ de classe C^∞.

1. On suppose
$$\forall x \in I, \forall n \in \mathbb{N}, \quad f^{(n)}(x) \geq 0.$$
Montrer que f est développable en série entière sur I.
2. On suppose désormais $I =]-a, a[$ et
$$\forall x \in]-a, a[, \forall n \in \mathbb{N}, \quad f^{(2n)}(x) \geq 0.$$
Montrer que f est développable en série entière autour de 0.[42]

Remarque 2.64. ** Le théorème de Bernstein s'intéresse à la caractérisation des transformées de Laplace des mesures de Borel positives sur $\mathbb{R}_+$. En effet, on peut montrer que si $f : \mathbb{R}_+^* \to \mathbb{R}$ est C^∞ sur $]0, +\infty[$, alors

$$\forall n \in \mathbb{N}, \ (-1)^n f^{(n)} \geq 0 \iff \exists g : \mathbb{R}_+ \to \mathbb{R} \text{ croissante telle que } \forall t \in]0, +\infty[, \ f(x) = \int_0^\infty e^{-tx} dg(x).$$

---

3 Fonctions holomorphes

Comme on l'a annoncé au paragraphe 1.2, nous allons nous concentrer sur la compréhension de la notion de dérivabilité au sens des complexes.

3.1 Définition

Définition 3.1. Soit U un ouvert[43] de $\mathbb{C}$, et $f : U \to \mathbb{C}$.

— On dit que f est $\mathbb{C}$-dérivable au point $z_0 \in U$ si la fonction
$$z \in U \setminus \{z_0\} \longmapsto \frac{f(z) - f(z_0)}{z - z_0}$$
admet une limite complexe en z_0. On notera alors $f'(z_0)$ cette limite.

— On dit que f est $\mathbb{C}$-dérivable sur U ou *holomorphe* sur U si elle est $\mathbb{C}$-dérivable en tout point de U. On peut alors définir $f' : U \to \mathbb{C}$ la dérivée de f.

Notez qu'on a utilisé la possibilité de diviser par un complexe non nul, c'est-à-dire la structure de corps de $\mathbb{C}$.

Exemple 3.2. Les polynômes sont des fonctions holomorphes sur $\mathbb{C}$. On a également vu (proposition 2.33) que les séries entières définissent des fonctions holomorphes sur leur disque de convergence. Finalement, les fonctions développables en série entière sur un ouvert U de $\mathbb{C}$, sont également holomorphes.

42. Indication : on commencera par traiter le cas de la fonction $F : x \mapsto f(x) + f(-x)$.
43. Attention, la notion de $\mathbb{C}$-dérivabilité n'a a priori pas de sens pour une fonction qui n'est pas définie sur un ouvert. Aussi, contrairement au cas de $\mathbb{R}$, on ne peut pas parler de dérivabilité à droite ou à gauche en un point.

3.2 Équations de Cauchy-Riemann

Etant donné un ouvert U de $\mathbb{C}$ et $f : U \to \mathbb{C}$, on peut voir U comme un sous-ensemble de $\mathbb{R}^2$ et f comme une fonction de cet ouvert dans $\mathbb{R}^2$. On discute ici la différence entre la $\mathbb{C}$-dérivabilité sur un ouvert de $\mathbb{C}$, et la différentiabilité sur ce même ouvert vu dans $\mathbb{R}^2$. On anticipe donc en utilisant des concepts du Chapitre VII ; mais nous n'irons pas très loin, et un lecteur familier avec la notion de dérivée partielle pourra avoir une idée assez bonne de ce qu'on discute ici (même si ce n'est pas suffisant : être différentiable en un point est un concept plus fort que d'avoir des dérivées partielles en ce point).

On obtient en fait la caractérisation suivante de la $\mathbb{C}$-dérivabilité :

Théorème 3.3. *Soit U un ouvert de $\mathbb{C}$, $z_0 = x_0 + iy_0 \in U$ (où $(x_0, y_0) \in \mathbb{R}^2$) et $f : U \to \mathbb{C}$. On note $P : \mathrm{Re}(f)$ et $Q = \mathrm{Im}(f)$ de sorte que $f = P + iQ$. Alors f est $\mathbb{C}$-dérivable en z_0 si et seulement si*

— *f est différentiable (au sens d'une fonction de $U \subset \mathbb{R}^2$ dans $\mathbb{R}^2$) en z_0*

— *et les dérivées partielles de f en z_0 satisfont*[44] *:*

$$\frac{\partial P}{\partial x}(x_0, y_0) = \frac{\partial Q}{\partial y}(x_0, y_0), \qquad et \qquad \frac{\partial P}{\partial y}(x_0, y_0) = -\frac{\partial Q}{\partial x}(x_0, y_0). \tag{3.1}$$

Dans ce cas, $df(z_0) : h \in \mathbb{C} \mapsto f'(z_0).h$ et $f'(z_0) = df(z_0)(1)$.

Remarque 3.4. On peut également écrire les deux équations de (3.1) par

$$\frac{\partial f}{\partial y}(z_0) = i\frac{\partial f}{\partial x}(z_0).$$

Il arrive qu'on définisse

$$\frac{\partial f}{\partial \bar{z}}(z_0) = \frac{1}{2}\left(\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y}\right)(z_0)$$

et (3.1) devient $\frac{\partial f}{\partial \bar{z}}(z_0) = 0$.

Remarque 3.5. Avec les notations du théorème 3.3, si f est différentiable en z_0, on a

$$df(z_0) = \begin{pmatrix} \frac{\partial P}{\partial x}(x_0, y_0) & \frac{\partial P}{\partial y}(x_0, y_0) \\ \frac{\partial Q}{\partial x}(x_0, y_0) & \frac{\partial Q}{\partial y}(x_0, y_0) \end{pmatrix}.$$

Cette matrice représente une application $\mathbb{R}$-linéaire de $\mathbb{R}^2$ dans $\mathbb{R}^2$. Les équations de Cauchy-Riemann sont équivalentes à dire que $df(z_0)$ vue de $\mathbb{C}$ dans $\mathbb{C}$ est une application $\mathbb{C}$-linéaire (si on veut, un élément de $M_{1,1}(\mathbb{C})$), qui est une condition plus forte que la $\mathbb{R}$-linéarité. Autrement dit on peut écrire $df(z_0)(h) = A.h$ où $A \in \mathbb{C}$, voir aussi la question 2 de l'Exercice 1.4.

Une telle transformation vue géométriquement est appelée une similitude directe[45], c'est-à-dire une composition entre une homothétie et une rotation, ce qui du point de vue du nombre complexe $A = \rho e^{i\theta}$ cité précédemment et qu'on a écrit sous forme trigonométrique, consiste à dire qu'on fait une homothétie de rapport $\rho = |A|$ et d'une rotation d'angle $\theta = \mathrm{Arg}(A)$.

Démonstration. Supposons f $\mathbb{C}$-dérivable en z_0. Alors de façon similaire à la remarque 3.36 du Chapitre I, on a le développement limité[46]

$$f(z_0 + h) = f(z_0) + f'(z_0)h + o(h)$$

44. Ces conditions sont appelées équations de Cauchy-Riemann.

45. Attention, rigoureusement l'application nulle n'est pas une similitude par définition, car on demande que l'homothétie ait un rapport strictement positif. Donc ici il faudrait dire que cette transformation est nulle ou une similitude directe.

46. C'est aussi le moment de remarquer que si $z \in \mathbb{C}$, alors son module est égal à la norme euclidienne de z vu comme élément de $\mathbb{R}^2$, donc la notation $o(h)$ est la même dans $\mathbb{C}$ et dans $\mathbb{R}^2$.

et comme l'application $h \mapsto f'(a)h$ est $\mathbb{R}$-linéaire (et continue[47]), f est bien différentiable en z_0 et $df(z_0) : h \in \mathbb{C} \mapsto f'(a).h$. On peut aussi constater

$$\frac{\partial f}{\partial x}(x_0, y_0) = df(z_0)(1) = -i.df(z_0)(i) = -i.\frac{\partial f}{\partial y}(x_0, y_0)$$

d'où les équations (3.1), voir aussi la remarque 3.4.

Réciproquement, si f est différentiable en z_0 alors

$$f(z_0 + h) = f(z_0) + df(z_0)(h) + o(h).$$

Attention, arrivé ici on ne peut pas diviser par h a priori, car on sait seulement que $df(z_0) : \mathbb{R}^2 \to \mathbb{R}^2$ est $\mathbb{R}$-linéaire. Mais si $h = x + iy$ avec $(x, y) \in \mathbb{R}^2$, avec la $\mathbb{R}$-linéarité combinée à (3.1) on peut écrire :

$$\begin{aligned} df(z_0)(h) &= xdf(z_0)(1) + ydf(z_0)(i) = x\frac{\partial f}{\partial x}(x_0, y_0) + y\frac{\partial f}{\partial y}(x_0, y_0) \\ &= x\frac{\partial f}{\partial x}(x_0, y_0) + iy\frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)(x + iy) \end{aligned}$$

et donc on peut écrire

$$f(z_0 + h) = f(z_0) + \frac{\partial f}{\partial x}(x_0, y_0).h + o(h),$$

d'où la $\mathbb{C}$-dérivabilité de f en z_0 et le fait que $f'(z_0) = \frac{\partial f}{\partial x}(x_0, y_0) = df(z_0)(1)$.

□

Corollaire 3.6. *Avec les notations du théorème 3.3, f est holomorphe sur U si et seulement si elle est différentiable en tout point $(x_0, y_0) \in U$ et*

$$\frac{\partial P}{\partial x} = \frac{\partial Q}{\partial y}, \qquad \text{et} \qquad \frac{\partial P}{\partial y} = -\frac{\partial Q}{\partial x} \qquad \text{sur } U. \tag{3.2}$$

3.3 Opérations et exemples

De façon analogue au cas réel (voir paragraphe 3.3.2 au chapitre I), on montre (U désigne un ouvert de $\mathbb{C}$) la stabilité des fonctions holomorphes par addition, multiplication par un complexe, produit, quotient si la fonction par laquelle on divise ne s'annule pas, et composée si les ensembles de définition sont compatibles, le tout avec les formules qui accompagnent ces énoncés, voir les propositions 3.37 et 3.38 du Chapitre I. En particulier l'espace des fonctions holomorphes sur U est une sous-algèbre de l'espace des fonctions continues de U dans $\mathbb{C}$.

Par exemple, à partir de l'exemple 3.2, on observe qu'une fraction rationnelle $\frac{P}{Q} \in \mathbb{C}(z)$ sera holomorphe en dehors de ses pôles, c'est-à-dire en dehors des points où Q s'annule (qui sont en nombre fini), comme quotient de deux fonctions holomorphes.

On conclut par un résultat qui sera généralisé au Chapitre VII :

Proposition 3.7. *Soit $f : U \to \mathbb{C}$ holomorphe sur un ouvert $U \subset \mathbb{C}$ connexe. Si $f' = 0$ sur U, alors f est constante.*

Démonstration. Commençons par montrer que f est localement constante : soit $z_0 \in U$, et $r > 0$ tel que $D(z_0, r) \subset U$. Soit $z \in D(z_0, r)$. On écrit $z_0 = x_0 + iy_0$ et $z = x + iy$, où $(x_0, y_0, x, y) \in \mathbb{R}^4$. On pose $\tilde{z} = x_0 + iy$. Il n'est pas dur de voir que $\tilde{z} \in U$, et donc les segments $[z_0, \tilde{z}]$ et $[\tilde{z}, z]$ sont aussi dans U. Comme f est différentiable, on a d'après le théorème 3.45 du Chapitre I,

$$\frac{\partial f}{\partial y} \equiv 0, \text{ donc } f(z_0) = f(x_0 + iy_0) = f(x_0 + iy) = f(\tilde{z})$$

47. Dans la définition de la différentiabilité, on demandera que l'application linéaire qui est candidate à être la différentielle, soit linéaire et continue. Quand l'espace de départ sera de dimension finie, comme ici, il arrivera que le mot "continue" passe à la trappe puisqu'une application définie et linéaire sur un espace de dimension finie est toujours continue, voir le paragraphe 2.5 du Chapitre III.

puis
$$\frac{\partial f}{\partial x} \equiv 0, \text{ donc } f(\widetilde{z}) = f(x_0 + iy) = f(x + iy) = f(z).$$
ce qui donne $f(z_0) = f(z)$ pour tout $z \in D(z_0, r)$, et ce pour tout z_0, r tel que $D(z_0, r) \subset U$.

Pour conclure, le cas $U = \emptyset$ étant trivial, considérons que U est non vide et donc il existe $z_0 \in U$. Nous posons $A = \{z \in U, f(z) = f(z_0)\}$. Par continuité de f, A est un fermé de U [48]. Et par ce qui précède, A est un ouvert. En effet, si $z_1 \in A$, alors il existe $\alpha > 0$ tel que $D(z_1, \alpha) \subset U$ (car U ouvert), et par l'étape précédente appliquée à ce disque, on a que $f(z) = f(z_1) = f(z_0)$ pour tout $z \in D(z_1, \alpha)$, d'où $D(z_1, \alpha) \subset A$ et le fait que A est ouvert. Par connexité de U, on obtient $A = U$, d'où le résultat. □

3.4 Formule de Cauchy

3.4.1 Intégration sur un chemin, indice

Définition 3.8. On appelle chemin [49] une application $\gamma : [a, b] \to \mathbb{C}$ qui est continue. On dira que ce chemin est C^1 par morceaux si γ est C^1 par morceaux [50] sur le segment $[a, b] \subset \mathbb{R}$. On dit que γ est un chemin fermé ou un lacet si $\gamma(a) = \gamma(b)$. On appelle image de γ l'ensemble $\mathrm{Im}(\gamma)$, qui est un compact de U [51].

Exemple 3.9 (Segment)**.** Etant donnés $(z_0, z_1) \in \mathbb{C}^2$, l'application
$$t \in [0, 1] \mapsto (1 - t)z_0 + tz_1$$
définit un chemin dont l'image est le segment orienté entre z_0 et z_1. On notera $[z_0, z_1]$ ce chemin.

Exemple 3.10 (Cercle)**.** Etant donné $z_0 \in \mathbb{C}$ et $r > 0$, l'application
$$t \in [0, 2\pi] \mapsto z_0 + re^{it}$$
définit un chemin dont l'image est le cercle de centre z_0 et de rayon r, parcouru dans le sens trigonométrique. On notera $C(z_0, r)$ ce chemin.

Exemple 3.11 (Concaténation de chemins)**.** Etant donnés $\gamma_1 : [a, b] \to U$ et $\gamma_2 : [c, d] \to U$ sont deux chemins à valeurs dans U tels que $\gamma_1(b) = \gamma_2(c)$. Alors on peut construire un chemin *composé* [52] de γ_1 et γ_2 comme étant le chemin [53] $\gamma : [a, b + d - c] \to U$ par :
$$\forall t \in [a, b + d - c], \gamma(t) = \begin{cases} \gamma_1(t) \text{ si } t \in [a, b] \\ \gamma_2(t + c - b) \text{ si } t \in [b, b + d - c]. \end{cases}$$
Si γ_1 et γ_2 sont C^1 par morceaux, alors γ l'est aussi.

Exemple 3.12. Si $\gamma : [a, b] \to U$ est un chemin dans U, on peut définir le chemin opposé $\widetilde{\gamma} : [-b, -a] \to U$ par $\forall t \in [-b, -a], \widetilde{\gamma}(t) = \gamma(-t)$: il s'agit du même chemin mais parcouru dans le sens opposé.

Définition 3.13. Soit $\gamma : [a, b] \to \mathbb{C}$ un chemin C^1 par morceaux à valeurs dans U, et $f : U \to \mathbb{C}$ continue. On définit l'intégrale de f sur γ par [54]
$$\int_\gamma f(z)dz := \int_a^b f(\gamma(t))\gamma'(t)dt.$$

48. Pour voir cela, vous pouvez connaître le résultat du Chapitre III qui dit que l'image réciproque d'un fermé par une fonction continue est un fermé, mais vous pouvez aussi simplement revenir à la définition.

49. Comme tous les chemins qu'on considère ici seront orientés, on a choisi de ne pas dire "chemin orienté".

50. Cela signifie que chaque fonction coordonnée est continue et C^1 par morceaux, voir aussi la remarque 3.32 du Chapitre I. Notez qu'on inclut dans le mot "chemin" le caractère continu de la fonction γ, car on considère le mot suffisamment explicite. On ne dira donc pas "chemin continu et C^1 par morceaux".

51. Car d'après le paragraphe 1.5 du Chapitre III, l'image d'un compact par une application continue est un compact.

52. J'aurais préféré qu'on dise concaténé, car il ne s'agit pas d'une composition au sens des fonctions, mais je respecte l'usage qui m'a semblé le plus répandu.

53. Vérifiez que l'on maintient la continuité dans ce recollement, ce qui autorise à parler de chemin.

54. Attention, ici on considère ce qu'on appelle une intégrale curviligne, c'est-à-dire l'intégrale d'une fonction le long d'un chemin. Notez qu'on utilise en fait l'intégrale de Riemann vu que le membre de droite est une fonction continue par

Remarque 3.14. Le choix de la paramétrisation n'a pas d'effet sur l'intégrale sur un chemin, à condition de respecter l'orientation : en effet si $\varphi : [c,d] \to [a,b]$ est C^1 et telle que $\varphi(c) = a$, $\varphi(d) = b$, alors
$$\int_{\gamma\circ\varphi} f(z)dz = \int_c^d f(\gamma(\varphi(t)))\gamma'(\varphi(t))\varphi'(t)dt = \int_a^b f(\gamma(u))\gamma'(u)du = \int_\gamma f(z)dz.$$

Remarque 3.15. Si $\gamma_1 : [a,b] \to U$ et $\gamma_2 : [c,d] \to U$ sont deux chemins C^1 par morceaux tels que $\gamma_1(b) = \gamma_2(c)$, on note γ le chemin composé (ou concaténation) de γ_1 et γ_2 et on vérifie que
$$\int_\gamma f(z)dz = \int_{\gamma_1} f(z)dz + \int_{\gamma_2} f(z)dz.$$
De plus, si $\gamma : [a,b] \to U$ est un chemin et $\widetilde{\gamma}$ sont chemin opposé, alors
$$\int_{\widetilde{\gamma}} f(z)dz = -\int_\gamma f(z)dz.$$

Exemple 3.16 (Intégrale sur un segment et sur un cercle)**.** Si γ est le segment $[z_0, z_1]$ (supposé dans U), alors
$$\int_{[z_0,z_1]} f(z)dz = \int_0^1 f((1-t)z_0 + tz_1).(z_1 - z_0)dt.$$
Si γ est le cercle $C(z_0, r) \subset U$, alors
$$\int_{C(z_0,r)} f(z)dz = \int_0^{2\pi} f(z_0 + re^{it})rie^{it}dt.$$

Remarque 3.17. Si $\gamma : [a,b] \to \mathbb{C}$ est un chemin C^1 par morceaux, on définit sa longueur par
$$\mathrm{Long}(\gamma) = \int_a^b |\gamma'(t)|dt.$$
Pour ce convaincre qu'il s'agit bien de la longueur intuitive, on vérifie que c'est une notion invariante par reparamétrisation, additive pour les concaténations, et enfin que la longueur du segment $[z_0, z_1]$ est $|z_1 - z_0|$. De ces faits on déduit que $\mathrm{Long}(\gamma)$ désigne bien la longueur de γ si celui-ci est affine par morceaux, et on conclut cette heuristique par le fait que tout chemin C^1 par morceaux peut être approché par un chemin affine par morceaux (voir le lemme 4.40, même si ce n'est pas suffisant ici, car il faudrait faire une "approximation C^1" ; notre but n'est pas de donner un énoncé précis ici, juste de convaincre le lecteur de la légitimité d'une telle notation). Concluons par l'inégalité suivante qu'on utilisera de temps en temps :
$$\left|\int_\gamma f(z)dz\right| \leq \int_a^b |f(\gamma(t))\gamma'(t)|dt \leq \left(\sup_{\mathrm{Im}(\gamma)} |f|\right).\mathrm{Long}(\gamma),$$
et qui implique en particulier que si f_n converge uniformément vers f sur $\mathrm{Im}(\gamma)$, alors $\int_\gamma f_n(z)dz$ converge vers $\int_\gamma f(z)dz$.

Définition-Proposition 3.18. Soit $\gamma : [a,b] \to \mathbb{C}$ un lacet C^1 par morceaux et $z \in \mathbb{C} \setminus \mathrm{Im}(\gamma)$. On pose
$$\mathrm{ind}_\gamma(z) := \frac{1}{2i\pi}\int_\gamma \frac{1}{\zeta - z}d\zeta = \frac{1}{2i\pi}\int_a^b \frac{\gamma'(t)}{\gamma(t) - z}dt.$$
L'application $ind_\gamma : \mathbb{C} \setminus \mathrm{Im}(\gamma) \to \mathbb{C}$ est continue et à valeurs dans $\mathbb{Z}$.

morceaux d'une variable réelle, et à valeurs dans $\mathbb{C}$. La notation dz ne doit donc pas être confondue avec une intégrale multiple sur un ensemble de $\mathbb{C}$. D'ailleurs, dans le corollaire 3.39 (et seulement à cet endroit), on utilisera une intégrale multiple, et la notation dz renverra cette fois-là à la mesure de Lebesgue dans $\mathbb{R}^2$, c'est-à-dire qu'on pourrait noter $dxdy$ si on voit $z = x + iy \in \mathbb{R}^2$. La raison pour laquelle il ne devrait pas y avoir de confusion entre les deux objects tient à la nature de l'ensemble sur lequel on intègre ; s'il s'agit d'un chemin, alors on est dans une intégrale curviligne comme définie ici, s'il s'agit d'un ensemble "surfacique", comme par exemple un disque, alors on est sur un intégrale multiple.

Remarque 3.19. Intuitivement, $\mathrm{ind}_\gamma(z)$ décrit le nombres de tours (orientés dans le sens trigonométriques) que la courbe décrite par γ fait autour de z. L'exposant -1 n'est pas un choix anodin, voir l'exercice 3.62 et la proposition 3.22.

Remarque 3.20. Comme l'image d'un connexe par une fonction continue est connexe (voir paragraphe 1.6 du Chapitre III), la fonction ind_γ est constante sur chaque composante connexe de $\mathbb{C} \setminus \mathrm{Im}(\gamma)$.

Démonstration. Soit $\varepsilon > 0$. On pose $A_\varepsilon = \{z \in \mathbb{C}, d(z, \mathrm{Im}(\gamma)) > \varepsilon\}$ [55]. Pour $(z, z') \in A_\varepsilon$, on a $|\zeta - z| > \varepsilon$ et $|\zeta - z'| > \varepsilon$ pour tout $\zeta \in \mathrm{Im}(\gamma)$. On obtient ainsi

$$\forall (z, z') \in A_\varepsilon, \quad |\mathrm{ind}_\gamma(z') - \mathrm{ind}_\gamma(z)| \leq \frac{\mathrm{Long}(\gamma)}{2\pi\varepsilon^2}|z' - z|$$

ce qui prouve le caractère lipschitz et donc la continuité de ind_γ sur A_ε. Comme $\cup_{\varepsilon>0} A_\varepsilon = \mathbb{C} \setminus \mathrm{Im}(\gamma)$ [56], on a bien prouvé la continuité de ind_γ [57].

Pour voir que ind_γ est à valeurs dans $\mathbb{Z}$, on introduit $z_0 \in \mathbb{C} \setminus \mathrm{Im}(\gamma)$ et on va vérifier que $\exp(2i\pi \mathrm{ind}_\gamma(z_0)) = 1$. On définit

$$\begin{array}{rccl} \varphi : & [a, b] & \longrightarrow & \mathbb{C} \\ & t & \longmapsto & \exp\left(\displaystyle\int_a^t \frac{\gamma'(s)}{\gamma(s) - z_0} ds\right). \end{array}$$

Comme γ est C^1 par morceaux, φ est continue et C^1 par morceaux (elle fait intervenir la primitive d'une fonction continue par morceaux), et sur chaque segment où elle est C^1, on a

$$\varphi'(t) = \frac{\gamma'(t)}{\gamma(t) - z_0}\varphi(t), \qquad \text{i.e.} \qquad \left(\frac{\varphi}{\gamma - z_0}\right)' = 0$$

donc $t \mapsto \frac{\varphi}{\gamma - z_0}$ est constante sur chaque intervalle où elle est C^1, et comme elle est continue, elle est finalement constante sur $[a, b]$, d'où avec $\gamma(a) = \gamma(b)$, $\varphi(b) = \varphi(a) = 1$, ce qu'il fallait démontrer. □

3.4.2 Primitive et formule de Cauchy

Comme on l'a fait pour les intégrales réelles, on cherche à trouver des primitives de f pour calculer son intégrale sur un chemin. En effet, on a le résultat élémentaire suivant :

Proposition 3.21. *Soit U un ouvert de $\mathbb{C}$, $\gamma : [a, b] \to U$ un chemin C^1 par morceaux et $f : U \to \mathbb{C}$ continue. Si F est une primitive de f sur U, c'est-à-dire que $F : U \to \mathbb{C}$ est $\mathbb{C}$-dérivable et $F' = f$, alors*

$$\int_\gamma f(z)dz = F(\gamma(b)) - F(\gamma(a)).$$

En particulier $\int_\gamma f(z)dz = 0$ si γ est un lacet.

Démonstration. Il suffit d'écrire le calcul suivant :

$$\int_\gamma f(z)dz = \int_a^b f(\gamma(t))\gamma'(t)dt = \int_a^b (F \circ \gamma)'(t)dt = F \circ \gamma(b) - F \circ \gamma(a).$$

□

Revenons sur l'exemple fondamental de l'exercice 3.62 qu'on avait promis d'éclaircir :

55. On peut préférer introduire $z_0 \in U \setminus \mathrm{Im}(\gamma)$ et trouver $r_0 > 0$ tel que ind_γ sera lipschitzienne sur $D(z_0, r_0)$, avec un calcul similaire.
56. Voir l'exercice 1.6.
57. On aurait pu utiliser un résultat sur les intégrales à paramètre. Attention a priori, les résultats de l'Exercice 5.28 du Chapitre I ne s'appliquent pas directement, il faudrait appliquer un résultat du Chapitre IV. C'est un bon réflexe à avoir, et la preuve fonctionne très bien, même si on propose ici un calcul sans doute plus élémentaire (et tiré de [Tau06, Théorème 6.3.2], mais attention, la rédaction y est à mon avis confuse car la lettre d (qui remplace notre ε) est introduite après (z, z'), alors qu'il aurait fallu l'introduire avant comme on l'a fait ici).

Proposition 3.22. *Soit $n \in \mathbb{Z} \setminus \{-1\}$, γ un lacet C^1 par morceaux, et $z_0 \in \mathbb{C}$. Si*
- *$n \geq 0$,*
- *ou si $n \leq -2$ et $z_0 \notin \mathrm{Im}(\gamma)$*

alors

$$\int_\gamma (z - z_0)^n dz = 0$$

Démonstration. Cela vient de la proposition 3.21 car si $z \mapsto \frac{(z-z_0)^n}{n+1}$ est une primitive de $z \mapsto z^n$ sur $\mathbb{C}$ si $n \geq 0$, et sur $\mathbb{C} \setminus \{z_0\}$ si $n \leq -2$. □

Le dernier énoncé de la proposition 3.21 est en fait une caractérisation de l'existence d'une primitive : en effet

Théorème 3.23. *Soit U un ouvert de $\mathbb{C}$ et $f : U \to \mathbb{C}$ continue. Alors*

$$\Big[\ \textit{Il existe une primitive de } f \textit{ sur } U\ \Big] \Longleftrightarrow \Big[\ \textit{Pour tout lacet } \gamma\ \ C^1 \textit{ par morceaux, } \int_\gamma f(z)dz = 0\ \Big].$$

Démonstration. On a déjà prouvé le sens $\Rightarrow$ (Proposition 3.21), on s'intéresse donc à la réciproque. Les composantes connexes de U étant connexes et disjointes, il suffit de construire une primitive de f sur chacune d'elle, donc on peut supposer U connexe, et donc connexe par arcs vu qu'il est ouvert (voir proposition 1.3). On fixe $z_0 \in U$: pour tout $z \in U$, il existe $\gamma : [a, b] \to U$ qui joint z_0 à z. On définit alors

$$F(z) = \int_\gamma f(\zeta)d\zeta.$$

Ceci est légitime car on va voir que ça ne dépend pas du chemin utilisé pour joindre z_0 à z. En effet, si γ_1 et γ_2 sont deux chemins joignant z_0 et z, on considère γ la concaténation de γ_1 et de l'opposé de γ_2 consitute un lacet de U, et donc par hypothèse on a

$$0 = \int_\gamma f(\zeta)d\zeta = \int_{\gamma_1} f(\zeta)d\zeta - \int_{\gamma_2} f(\zeta)d\zeta$$

ce qui fournit bien l'égalité attendue. Il reste donc à montrer que F est $\mathbb{C}$-dérivable en z et que $F'(z) = f(z)$. Soit donc $r > 0$ tel que $D(z, r) \subset U$, et $h \in \mathbb{C}^*$ tel que $|h| < r$. On considère γ_z et γ_{z+h} deux chemins qui joignent z_0 avec z et $z + h$ respectivement. En considérant le lacet γ qui est la concaténation de γ_z, du segment $[z, z + h]$ et du chemin opposé à γ_{z+h}, on obtient

$$0 = \int_{\gamma_z} f(\zeta)d\zeta + \int_{[z,z+h]} f(\zeta)d\zeta - \int_{\gamma_{z+h}} f(\zeta)d\zeta$$

d'où

$$F(z + h) - F(z) = \int_{[z,z+h]} f(\zeta)d\zeta.$$

Constatons désormais que $\int_{[z,z+h]} 1d\zeta = h$, et donc on peut écrire :

$$\frac{F(z + h) - F(z)}{h} - f(z) = \frac{1}{h} \int_{[z,z+h]} \big[f(\zeta) - f(z)\big]d\zeta.$$

Comme f est continue en z, pour $\varepsilon > 0$ il existe $\alpha > 0$ tel que $D(z, \alpha) \subset U$ et $\forall \zeta \in D(z, \alpha)$, $|f(\zeta) - f(z)| \leq \varepsilon$. Si on considère la construction précédente avec $r = \alpha$, on obtient

$$\forall h \in D(0, \alpha), \ \left|\frac{F(z + h) - F(z)}{h} - f(z)\right| \leq \left|\frac{Long([z, z + h])}{h}\right| \varepsilon = \varepsilon,$$

d'où le résultat. □

Nous allons maintenant voir que sous certaines conditions géométriques sur les ensembles de définition, les fonctions holomorphes ont des primitives. Ceci aura pour conséquence la formule de Cauchy, qui est le résultat le plus important de ce chapitre.

Théorème 3.24 (Théorème et formule de Cauchy, première version)**.** *Soit U un ouvert* **étoilé** *de $\mathbb{C}$, et $f : U \to \mathbb{C}$ holomorphe*[58]*. Alors*

— *f admet une primitive sur U ;*

— *si γ est un lacet C^1 par morceaux et à valeurs dans U, alors $\int_\gamma f(z)dz = 0$;*

— *si de plus $z \in U \setminus \mathrm{Im}(\gamma)$, alors*

$$f(z)\mathrm{ind}_\gamma(z) = \frac{1}{2i\pi}\int_\gamma \frac{f(\zeta)}{\zeta - z}d\zeta.$$

Ce théorème est assez délicat à démontrer ; on prouve ci-dessous la fin de l'énoncé, et au paragraphe 4.7 on en prouvera la première partie.

Remarque 3.25. Remarquez qu'on a fait une hypothèse sur la nature de U, à savoir son caractère étoilé, ce qui signifie qu'il existe z_0 tel que[59]

$$\forall z \in U, \quad [z_0, z] \subset U.$$

Il s'agit d'une hypothèse plus faible que la convexité (en fait un domaine est convexe s'il est étoilé par rapport à tous ses points). Au paragraphe 4.8, on verra qu'on peut affaiblir encore l'hypothèse. Mais néanmoins ce résultat n'est pas valable sans hypothèse. Et pour cause, sinon on pourrait appliquer le résultat à la fonction $\zeta \in U \setminus \{z\} \mapsto \frac{f(\zeta)}{\zeta - z}$, ce qui donnerait $f(z)\mathrm{ind}_\gamma(z) = 0$, ce qui ne peut pas être vrai si par exemple γ est d'indice 1 (un cercle par exemple). On vient donc de voir que le résultat n'est pas valable pour un ouvert du type $U \setminus \{z\}$ et un chemin qui "tourne" autour de z. Encore une fois, on invite le lecteur à comparer avec l'exercice 3.62.

Notons que même si on a fait une hypothèse sur U, cela suffira à obtenir les conséquences les plus importantes de la formule de Cauchy. On l'appliquera souvent dans des disques inclus dans U (qui sont bien convexes), ce qui explique qu'on qualifie parfois de "local" l'énoncé précédent.

Preuve de la dernière partie du théorème 3.24 : Définissons

$$\begin{array}{rccl} g : & U & \longrightarrow & \mathbb{C} \\ & \zeta & \longmapsto & \begin{cases} \frac{f(\zeta)-f(z)}{\zeta - z} & \text{si } \zeta \neq z \\ f'(z) & \text{sinon} \end{cases} \end{array} \tag{3.3}$$

Cette fonction est continue sur U et holomorphe sur $U \setminus \{z\}$, donc par la première partie de l'énoncé et la note 58, on a $\int_\gamma g(\zeta)d\zeta = 0$, ce qui donne

$$\int_\gamma \frac{f(\zeta)}{\zeta - z}d\zeta = \int_\gamma \frac{f(z)}{\zeta - z}d\zeta = 2i\pi f(z)\mathrm{ind}_\gamma(z)$$

□

Corollaire 3.26 (Formule de Cauchy sur les cercles)**.** *Soit U un ouvert de $\mathbb{C}$, et $f : U \to \mathbb{C}$. Soit $z_0 \in U$ et $r > 0$ tel que le disque fermé $\overline{D}(z_0, r)$ est inclus dans U. Alors*

$$\forall z \in D(z_0, r), \quad f(z) = \frac{1}{2i\pi}\int_{C(z_0,r)} \frac{f(\zeta)}{\zeta - z}d\zeta = \frac{1}{2\pi}\int_0^{2\pi} \frac{f(z_0 + re^{it})}{z_0 + re^{it} - z} re^{it}dt. \tag{3.4}$$

Remarque 3.27. Notez bien qu'on n'a pas fait d'hypothèse sur U, mais on a supposé que le disque (et pas seulement le cercle) considéré était inclus dans U.

Démonstration. Comme $\overline{D}(z_0, r) \subset U$, il existe $\varepsilon > 0$ tel que $D(z_0, r + \varepsilon) \subset U$[60]. On peut donc appliquer le théorème 3.24 dans le convexe $D(z_0, r + \varepsilon)$, appliquer la formule de Cauchy sur le cercle $C(z_0, r)$ pour lequel l'indice de $z \in D(z_0, r)$ est $+1$. □

58. On verra un résultat un peu plus fort, à savoir qu'il suffit de supposer f continue sur U et holomorphe sur U privé d'un point ; cela sera particulièrement important pour la conséquence finale de l'énoncé.
59. On dit que le domaine est étoilé par rapport à z_0.
60. Voir l'exercice 1.6.

3.5 Conséquences de la formule de Cauchy

Analyticité des fonctions holomorphes : La conséquence la plus célèbre du théorème 3.24[61] est la suivante :

Théorème 3.28. *Soit U un ouvert de $\mathbb{C}$, $f : U \to \mathbb{C}$ holomorphe. Alors f est analytique sur U, et le développement en série entière de f en $z_0 \in U$ a un rayon de convergence supérieur ou égal à $d(z_0, \mathbb{C} \setminus U)$*[62].

Remarque 3.29. Ainsi, combiné avec la proposition 2.13, on obtient finalement qu'une fonction $f : U \to \mathbb{C}$ est analytique sur U si et seulement si elle est holomorphe sur U.

Démonstration. Soit $z_0 \in U$, et $r < d(z_0, \mathbb{C} \setminus U)$. Comme $D(z_0, r) \subset U$, par le corollaire 3.26,

$$\forall z \in D(z_0, r), \quad f(z) = \frac{1}{2i\pi} \int_{C(z_0,r)} \frac{f(\zeta)}{\zeta - z} d\zeta.$$

Fixons $z \in D(z_0, r)$:

$$\forall \zeta \in C(z_0, r), \frac{f(\zeta)}{\zeta - z} = \frac{f(\zeta)}{\zeta - z_0} . \frac{1}{1 - \frac{z - z_0}{\zeta - z_0}} = \frac{f(\zeta)}{\zeta - z_0} . \sum_{n \in \mathbb{N}} \left(\frac{z - z_0}{\zeta - z_0} \right)^n . \tag{3.5}$$

Cette dernière série converge normalement (en la variable ζ) sur $C(z_0, r)$ car

$$\sup_{\zeta \in C(z_0,r)} \left| \frac{z - z_0}{\zeta - z_0} \right|^n = \left(\frac{|z - z_0|}{r} \right)^n .$$

On peut donc intervertir l'intégrale et la somme (voir l'exercice 5.22), pour obtenir :

$$\forall z \in D(z_0, r), \; f(z) = \sum_{n \in \mathbb{N}} \left(\frac{1}{2i\pi} \int_{C(z_0,r)} \frac{1}{(\zeta - z_0)^{n+1}} d\zeta \right) (z - z_0)^n.$$

Ceci prouve bien que f est développable en série entière autour de z_0. Notez que les coefficients qui apparaissent ne dépendent en fait pas de r, par unicité du développement en série entière. Comme la formule précédente est valable pour tout $r < d(z_0, \mathbb{C} \setminus U)$, le rayon de convergence de cette série est donc supérieur ou égal à $d(z_0, \mathbb{C} \setminus U)$. □

Remarque 3.30. On obtient en prime une expression intégrale du coefficient dans le développement en série entière de f autour de z_0, à savoir si $f(z) = \sum_{n \in \mathbb{N}} a_n (z - z_0)^n$,

$$a_n = \frac{1}{2i\pi} \int_{C(z_0,r)} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta.$$

Notez qu'on peut retrouver cette formule (au moins formellement, voir la remarque 3.45) en dérivant n fois la formule (3.4), puisque $a_n = \frac{f^{(n)}(z_0)}{n!}$.

Remarque 3.31. Une façon de comprendre l'importance de la fin de l'énoncé du théorème 3.28 est de voir que si une fonction $f : \mathbb{C} \to \mathbb{C}$ est analytique (c'est-à-dire qu'en tout point il y a un développement en série entière de rayon supposé strictement positif), alors en fait le rayon du développement en série entière en tout point sera $+\infty$! Cela motive notamment l'étude des fonctions développables en série entière sur $\mathbb{C}$ plutôt que sur $\mathbb{R}$. Prenons par exemple la fonction $x \in \mathbb{R} \mapsto \frac{1}{1+x^2}$ qui est développable en série entière sur $\mathbb{R}$. Si on veut comprendre le rayon de convergence du développement en série entière autour d'un point (réel ou complexe, peu importe), il faut tracer le plus grand cercle centré en ce point et qui ne contient pas les points $i, -i$ qui sont les deux pôles de cette fonction vue comme une fonction définie sur un sous-ensemble de $\mathbb{C}$. Par exemple, si on développe autour de 2, alors le rayon sera $|2 - i| = \sqrt{5}$.

61. En fait on n'utilisera que le corollaire 3.26.

62. Ne négligez pas cette deuxième partie de l'énoncé, qui se comprend très bien par un dessin. C'est en effet une partie étonnante du résultat, que nous n'avions pas présagée dans l'étude des fonctions analytiques au paragraphe 2.6.

Une conséquence du théorème 3.28 est le fait que si f est holomorphe en dehors d'un point, mais a un comportement raisonnable au voisinage de ce point, alors f se prolonge de façon holomorphe en ce point [63].

Proposition 3.32. *Soit U un ouvert de $\mathbb{C}$ et $z_0 \in U$. On suppose que $f : U \setminus \{z_0\} \to \mathbb{C}$ est holomorphe sur $U \setminus \{z_0\}$ et bornée sur un voisinage [64] de z_0, alors f se prolonge de façon holomorphe sur tout U.*

Remarque 3.33. Un cas particulier classique est le suivant : si $f : U \to \mathbb{C}$ est continue et holomorphe sur $U \setminus \{z_0\}$, alors f est holomorphe sur tout U. Ainsi les hypothèses raffinées du lemme de Goursat paraissent inutiles. Mais bien sûr, il ne faut pas raisonner à l'envers. On a utilisé le lemme de Goursat pour prouver la formule de Cauchy et donc le théorème 3.28, qu'on utilise pour prouver la proposition 3.32. On a donc bien utilisé ce cas raffiné pour arriver jusque là !

Tant qu'on y est, précisons que le cas particulier cité au début de cette remarque a une autre preuve : par le raffinement du lemme 4.29 de Goursat, f est d'intégrale nulle sur les triangles, donc pas le théorème 4.32 de Morera, elle est holomorphe.

Remarque 3.34. On précisera les comportements possibles d'une fonction f holomorphe en dehors d'un point au paragraphe 3.8.

Démonstration. On pose

$$\forall z \in U, \quad g(z) = \begin{cases} (z-z_0)^2 f(z) & \text{si } z \neq z_0 \\ 0 & \text{si } z = z_0 \end{cases}$$

Par les opérations classiques, g est holomorphe sur $U \setminus \{z_0\}$, et comme f est bornée, on a

$$\frac{g(z_0+h) - g(z_0)}{h} = hf(z_0+h) \xrightarrow[h\to 0]{} 0$$

donc g est $\mathbb{C}$-dérivable en z_0 et $g'(z_0) = 0$. Par le théorème 3.28, g est développable en série entière au voisinage de z_0, et comme $g(z_0) = g'(z_0) = 0$, il existe $(a_n)_{n\geq 2}$ tels que $g(z) = \sum_{n\geq 2} a_n (z-z_0)^n$ sur ce voisinage. Ainsi, si $z \neq z_0$ est dans ce voisinage, $f(z) = \sum_{n\geq 2} a_n (z-z_0)^{n-2}$ donc en posant $f(z_0) = a_2$, on a f [65] analytique (et donc holomorphe) sur ce voisinage de z_0. □

Inégalités de Cauchy : La formule de Cauchy donne une représentation intégrale de $z \mapsto f(z)$. Notez bien que dans l'intégrale, f apparait via la variable d'intégration seulement. Donc quand on dérive cette formule par rapport à z, on obtient une représentation intégrale de f' qui dépend seulement de f. Ceci est très fort, car cela implique qu'on peut contrôler f' à partir de f. Ce phénomène est extrêmement différent du cas des fonctions dérivables réelles, pour lesquelles f' peut avoir un comportement très chaotique en comparaison du comportement de f (je pense à une fonction du type $x \mapsto x^3 \sin(\frac{1}{x})$ au voisinage de 0 dans $\mathbb{R}$).

Voici une façon d'exprimer cette remarque :

Proposition 3.35 (Inégalités de Cauchy)**.** *Soit U un ouvert de $\mathbb{C}$, $f : U \to \mathbb{C}$ holomorphe. Soit $z_0 \in U$ et $r > 0$ tel que $\overline{D}(z_0, r) \subset U$. Alors*

$$\forall n \in \mathbb{N}, \quad |f^{(n)}(z_0)| \leq \frac{n!}{r^n} \sup_{C(z_0,r)} |f|. \tag{3.6}$$

Démonstration. Soit $n \in \mathbb{N}$. Via la remarque 3.30, on a

$$f^{(n)}(z_0) = \frac{n!}{2i\pi} \int_{C(z_0,r)} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta$$

63. Encore une fois, on voit le comportement radicalement différent entre les fonctions $\mathbb{R}$-dérivables et $\mathbb{C}$-dérivables ; ici dans le cas complexe, une seule condition sur f, va permettre de prolonger f de façon dérivable en le point singulier, alors que dans le cas réel, on aura besoin de condition sur f'.

64. Attention, f n'est pas supposée définie en z_0, donc ce voisinage est épointé, par exemple de la forme $D(z_0, \varepsilon) \setminus \{z_0\}$.

65. Attention à l'abus de notation, on a encore noté f le prolongement ; le lecteur plus rigoureux et/ou perturbé par un tel abus pourra prendre une autre notation.

d'où
$$|f^{(n)}(z_0)| \leq \frac{n!}{2\pi}\int_0^{2\pi}\frac{|f(z_0+re^{it})}{r^{n+1}}rdt \leq \frac{n!}{r^n}\sup_{\zeta\in C(z_0,r)}|f(\zeta)|.$$

□

Corollaire 3.36 (Théorème de Liouville). *Une fonction holomorphe sur $\mathbb{C}$ qui est bornée est constante.*

Démonstration. Soit f holomorphe sur $\mathbb{C}$, telle que $\forall z \in \mathbb{C}, |f(z)| \leq M$ pour $M \in \mathbb{R}$. Par l'inégalité de Cauchy pour $n = 1$ sur tout disque $D(z,r) \subset \mathbb{C}$, on obtient

$$\forall z \in \mathbb{C},\quad \forall r > 0,\quad |f'(z)| \leq \frac{1}{r}M$$

ce qui donne avec $r \to +\infty$, $f'(z) = 0$ pour tout $z \in \mathbb{C}$. Ainsi, $\mathbb{C}$ étant connexe, par la proposition 3.7, f est constante. □

Corollaire 3.37 (Théorème de d'Alembert-Gauss). *Soit P un polynôme complexe non constant. Alors P admet au moins une racine dans $\mathbb{C}$.*

Démonstration. Supposons par l'absurde que P est un polynôme non constant qui ne s'annule pas sur $\mathbb{C}$. On peut alors considérer la fonction $\frac{1}{P}$ qui est holomorphe sur $\mathbb{C}$. Comme P est de degré supérieur ou égal à 1, on voit que $|P(z)| \xrightarrow[|z|\to+\infty]{} +\infty$, donc f tend vers 0 quand $|z| \to +\infty$. Or une fonction continue qui tend vers 0 à l'infini est bornée [66]. Cela implique par le théorème de Liouville que f est constante, ce qui est une contradiction. □

Remarque 3.38. Remarquons qu'avec le théorème de Parseval qu'on verra au paragraphe 1 du Chapitre VIII, on peut affiner (3.6) : en effet, pour $r > 0$ tel que $\overline{D}(z_0,r) \subset U$ l'égalité

$$\forall t \in \mathbb{R},\qquad f(z_0+re^{it}) = \sum_{n\in\mathbb{N}}\frac{f^{(n)}(z_0)}{n!}r^ne^{int}$$

peut être vue comme un développement en série de Fourier de la fonction $t \mapsto f(z_0+re^{it})$, et donc par l'égalité de Parseval,
$$\sum_{n\in\mathbb{N}}\left|\frac{f^{(n)}(z_0)}{n!}\right|^2 r^{2n} = \frac{1}{2\pi}\int_0^{2\pi}|f(z_0+re^{it})|^2dt. \tag{3.7}$$
Comme
$$\frac{1}{2\pi}\int_0^{2\pi}|f(z_0+re^{it})|^2dt \leq \left(\sup_{C(z_0,r)}|f|\right)^2$$
cette égalité implique bien (3.6).

Propriété de la moyenne et principe du maximum : Étant donnée une fonction $f : U \to \mathbb{C}$ holomorphe et z_0, r tels que $\overline{D}(z_0,r) \subset U$, en écrivant la formule (3.4) pour $z_0 = z$, on obtient

$$f(z_0) = \frac{1}{2\pi}\int_0^{2\pi}f(z_0+re^{it})dt$$

qui exprime que $f(z_0)$ est égal à la moyenne de f sur un cercle de rayon r. On peut obtenir le même résultat sur un disque :

Corollaire 3.39. *Soit U un ouvert de $\mathbb{C}$ contenant le disque $\overline{D}(z_0,r)$, et $f : U \to \mathbb{C}$ holomorphe. Alors*

$$f(z_0) = \frac{1}{\pi r^2}\int_{D(z_0,r)}f(z)dz$$

où $\int_{D(z_0,r)}f(z)dz$ représente l'intégrale double de f sur $D(z_0,r)$, voir aussi la note 54.

66. Voir l'exercice 3.23 au Chapitre I.

Démonstration. On écrit l'intégrale en coordonnées polaires (voir le Chapitre IV) :

$$\int_{D(z,r)} f(z)dz = \int_0^r \int_0^{2\pi} f(z+\rho e^{i\theta})\rho d\theta d\rho = \int_0^r 2\pi f(z)\rho d\rho = \pi r^2 f(z)$$

□

Proposition 3.40 (Principe du maximum)**.** *Soit U un ouvert connexe de $\mathbb{C}$.*

1. *Si $f : U \to \mathbb{C}$ holomorphe sur U, et $|f|$ admet un maximum local dans U, alors f est constante.*
2. *Si de plus*[67] *U est borné et $f : \overline{U} \to \mathbb{C}$ est continue sur $\overline{U}$, alors*

$$\max_{\overline{U}} |f| = \max_{\partial U} |f|. \tag{3.8}$$

Démonstration. Supposons que $z_0 \in U$ est un maximum local de $|f|$. On note $M = |f(z_0)|$. Alors par définition il existe $r_0 > 0$ tel que $\overline{D}(z_0, r_0) \subset U$ et

$$\forall z \in D(z_0, r_0), \quad |f(z)| \leq M,$$

d'où d'après le corollaire (3.39),

$$M = |f(z_0)| = \frac{1}{\pi r^2}\left|\int_{D(z_0,r_0)} f(z)dz\right| \leq \frac{1}{\pi r^2}\int_{D(z_0,r_0)} |f(z)|dz \leq \frac{1}{\pi r^2}\int_{D(z_0,r_0)} Mdz = M.$$

Il y a donc égalité dans ces inégalités, ce qui implique que $|f|$ est constante égale à M sur $D(z_0, r_0)$. Si $M = 0$, alors f est nulle sur $D(z_0, r_0)$ et donc sur U par théorème de prolongement analytique. Si $M \neq 0$, alors on constate alors que $\overline{f} = \frac{M^2}{f}$ est aussi holomorphe sur $D(z_0, r_0)$, ce qui d'après les équations de Cauchy Riemann pour f et $\overline{f}$ donne $f' = 0$ sur $D(z_0, r_0)$, et donc que f est constante sur $D(z_0, r_0)$, et par suite sur U d'après le théorème de prolongement analytique.

Pour le second point : la fonction $|f|$ est continue sur $\overline{U}$ qui est compact, donc $|f|$ atteint son maximum. Si ce maximum est dans ∂U, l'énoncé est prouvé ; si par contre ce maximum est dans U, alors d'après la première partie de l'énoncé, f est contante, auquel cas l'égalité (3.8) est encore valable. □

Remarque 3.41. On utilise souvent cet énoncé sous la forme suivante : si f est holomorphe sur U et non constante, alors pour tout Ω ouvert tel que $\overline{\Omega} \subset U$, on a d'après la proposition 3.40 :

$$\forall z \in \Omega, \quad |f(z)| < \max_{\partial\Omega} |f|.$$

3.6 Holomorphie des intégrales à paramètre :

Dans le paragraphe 5.1.3 du Chapitre I, on a donné des résultats sur les intégrales à paramètre, qui sont élémentaires : on verra des théorèmes plus efficaces au Chapitre IV, qui seront conséquences du théorème de convergence dominée. On va voir ce que deviennent ces résultats dans le cadre holomorphe. Le lecteur qui n'est pas (encore !) familier avec l'intégrale de Lebesgue peut sans trop de danger considérer les fonctions continues par morceaux par rapport au paramètre et considérer que les intégrales sont des intégrales de Riemann ; néanmoins, la force de l'énoncé et sa spécificité dans le contexte de l'holomorphie deviendront plus claires en comparaison du contexte de la dérivabilité dans $\mathbb{R}$, voir le théorème 3.21 du Chapitre IV. Vous pourrez quand même sentir cette différence en comparant le corollaire 3.46 ci-dessous avec la dernière question de l'exercice 5.22 du Chapitre I.

Théorème 3.42. *Soit U un ouvert de $\mathbb{C}$, et $(E, \mathcal{A}, \mu)$ un espace mesuré. Soit $f : E \times U \to \mathbb{C}$ telle que*

— *pour tout $z \in U$, l'application $t \mapsto f(t, z)$ est mesurable,*

— *pour tout $t \in E$, l'application $z \mapsto f(t, z)$ est holomorphe,*

67. Autrement dit on suppose encore U connexe et f holomorphe sur U.

— *il existe* $\varphi : E \to \mathbb{R}_+$ *intégrable*[68] *et*

$$\forall z \in U, \forall t \in E, \ \ |f(t,z)| \leq \varphi(t). \tag{3.9}$$

Alors l'application

$$z \in U \mapsto F(z) = \int_E f(t,z)d\mu(t)$$

est bien définie et holomorphe sur U. *De plus*

$$\forall k \in \mathbb{N}, \ \ \forall z \in U, \ \ F^{(k)}(z) = \int_E \frac{\partial^k f}{\partial z^k}(t,z)d\mu(t). \tag{3.10}$$

Remarque 3.43. Comme d'habitude en théorie de Lebesgue, on peut remplacer $\forall t \in E$ par "pour presque tout $t \in E$".

Remarque 3.44. L'hypothèse (3.9) est typique de la théorie de Lebesgue : on l'appelle hypothèse de domination. Dans ce type de théorème sur les intégrales à paramètre, il est très fréquent qu'on n'arrive pas à prouver directement une telle estimation pour tout $t \in E$ et tout $z \in U$. On trouve d'ailleurs souvent dans la littérature l'hypothèse suivante : pour tout K compact inclus dans U, il existe φ_K telle que

$$\forall z \in K, \forall t \in E, \ \ |f(t,z)| \leq \varphi_K(t), \qquad \text{et} \qquad \int_E \varphi_K(t)d\mu(t) < +\infty,$$

qui est plus faible que (3.9) dans laquelle le même φ vaut pour tout K. J'ai préféré donner un énoncé simplifié pour la raison suivante : en pratique, si l'hypothèse (3.9) est satisfaite, alors il n'y a pas de raison d'introduire ces compacts. Si par contre on n'arrive pas à prouver une telle estimation, alors on introduit des sous-ensembles ouverts U_ε (d'adhérence compacte ou non) de U tels que $\cup_{\varepsilon>0} U_\varepsilon = U$ et telle qu'on ait une estimation du type (3.9) sur U_ε pour une fonction φ_ε intégrable. Le théorème 3.42 s'applique alors sur cet ouvert U_ε, ce qui donne que F est holomorphe sur U_ε. Ceci étant vrai pour tout $\varepsilon > 0$, cela prouve que F est holomorphe sur U.

Remarque 3.45. Si γ est un chemin C^1 par morceaux, et $f : (\zeta, z) \in O \times U \to \mathbb{C}$, on peut envisager d'appliquer le théorème 3.42 à la fonction

$$z \longmapsto \int_\gamma f(\zeta, z)d\zeta = \int_a^b f(\gamma(t), z)\gamma'(t)dt$$

en adaptant simplement les hypothèses. Par exemple, si on regarde l'expression qui apparaît dans la formule de Cauchy : étant donnée $f : U \to \mathbb{C}$ est continue[69], on considère

$$F : z \longmapsto \int_\gamma \frac{f(\zeta)}{\zeta - z}d\zeta \tag{3.11}$$

et on voit que si $U_\varepsilon = \{z \in U, d(z,\gamma) > \varepsilon\}$, alors

$$\forall \zeta \in \mathrm{Im}(\gamma), \quad \forall z \in U_\varepsilon, \quad \left|\frac{f(\zeta)}{\zeta - z}\right| \leq \frac{1}{\varepsilon}\|f\|_{\infty,\mathrm{Im}(\gamma)} \text{ qui est intégrable sur } \gamma$$

et donc on peut appliquer le théorème 3.42 qui montre que F est holomorphe sur U_ε pour tout $\varepsilon > 0$, et donc sur $U \setminus \mathrm{Im}(\gamma)$. On obtient de plus

$$\forall k \in \mathbb{N}, \quad \forall z \in U \setminus \mathrm{Im}(\gamma), \quad F^{(k)}(z) = k! \int_\gamma \frac{f(\zeta)}{(\zeta - z)^{k+1}}d\zeta.$$

Au final, la démonstration du théorème 3.28 n'est pas si différente de la preuve 2 qu'on donne ci-dessous où on avait montré à la main le caractère développable en série entière de (3.11) dans le cas où γ est un cercle. Voir aussi la remarque 3.30 pour l'expression des dérivées.

68. C'est-à-dire $\int_E |\varphi(t)|d\mu(t) < +\infty$.

69. Je n'ai pas cherché à mettre des hypothèses optimales...

Démonstration. On peut donner plusieurs démonstrations, voir par exemple [God01, VIII.7] : on en détaille deux ici, qui seront à comparer avec la preuve du théorème 3.21 au Chapitre IV (le première est similaire, à ceci près qu'on prouve ici la domination de la variation de f, la seconde est radicalement différente et se base sur le développement en série entière).

(*Ce paragraphe demande un peu de familarité avec le théorème de convergence dominée pour la preuve 1, et le théorème d'interversion série-intégrale (cas particulier du théorème de Fubini, si on veut) pour la preuve 2).

Preuve 1 : Soit $z \in U$ tel que $D(z, 2\alpha) \subset U$. On a

$$\forall t \in E, \ \ \forall h \in D(0,\alpha), \quad \frac{F(z+h)-F(z)}{h} = \int_E \frac{f(t,z+h)-f(t,z)}{h} d\mu(t)$$

avec

$$\forall t \in E, \ \ \frac{f(t,z+h)-f(t,z)}{h} \xrightarrow[h\to 0]{} \frac{\partial f}{\partial z}(t,z).$$

On veut maintenant justifier l'interversion de limite en utilisant le théorème de convergence dominée ; il faut donc obtenir une domination : par le théorème des accroissements finis (voir le Chapitre VII) on a

$$\forall t \in E, \ \ \forall h \in D(0,\alpha), \quad \left| \frac{f(t,z+h)-f(t,z)}{h} \right| \leq \sup_{D(z,\alpha)} \left| \frac{\partial f}{\partial z}(t,z) \right|$$

et avec l'exercice 3.64 (la question 1 suffit), il existe C tel que[70]

$$\forall t \in E, \ \ \sup_{D(z,\alpha)} \left| \frac{\partial f}{\partial z}(t,z) \right| \leq C \sup_{D(z,2\alpha)} |f(t,z)| \leq C\varphi(t)$$

où on a utlisé l'hypothèse de domination faite sur f. Et donc le théorème de convergence dominée s'applique, et permet de voir que F est $\mathbb{C}$-dérivable en z et

$$F'(z) = \int_E \frac{\partial f}{\partial z}(t,z) d\mu(t).$$

Si on veut justifier la formule (3.10), il faut réitérer le processus : remarquons qu'on a obtenu ci-dessus

$$\forall t \in E, \ \ \sup_{D(z,\alpha)} \left| \frac{\partial f}{\partial z}(t,z) \right| \leq C\varphi(t),$$

qui montre qu'on a bien à nouveau les hypothèses faites sur f, au moins dans l'ouvert $D(z,\alpha)$. On laisse les détails au lecteur.

Preuve 2 :[71]Soit $z_0 \in U$ tel que $\overline{D}(z_0, 2\alpha) \subset U$[72]. On utilise ici le théorème 3.28, qui permet d'écrire :

$$\forall t \in E, \ \forall z \in D(z_0,\alpha), \quad f(t,z) = \sum_{n\in\mathbb{N}} \frac{\partial^n f}{\partial z^n}(t,z_0) \frac{(z-z_0)^n}{n!}.$$

On veut intervertir somme et intégrale, on verra au Chapitre IV que cela est possible si

$$\sum_{n\in\mathbb{N}} \int_E \left| \frac{\partial^n f}{\partial z^n}(t,z_0) \frac{(z-z_0)^n}{n!} \right| d\mu(t) < +\infty.$$

Avec les inégalités de Cauchy données à la Proposition 3.35 : pour r tel que $D(z_0, r) \subset U$,

$$\left| \frac{\partial^n f}{\partial z^n}(t,z_0) \right| \leq \frac{n!}{r^n} \sup_{C(z_0,r)} |f| \leq \frac{n!}{r^n}\varphi(t)$$

70. C'est ici que réside la spécificité au cadre holomorphe : on voit que la domination sur f implique une domination sur les variations de f ; dans le cas du théorème 3.21 du Chapitre IV, on devra en fait supposer une domination de la dérivée de f par rapport au paramètre.

71. Je n'ai pas vraiment trouvé cette preuve dans la littérature ; elle ressemble à [God01, Troisième démonstration page 65], mais je ne comprends pas très bien pourquoi il fait quelque chose de plus compliqué. Si quelqu'un trouve une faute dans mon raisonnement, qu'il me prévienne ! Si quelqu'un trouve une bonne référence, idem !

72. Voyez plus loin le pourquoi de ce choix !

On est donc tenté de prendre $r = 2\alpha$ et $z \in D(z_0, \alpha)$ qui donne [73] :

$$\sum_{n\in\mathbb{N}} \int_E \left| \frac{\partial^n f}{\partial z^n}(t, z_0) \frac{(z - z_0)^n}{n!} \right| d\mu(t) \leq \sum_{n\in\mathbb{N}} \int_E \frac{n!}{(2\alpha)^n} \varphi(t) \frac{\alpha^n}{n!} d\mu(t) = \left(\sum_{n\in\mathbb{N}} \frac{1}{2^n} \right) \left(\int_E \varphi(t) d\mu(t) \right) < +\infty.$$

Ainsi, on peut écrire

$$\forall z \in D(z_0, \alpha), \quad F(z_0) = \sum_{n\in\mathbb{N}} \left(\int_E \frac{\partial^n f}{\partial z^n}(t, z_0) d\mu(t) \right) \frac{(z - z_0)^n}{n!},$$

donc F est développable en série entière sur U et par identification des coefficients du développement, on récupère bien :

$$\forall k \in \mathbb{N}, \quad \forall z_0 \in U, \quad F^{(k)}(z_0) = \int_E \frac{\partial^k f}{\partial z^k}(t, z_0) d\mu(t).$$

□

Réécrivons l'énoncé précédent dans le cas où l'espace mesuré est $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathrm{Card})$, ce qui donne des séries :

Corollaire 3.46. *Soit U un ouvert de $\mathbb{C}$, et $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions holomorphes sur U. On suppose que la série $\sum_{n\in\mathbb{N}} f_n$ est normalement convergente* [74] *sur U. Alors*

$$z \in U \mapsto F(z) = \sum_{n\in\mathbb{N}} f_n(z)$$

est bien définie et holomorphe sur U. De plus

$$\forall k \in \mathbb{N}, \forall z \in U, \quad F^{(k)}(z) = \sum_{n\in\mathbb{N}} f_n^{(k)}(z).$$

3.7 Suites de fonctions holomorphes

On reprend le paragraphe 4 du Chapitre I à la lumière de la formule de Cauchy. Nous allons voir que les résultats deviennent autrement plus performants et simples :

Théorème 3.47. *Soit U un ouvert de $\mathbb{C}$, et $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions holomorphes sur U qui converge uniformément sur tout compact de U vers $f : U \to \mathbb{C}$.*

Alors f est holomorphe sur U. De plus, pour tout $k \in \mathbb{N}$, $(f_n^{(k)})_{n\in\mathbb{N}}$ converge uniformément vers $f^{(k)}$ sur tout compact de U.

Démonstration. Commençons par la seconde partie de l'énoncé, c'est-à-dire qu'on suppose la limite f holomorphe, et on montre que la convergence uniforme de f_n (sur tout compact) implique la convergence uniforme de ses dérivées (sur tout compact). Soit K un compact de U, et $k \in \mathbb{N}$. D'après l'exercice 1.6, $r_0 = d(K, \mathbb{C} \setminus U) > 0$. Soit $r \in]0, r_0[$. On pose

$$K_r = \{z \in \mathbb{C}, d(z, K) \leq r\} = \bigcup_{z\in K} \overline{D}(z, r).$$

Il est classique que K_r est aussi compact [75], et il est inclus dans U car $r < r_0$. Par les inégalités de Cauchy,

$$\forall z \in K, \quad |(f_n - f)^{(k)}(z)| \leq \frac{k!}{r^k} \sup_{C(z_0,r)} |f_n - f|, \qquad \text{d'où} \qquad \sup_K |f_n^{(k)} - f^{(k)}| \leq \frac{k!}{r^k} \sup_{K_r} |f_n - f|.$$

Le terme de droite tend vers 0, et donc $(f_n^{(k)})_{n\in\mathbb{N}}$ converge uniformément vers $f^{(k)}$.

73. Ici on a mis des modules à l'intérieur, toutes les interversions sont donc autorisées, voir le Chapitre IV.
74. On verra en fait au paragraphe 3.7 que la convergence uniforme (sur tout compact de U) suffit.
75. Voir le paragraphe 1.5 au Chapitre III.

On donne maintenant deux preuves du caractère holomorphe de f :

Preuve 1 : On se donne $z_0 \in U$ et $r_0 > 0$ tel que $\overline{D}(z_0, r_0) \subset U$. Alors par le corollaire 3.26 on peut écrire

$$\forall n \in \mathbb{N}, \forall z \in D(z_0, r_0), \quad f_n(z) = \frac{1}{2i\pi} \int_{C(z_0,r_0)} \frac{f_n(\zeta)}{\zeta - z} d\zeta.$$

Grâce à la convergence uniforme de f_n sur le compact $\overline{D}(z_0, r_0)$, on peut écrire la même formule pour f, ce qui implique par la preuve du Théorème 3.28 qu'elle est développable en série entière et donc holomorphe au voisinage de z_0, et ce pour tout $z_0 \in U$.

Preuve 2 : on peut aussi utiliser le théorème de Morera qui sera donné au paragraphe 4.7.

Comme pour tout $n \in \mathbb{N}$, f_n est continue, par le théorème 4.10 du Chapitre I, f est continue. Soit Δ un triangle plein inclus dans U. Alors pour tout $n \in \mathbb{N}$, f_n est holomorphe donc $\int_{\partial\Delta} f_n(z)dz = 0$. Par convergence uniforme des $(f_n)_{n\in\mathbb{N}}$, on en déduit (voir la remarque 3.17) que $\int_{\partial\Delta} f(z)dz = 0$. Par le théorème 4.32 de Morera, on en déduit que f est holomorphe.

□

Exemple 3.48. Ainsi, si pour tout $n \in \mathbb{N}$, $f_n : U \to \mathbb{C}$ est holomorphe, et si $\sum_n f_n$ converge uniformément sur tout compact de U, alors $S := \sum_{n\in\mathbb{N}} f_n$ est holomorphe sur U, et de plus

$$\forall k \in \mathbb{N}, \quad S^{(k)} = \sum_{n\in\mathbb{N}} f_n^{(}k)$$

où la convergence de la somme est uniforme sur tout compact.

3.8 Singularités

On a déjà vu à la remarque 3.25 que le théorème de Cauchy ne s'applique pas à une fonction qui est définie sur un ouvert privé d'un point isolé (par exemple $\mathbb{C}^*$). Nous allons nous intéresser aux singularités que peut avoir une fonction holomorphe en un tel point, et on donnera un résultat qui remplace la formule de Cauchy dans une telle situation.

3.8.1 Classification des singularités

Théorème 3.49. *Soit U un ouvert de $\mathbb{C}$, $z_0 \in U$, et f une fonction holomorphe sur $U \setminus \{z_0\}$. Alors seuls 3 cas se présentent :*

1. *La fonction f se prolonge en une fonction holomorphe sur U ; on parle de singularité effaçable*[76].
2. *On a $\lim_{z\to z_0} |f(z)| = +\infty$: dans ce cas, il existe $m \in \mathbb{N}^*$ et g holomorphe sur U tels que*

$$\forall z \in U \setminus \{z_0\}, \quad f(z) = \frac{g(z)}{(z - z_0)^m}, \qquad \text{et } g(z_0) \neq 0.$$

   On dit que z_0 est un pôle d'ordre m pour f.

3. *L'image par f de tout disque épointé $D(z_0, r)\setminus\{z_0\} \subset U$ est dense dans $\mathbb{C}$; on parle de singularité essentielle.*

Démonstration. Supposons qu'on n'est pas dans le 3ième cas. Alors il existe $r > 0$ tel que $f(D(z_0, r) \setminus \{z_0\})$ n'est pas dense dans $\mathbb{C}$, donc il existe $D(\zeta, \alpha) \subset \mathbb{C}$ qui est disjoint de $f(D(z_0, r) \setminus \{z_0\})$. Ainsi

$$\forall z \in D(z_0, r) \setminus \{z_0\}, \quad |f(z) - \zeta| \geq \alpha, \quad \text{donc on peut poser } \varphi(z) = \frac{1}{f(z) - \zeta}.$$

Comme φ est holomorphe et bornée sur $D(z_0, r) \setminus \{z_0\}$, par la proposition 3.32, on peut la prolonger en z_0 en une fonction $\widetilde{\varphi}$ holomorphe sur $D(z_0, r)$. On distingue deux cas :

76. On peut aussi dire apparente.

1. Si $\widetilde{\varphi}(z_0) \neq 0$, alors on peut écrire $f = \zeta + \dfrac{1}{\widetilde{\varphi}}$ sur $D(z_0, r') \setminus \{z_0\}$ où $r' \leq r$ est choisi tel que $\widetilde{\varphi}$ ne s'annule pas sur $D(z_0, r')$, ce qui montre que f se prolonge de façon holomorphe en z_0 et on est dans la situation 1.
2. Si $\widetilde{\varphi}(z_0) = 0$, alors il existe $m \in \mathbb{N}^*$ tel que $\widetilde{\varphi}(z) = (z - z_0)^m h(z)$ avec h holomorphe et $h(z_0) \neq 0$ (voir l'exercice 2.58). Par continuité de h, il existe $r' \leq r$ tel que h ne s'annule pas sur $D(z_0, r')$, et alors
$$\forall z \in D(z_0, r') \setminus \{z_0\}, \quad (z - z_0)^m f(z) = \frac{1}{h(z)} + \zeta(z - z_0)^m.$$
Ainsi la fonction $g : z \in D(z_0, r') \setminus \{z_0\} \mapsto \frac{1}{h(z)} + \zeta(z - z_0)^m$ se prolonge en une fonction holomorphe sur $D(z_0, r')$, telle que $g(z_0) = \frac{1}{h(z_0)} \neq 0$. Enfin
$$f(z) \underset{z \to z_0}{\sim} \frac{g(z_0)}{(z - z_0)^m} \text{ donc } |f(z)| \xrightarrow[z \to z_0]{} +\infty$$
et donc on est bien dans la situation 2.

□

Remarque 3.50. On retrouve ici une formulation améliorée de la proposition 3.32 : en effet si f est bornée, on ne peut pas être dans les situations 2 ou 3. Mais attention, on utilise la proposition 3.32 dans la preuve de cet énoncé !

Remarque 3.51. Le 3ième cas est appelé théorème de Casorati-Weierstrass. Le comportement décrit est très différent des deux autres : dans le premier, $f(z)$ se rapproche d'une limite complexe quand $z \to z_0$, et dans le second, $|f(z)|$ tend vers $+\infty$. Dans ce dernier cas, pour tout $z' \in \mathbb{C}$, il existe $(z_n)_{n \in \mathbb{N}}$ qui converge vers z_0 telle que $f(z_n)$ tend vers z'.

Remarque 3.52. Dans le deuxième cas, c'est-à-dire si z_0 est un pôle d'ordre m, on peut montrer qu'il existe $a_{-m}, \ldots, a_{-1} \in \mathbb{C}^m$ et h holomorphe sur U tels que[77]
$$\forall z \in U, \quad f(z) = \frac{a_{-m}}{(z - z_0)^m} + \frac{a_{-m+1}}{(z - z_0)^{m-1}} + \ldots + \frac{a_{-1}}{z - z_0} + h(z), \tag{3.12}$$
En effet, avec les notations du théorème 3.49, la fonction g est holomorphe sur U, donc on peut la développer autour de z_0, et en isolant les premiers termes du développement en série, on obtient qu'il existe $a_{-m}, \ldots, a_{-1} \in \mathbb{C}^m$ tels que
$$g(z) = a_{-m} + a_{-m+1}(z - z_0) + \ldots a_{-1}(z - z_0)^{m-1} + (z - z_0)^m h(z)$$
où a priori h est défini sur un voisinage de z_0. On obtient donc (3.12). Pour étendre h en dehors du voisinage de z_0, il suffit de la définir par $f(z) - \frac{a_{-m}}{(z-z_0)^m} - \frac{a_{-m+1}}{(z-z_0)^{m-1}} - \ldots - \frac{a_{-1}}{z-z_0}$ qui est holomorphe sur $U \setminus \{z_0\}$ (n'oublions pas que par le théorème de prolongement analytique, il y a unicité du prolongement de h).

Exemple 3.53. Donnons 3 exemples :
1. La fonction $f : z \in \mathbb{C}^* \mapsto \frac{e^z - 1}{z}$ a une singularité effaçable en 0.
2. La fonction $f : z \in \mathbb{C}^* \mapsto \frac{1}{z^m}$ pour $m \in \mathbb{N}^*$ a un pôle d'ordre m en 0.
3. La fonction $f : z \in \mathbb{C}^* \mapsto e^{\frac{1}{z}}$ a une singularité essentielle en 0. On propose deux méthodes pour le voir :
   - soit on accepte d'utiliser le théorème 3.49 et on observe que 0 n'est ni effaçable (car $\left|f\left(\dfrac{1}{n}\right)\right| \xrightarrow[n \to +\infty]{} +\infty$) ni un pôle (car $\left|f\left(\dfrac{-1}{n}\right)\right| \xrightarrow[n \to +\infty]{} 0$). C'est donc une singularité essentielle.
   - soit on veut se convaincre à la main que l'image d'un disque épointé est dense dans $\mathbb{C}$. On vérifie en fait facilement que
$$f(D(0, r) \setminus \{0\}) = \mathbb{C}^*.$$

77. On appelle parfois $\sum_{k=1}^{m} \frac{a_{-k}}{(z-z_0)^k}$ la partie principale de f en z_0.

3.8.2 Développement en série de Laurent

Définition 3.54. Une série de Laurent est une série de fonctions s'écrivant

$$z \mapsto \sum_{n\in\mathbb{Z}} a_n z^n := \sum_{n\in\mathbb{N}} a_n z^n + \sum_{n\in\mathbb{N}^*} a_{-n} z^{-n},$$

où $(a_n)_{n\in\mathbb{Z}} \in \mathbb{C}^{\mathbb{Z}}$. On dit que la série est convergente si les deux séries $\sum_{n\in\mathbb{N}} a_n z^n$ et $\sum_{n\in\mathbb{N}^*} a_{-n} z^{-n}$ convergent.

Si on note R le rayon de convergence de la série entière $z \mapsto \sum_{n\in\mathbb{N}} a_n z^n$ et $\frac{1}{r}$ le rayon de convergence de la série entière $w \mapsto \sum_{n\in\mathbb{N}^*} a_{-n} w^n$, alors la série de Laurent $z \mapsto \sum_{n\in\mathbb{Z}} a_n z^n$ converge dans l'anneau $A(0, r, R)$ où

$$A(z_0, r, R) := D(z_0, R) \setminus \overline{D}(z, r),$$

qui est non vide seulement si $R > r$.

En appliquant les résultats sur les séries entières à chacune de ces séries, on déduit que la convergence est normale sur tout anneau $\overline{A}(0, r_1, r_2)$ tel que $r < r_1 < r_2 < R$, et que la somme d'une série de Laurent est holomorphe sur $A(0, r, R)$.

On a vu à la remarque 3.52 que si f admet un pôle en un point, alors elle admet un développement de Laurent avec un nombre fini d'exposants négatifs. On peut généraliser ce résultat :

Théorème 3.55. *Soit f holomorphe sur un anneau $A(z_0, r, R)$. Alors f admet un unique développement en série de Laurent : il existe une unique suite $(a_n)_{n\in\mathbb{Z}}$ telle que*

$$\forall z \in A(z_0, r, R), \quad f(z) = \sum_{n\in\mathbb{Z}} a_n (z - z_0)^n.$$

De plus, les coefficients sont donnés par :

$$\forall n \in \mathbb{Z}, a_n = \frac{1}{2i\pi} \int_{C(z_0, r')} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta \tag{3.13}$$

pour n'importe quel $r' \in]r, R[$.

Démonstration. On renvoit donc le lecteur au paragraphe 4.8.4 pour une démonstration.[78] □

Remarque 3.56. Le développement en série de Laurent dans le cas d'un disque épointé, c'est-à-dire $D(z_0, R) \setminus \{z_0\} = A(z_0, 0, R)$ permet de relire la classification des singularités donnée au théorème 3.49 : en effet, si $f(z) = \sum_{n\in\mathbb{Z}} a_n (z - z_0)^n$ pour $z \in D(z_0, R) \setminus \{z_0\}$, alors

1. la fonction f a une singularité effaçable en z_0 si et seulement si f admet un développement en série entière au voisinage de 0, c'est-à-dire si et seulement si

$$\forall n \in \mathbb{Z} \setminus \mathbb{N}, \quad a_n = 0.$$

2. la fonction f a un pôle d'ordre m en z_0 si et seulement si f a un développement de la forme (3.12), c'est-à-dire si et seulement si

$$\forall n \in \mathbb{Z} \text{ tel que } n < -m, \quad a_n = 0$$

3. la fonction f a une singularité essentielle si et seulement si les deux cas précédents ne se présentent pas, c'est-à-dire si et seulement si

$$\text{il y a une infinité de } n \text{ négatifs tels que } a_n \neq 0.$$

78. A venir, deux autres preuves

3.8.3 Théorème des résidus

Définition 3.57 (Résidu)**.** Soit U un ouvert de $\mathbb{C}$, $z_0 \in U$, et $f : U \setminus \{z_0\}$ holomorphe. On écrit $\sum_{n\in\mathbb{Z}} a_n(z-z_0)^n$ le développement en série de Laurent de f donné par le Théorème 3.55. On définit

$$\operatorname{Res}(f, z_0) = a_{-1}.$$

le résidu de f en z_0.

Théorème 3.58 (Théorème des résidus, première version)**.** *Soit U un ouvert étoilé, $(z_1, z_2, \dots, z_p\}$ un nombre fini de points de U deux à deux distincts. Soit $f : U \setminus \{z_1, z_2, \dots, z_p)$ holomorphe. Soit γ un lacet C^1 par morceaux et à valeurs dans $U \setminus \{z_1, z_2, \dots, z_p\}$. Alors*

$$\frac{1}{2i\pi}\int_\gamma f(z)dz = \sum_{k=1}^{p} \operatorname{Res}(f, z_k)\operatorname{ind}_\gamma(z_k) \tag{3.14}$$

Démonstration. Pour chaque $k \in [\![1, p]\!]$, on note f_k la partie du développement en série de Laurent de f avec des exposants négatifs, qui converge sur tout compact de $\mathbb{C} \setminus \{z_k\}$. Ainsi $f - \sum_{k=1}^{p} f_k$ se prolonge en une fonction holomorphe sur U. On applique le théorème 4.46 à cette fonction, ainsi que la remarque que pour $k \in [\![1, p]\!]$, si on écrit $f_k(z) = \sum_{n<0} a_n(z-z_0)^n$ alors

$$\int_\gamma f_k(z)dz = \sum_{n<0}\int_\gamma a_n(z-z_0)^n dz = a_{-1}\int_\gamma \frac{1}{z-z_k}dz = 2i\pi\operatorname{Res}(f, z_k)\operatorname{ind}_\gamma(z_k)$$

où on a pu échanger somme et intégrale car la convergence est normale sur tout compact de $\mathbb{C}\setminus\{z_k\}$. □

3.9 Exercices

Exercice 3.59. Montrer que la fonction $z \in \mathbb{C} \mapsto \overline{z}$ n'est $\mathbb{C}$-dérivable en aucun point de $\mathbb{C}$.

Exercice 3.60. Soit U un ouvert connexe, et $f : U \to \mathbb{C}$ telle que f et $\overline{f}$ sont holomorphes. Montrer que f est constante sur U.

Exercice 3.61 (Nouvelle construction du Logarithme principal)**.** On peut donner une autre définition expéditive de la fonction Logarithme principal introduite au paragraphe 2.5, et qui ne nécessite pas d'avoir construit l'exponentielle : posons

$$\forall z \in \mathbb{C} \setminus \mathbb{R}_-, \quad \mathrm{L}(z) = \int_{[1,z]} \frac{1}{\zeta}d\zeta = \int_0^1 \frac{1}{1-t+tz}(z-1)dt.$$

Sauf pour la dernière question, on n'utilisera pas la fonction Log du paragraphe 2.5 pour se convaincre qu'on peut bien utiliser cette définition pour construire le Logarithme principal.

1. Montrer que la définition précédente a bien un sens.
2. Soit $z \in \mathbb{C} \setminus \mathbb{R}_-$, et h tel que $|h| < |z|$. Montrer

$$\mathrm{L}(z+h) - \mathrm{L}(z) = \int_{[z,z+h]} \frac{d\zeta}{\zeta}$$

   et en déduire que L est $\mathbb{C}$-dérivable en z, avec $\mathrm{L}'(z) = \frac{1}{z}$.
3. On va montrer de deux façons que

$$\forall z \in \mathbb{C} \setminus \mathbb{R}_-, \quad \exp(\mathrm{L}(z)) = z. \tag{3.15}$$

(a) Calculer la dérivée de $z \in \mathbb{C} \setminus \mathbb{R}_- \mapsto z \exp(-\mathrm{L}(z))$ et en déduire (3.15) [79]
(b) Montrer que $\mathrm{L}(x) = \ln(x)$ si $x \in]0, +\infty[$, et en déduire (3.15).

4. Déduire de la question précédente que exp est surjective [80].
5. Montrer que si $|z| < 1$, alors
$$\mathrm{L}(1+z) = \sum_{k \in \mathbb{N}^*} (-1)^{k-1} \frac{z^k}{k}.$$
6. En supposant construite la fonction Log du paragraphe 2.5, montrer soit avec (3.15), soit avec la question précédente, que $\mathrm{L} = \mathrm{Log}$.

Exercice 3.62 (Exemple fondamental)**.** Le premier calcul sera très important à retenir, il donne une lumière non négligeable sur tous les éléments qui suivent dans ce chapitre. Le second calcul est plus technique, et on l'éclaircira par la théorie :

1. Pour $n \in \mathbb{Z}$ et $r > 0$, montrer que
$$\int_{C(0,r)} z^n dz = \begin{cases} 0 & \text{si } n \neq -1 \\ 2i\pi & \text{si } n = -1 \end{cases}.$$
2. *Pour $n \in \mathbb{Z}$, $r > 0$ et z_0 tel que $\left[0 \notin C(z_0, r) \text{ si } n < 0\right]$, calculer
$$\int_{C(z_0,r)} z^n dz.$$

Exercice 3.63. On suppose que γ décrit un cercle de centre z_0 et de rayon $r > 0$, comme dans l'Exemple 3.10. Montrer que $\mathrm{ind}_\gamma(z) = 1$ si $z \in D(z_0, r)$ et $\mathrm{ind}_\gamma(z) = 0$ si $z \in \mathbb{C} \setminus \overline{D}(z_0, r)$. Comparez avec l'exercice 3.62.

Exercice 3.64. Dans la formule de Cauchy, on contrôle la valeur de $f^{(n)}(z_0)$ par la valeur de f sur un cercle qui entoure z_0, et on ne peut pas faire tendre le rayon de ce cercle vers 0 à cause de la constante $\frac{1}{r^n}$. On voit ici deux façons d'exprimer cette idée :

1. Soit $z_0 \in \mathbb{C}$, $r > 0$ et $\varepsilon > 0$ tels que $D(z_0, r+\varepsilon) \subset U$. Montrer qu'il existe C qui dépend de ε et n telle que pour tout $f : U \to \mathbb{C}$ holomorphe,
$$\|f^{(n)}\|_{\infty, D(z_0,r)} \leq C \|f\|_{\infty, D(z_0, r+\varepsilon)}.$$
2. (*)Plus généralement, on se donne K et L deux compacts de U tels que $K \subset \mathring{L} \subset L \subset U$, et $n \in \mathbb{N}$. Montrer qu'il existe une constante C [81] telle que pour tout $f : U \to \mathbb{C}$ holomorphe,
$$\|f^{(n)}\|_{\infty, K} \leq C \|f\|_{\infty, L}.$$

Exercice 3.65 (Autre preuve du principe du maximum)**.** Soit U un ouvert connexe de $\mathbb{C}$, $f : U \to \mathbb{C}$ holomorphe sur U telle que f n'est pas constante. Soit $z_0 \in U$

1. Montrer que f n'est pas constante sur un voisinage de z_0.
2. Montrer qu'il existe $k \in \mathbb{N}^*$ et $\alpha \neq 0$ tel que
$$f(z_0 + h) = f(z_0) + \alpha h^k + o(h^k).$$
3. Montrer que z_0 n'est pas un maximum local de $|f|$. [82]On a ainsi montré la contraposée du premier point de la proposition 3.40.

---

79. Comparez avec le début de la preuve 2 du théorème 2.22.
80. C'est en fait ce qu'on a fait dans la preuve 2 du théorème 2.22.
81. Cette constante dépend de K, L et n, mais pas de f.
82. On pourra faire deux cas suivant que $f(z_0) = 0$ ou $f(z_0) \neq 0$, et dans ce second cas écrire
$$\frac{f(z_0+h)}{f(z_0)} = 1 + \beta h^k + o(h^k)$$
et choisir judicieusement h.

Exercice 3.66 ("Principe du minimum" ou théorème de module minimal)**.** Soit U un ouvert connexe de $\mathbb{C}$, $f : U \to \mathbb{C}$ holomorphe sur U telle que f n'est pas constante.

1. Montrer que si $z_0 \in U$ est un minimum local de $|f|$, alors $f(z_0) = 0$.[83]
2. En déduire une preuve alternative du théorème de d'Alembert-Gauss.

Exercice 3.67 (Encore une autre preuve du principe du maximum, à partir de la remarque 3.38)**.** Soit U un ouvert connexe de $\mathbb{C}$, $f : U \to \mathbb{C}$ holomorphe sur U, et $z_0 \in U$ un maximum local de $|f|$.

1. Montrer qu'il existe $r > 0$ tel que $D(z_0, r) \subset U$ et
$$|f(z_0)|^2 \geq \frac{1}{2\pi} \int_0^{2\pi} |f(z_0 + re^{it})|^2 dt.$$
2. En appliquant (3.7), en déduire que $f^{(n)}(z_0) = 0$ pour tout $n \in \mathbb{N}^*$.
3. Conclure.

Exercice 3.68 (Fonctions "Gamma" et "Zeta")**.**

1. On pose
$$\forall z \in \mathbb{C} \text{ tel que } \mathrm{Re}(z) > 0, \quad \Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt.$$
Montrer que Γ est holomorphe sur $\{z \in \mathbb{C}, \mathrm{Re}(z) > 0\}$.
2. On pose
$$\forall z \in \mathbb{C} \text{ tel que } \mathrm{Re}(z) > 1, \quad \zeta(z) = \sum_{n \in \mathbb{N}^*} \frac{1}{n^z}.$$
Montrer que ζ est holomorphe sur $\{z \in \mathbb{C}, \mathrm{Re}(z) > 1\}$.

Exercice 3.69. Donnez le développement en série de Laurent en 0 de la fonction $z \in \mathbb{C}^* \mapsto \exp\left(\frac{1}{z}\right)$.

Exercice 3.70. Soit $f = \frac{g}{h}$ où (g, h) sont deux fonctions holomorphes sur $D(z_0, r)$, telle que $h(z_0) = 0$ et $h'(z_0) \neq 0$.

1. Montrer qu'il existe $r' \in]0, r]$ tel que f est holomorphe sur $D(z_0, r') \setminus \{z_0\}$.
2. Montrer que
$$\mathrm{Res}(f, z_0) = \frac{g(z_0)}{h'(z_0)}.$$

---

4 Sujets de travail et de développement

4.1 *Analyse d'une série entière au bord du disque de convergence

Il s'agit d'un sujet vaste, qui peut fournir de nombreuses idées de développement :

— Le théorème d'Abel non tangentiel (dont un cas particulier est appelé théorème d'Abel radial) est un grand classique : il affirme que si une série entière converge en un point z_0 qui est au bord du disque de convergence, alors il y a continuité de la fonction f définie par cette série entière sur tout secteur angulaire centré en z_0 et qui forme un cône d'angle strictement inférieur à π et dont les bords évitent la tangente au cercle au z_0. Voir par exemple [Gou08b, page 249]. Un autre grand classique est d'étudier des réciproques partielles de cet énoncé : il en existe deux versions, le théorème taubérien faible (voir [Gou08b, page 253]) et taubérien fort ([Gou08b, page 259]).

— Dans [Gou08b, Problème 10] on trouve l'étude d'une série entière qui est semi-convergente sur tout le bord du disque de convergence.

— Le paragraphe IV du Chapitre III de [QZ13] fournit une étude de ces questions. En particulier le théorème des lacunes de Hadamard fournit une condition suffisante pour avoir une série entière qui n'admet un prolongement au voisinage d'aucun point du bord du disque de convergence.

83. On pourra faire deux preuves, l'une en utilisant le principe du maximum, l'autre en utilisant la même méthode qu'à l'exercice 3.65.

4.2 Utilisation des développements en série entière

Il est intéressant d'avoir quelques applications en tête de la représentation en série entière d'une fonction :

— on peut se servir des séries entières pour chercher des solutions (explicites ou non explicites) à des équations différentielles :

  — voir par exemple [QZ13, Théorème IV.7 page 408], [Pom94, 25.3.2] pour des cas linéaires simples.

  — **on peut aller plus loin et regarder la méthode des séries majorantes dans [Car61, VII-1], ce qui permet de montrer un théorème de Cauchy-Lipschitz “analytique” sans beaucoup de technicité.

— les séries entières sont utiles pour la résolution de récurrences et donc en particulier au dénombrement [84] : étant donnée une suite $(u_n)_{n\in\mathbb{N}}$, il peut arriver qu'on puisse la comprendre, voire la calculer en étudiant la fonction génératrice définie par $z \mapsto \sum_{n\in\mathbb{N}} u_n z^n$. À nouveau, les exemples seront plus parlants que la théorie : on renvoit par exemple aux exemples cités dans [Pom94, 25.3.3]. Le calcul du nombre de partitions d'un entier est un classique ([Gou08b, page 249] par exemple), citons aussi les nombres de Bell ([FGN07a]).

On propose deux exercices :

Exercice 4.1 (Calcul des termes de la suite de Fibonacci)**.** On pose $u_0 = 0, u_1 = 1, \quad \forall n \in \mathbb{N}, u_{n+2} = u_{n+1} + u_n$.

  — On pose $F : z \mapsto \sum_{n\in\mathbb{N}} u_n z^n$. Montrer que le rayon de convergence de F est strictement positif, et calculer F sur un voisinage de 0.

  — En déduire que

$$\forall n \in \mathbb{N}, \quad u_n = \frac{1}{\sqrt{5}}\left(\varphi^n - (-1)^n \frac{1}{\varphi^n}\right), \qquad \text{où } \varphi = \frac{1+\sqrt{5}}{2}.$$

Exercice 4.2 (*Nombres de Catalan)**.** On pose $C_0 = 1$ et

$$\forall n \in \mathbb{N}, \quad C_{n+1} = \sum_{k=0}^{n} C_k C_{n-k}.$$

En étudiant la série $z \mapsto \sum_{n\in\mathbb{N}} C_n z^n$, montrer que

$$\forall n \in \mathbb{N}, \quad C_n = \frac{1}{n+1}\binom{2n}{n}.$$

— Via la notion de fonction génératrice des moments d'une variable aléatoire, à savoir $z \mapsto \mathbb{E}(z^X)$ qui se réécrit

$$z \mapsto \sum_{n\in\mathbb{N}} \mathbb{P}(X = n) z^n$$

dans le cas où X est à valeur entière, il est souvent naturel d'avoir recours aux séries entières en théorie des probabilités. Dans le cadre de variables aléatoires non discrètes, on remplacera cet outil par la fonction caractéristique, qui n'est autre que la restriction de la fonction précédente à $\mathcal{U}$.

84. En réalité, ce sont plutôt les séries formelles qui servent, la plupart du temps. Une série formelle est à peu de choses près la même chose qu'une série entière, à ceci près qu'on ne se soucie pas de montrer que la série a un rayon de convergence non nul. On peut tout-à-fait formaliser ça algébriquement, un peu comme on le fait pour les polynômes. D'ailleurs, l'espace des séries formelles s'écrit $\mathbb{C}[\![X]\!]$. C'était au programme de l'agrégation (et même le sujet d'une leçon d'algèbre) jusqu'en 2016, semble-t-il.

4.3 Calculs d'intégrales

Dans [Tau06, Chapitre 15], [AM04, 8.4], [God01, VIII.3], [Mar09, IV.25.II.3 page 474] on trouve des remarques assez générales pour calculer des intégrales. À mon avis, il vaut mieux prendre le temps de traiter quelques exemples variés afin d'intégrer les méthodes et le type de calculs qui interviennent, plutôt que d'avoir une approche générale.

Le plus délicat au fond, dans ces méthodes de calcul, consiste à prendre la décision d'utiliser l'analyse complexe pour calculer une intégrale (ce n'est pas toujours la meilleure méthode), et une fois cette décision prise, de trouver le bon contour à utiliser. Dans les épreuves écrites, ces deux choix sont souvent faits pour vous.

Exercice 4.3 (Intégrales trigonométriques)**.** Soit $a > 1$. En exprimant $t \mapsto \sin(t)$ en fonction de $t \mapsto e^{it}$ et en écrivant l'intégrale suivante comme une intégrale complexe sur le cercle unité, montrer

$$\int_0^{2\pi} \frac{1}{a+\sin(t)} dt = \frac{2\pi}{\sqrt{a^2-1}}.$$

Exercice 4.4 (Fonction caractéristique/transformation de Fourier de la loi de Cauchy)**.**

1. En intégrant la fonction $z \mapsto \frac{1}{1+z^2}$ sur le contour C_R qui est la concaténation du segment $[-R, R]$ et du demi-cercle $C(0, R) \cap \{z \in \mathbb{C}, \operatorname{Im}(z) \geq 0\}$, retrouvez la formule classique

$$\int_{\mathbb{R}} \frac{1}{1+x^2} dx = \pi.$$

2. En réitérant avec la fonction $z \mapsto \frac{e^{i\xi z}}{1+z^2}$ où $\xi \in \mathbb{R}_+$, montrer que

$$\int_{\mathbb{R}} \frac{e^{i\xi x}}{1+x^2} dx = \pi e^{-\xi}.$$

3. Conclure

$$\forall \xi \in \mathbb{R}, \quad \int_{\mathbb{R}} \frac{e^{i\xi x}}{1+x^2} dx = \pi e^{-|\xi|}.$$

Exercice 4.5 (Intégrales de fractions rationnelles)**.**

1. En intégrant la fonction $z \mapsto \frac{1}{1+z^6}$ sur le contour C_R de l'exercice 4.4, montrer

$$\int_0^{+\infty} \frac{1}{1+t^6} dt = \frac{\pi}{3}.$$

2. Plus généralement montrer que si $n \in \mathbb{N}^*$

$$\int_0^{+\infty} \frac{1}{1+t^{2n}} dt = \frac{\pi}{2n \sin\left(\frac{\pi}{2n}\right)}.$$

Exercice 4.6. Soit $a \in]0, 1[$. En intégrant la fonction $z \mapsto \frac{e^{az}}{1+e^z}$ sur le rectangle dont les sommets ordonnés sont $R, R + 2i\pi, -R + 2i\pi, -R$, montrer que

$$\int_{\mathbb{R}} \frac{e^{ax}}{1+e^x} dx = \frac{\pi}{\sin(\pi a)}.$$

Formule des compléments : Voilà deux formules intégrales qui demandent un peu d'efforts :

Théorème 4.7. — *Pour $\alpha \in]0, 1[$, on a*

$$\int_0^{+\infty} \frac{dt}{t^\alpha(1+t)} = \frac{\pi}{\sin(\pi\alpha)}$$

— *Pour tout $z \in \mathbb{C}$ tel que $0 < \operatorname{Re}(z) < 1$,*

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}.$$

Voir [AM04, page 249] : la première formule est obtenue grâce au théorème 3.58, la seconde via un changement de variable dans $\mathbb{R}^2$.

4.4 *Déterminations continues du logarithme

On a donné au paragraphe 2.5 et à l'exercice 3.61 deux construction du Logarithme principal, qui était défini sur $\mathbb{C} \setminus \mathbb{R}_-$. Mais le choix de cet ensemble était un peu arbitraire ; par exemple, retirer une autre demi-droite pourrait paraître tout à fait acceptable. Par contre, on a remarqué que l'on ne pouvait pas raisonnablement définir un logarithme sur $\mathbb{C}^*$. Nous allons préciser un peu tout cela dans ce paragraphe (même si lorsque c'est possible, il est bien suffisant de travailler avec le Logarithme principal).

Définition 4.8. Etant donné un ouvert connexe U de $\mathbb{C}^*$, on dit que $L : U \to \mathbb{C}$ est une détermination continue du logarithme sur U si L est continue sur U et $\forall z \in U,\ \exp(L(z)) = z$.

Exemple 4.9. Le Logarithme principal est une détermination continue du logarithme sur $\mathbb{C} \setminus \mathbb{R}_-$.

Exercice 4.10. Etant donné $\theta \in \mathcal{U}$, on peut chercher une détermination continue du logarithme sur $\mathbb{C} \setminus (e^{i\theta}\mathbb{R}_-)$. Montrer que la fonction

$$\forall z \in \mathbb{C} \setminus (e^{i\theta}\mathbb{R}_-), \quad L_\theta(z) = \mathrm{Log}(e^{-i\theta}z) + i\theta$$

convient.

Exercice 4.11. Etant donné un ouvert connexe U de $\mathbb{C}^*$, montrez que deux déterminations du logarithme sur U sont égales à une constante près, qui est dans $2i\pi\mathbb{Z}$.

Exercice 4.12. Soit U un ouvert connexe de $\mathbb{C}^*$, et L une détermination continue du logarithme sur U.

1. Montrer que L est holomorphe sur U, et $\forall z \in U,\ L'(z) = \frac{1}{z}$.
2. Montrer que si F est holomorphe sur U et $\forall z \in U,\ F'(z) = \frac{1}{z}$, alors il existe une constante $c \in \mathbb{C}$ tel que $F + c$ est une détermination continue du logarithme.

On verra au paragraphe 4.8 et l'exercice 4.42 :

Théorème 4.13. *Si U est un ensemble simplement connexe de $\mathbb{C}$, alors il existe une détermination continue du logarithme sur U.*

Il suffit en effet de prendre le cas particulier $\phi(z) = z$ dans l'exercice 4.42.

Remarque 4.14 (Détermination continue de la racine k-ième)**.** Rappelons enfin que comme à la Remarque 2.30, étant donné L une détermination continue du logarithme sur U, on peut en déduire une fonction puissance sur U en posant, pour $\alpha \in \mathbb{C}$,

$$\forall z \in U,\ \ z^\alpha = e^{\alpha L(z)}. \tag{4.1}$$

Comme on l'a fait pour le logarithme, étant donné un ouvert connexe $U \subset \mathbb{C}^*$, on peut appeller détermination continue de $z^{\frac{1}{k}}$ une fonction $g : U \to \mathbb{C}$ continue et telle que

$$\forall z \in U,\ \ (g(z))^k = z.$$

La formule (4.1) avec $\alpha = \frac{1}{k}$ définit une telle détermination.

Pourquoi préférer un logarithme plutôt qu'un autre ? Cela dépend de la situation dans laquelle on se trouve. Par exemple, pour les calculs d'intégrales, il peut arriver que l'ensemble $\mathbb{C} \setminus \mathbb{R}_-$ ne soit pas propice, comme le montre l'exemple suivant :

Exercice 4.15. Voir [Tau06, page 192]. Soit $\alpha \in]-1,1[$. Pour $0 < r < \frac{1}{2}$ et $R > 2$, on pose $\Gamma_{r,R}$ la concaténation des chemins :

$$C_+(0,r), [r, 1-r], C_+(1,r), [r+1, R], C_+(0,R), [-R, -1-r], C_+(-1,r), [-1+r, -r]$$

où pour $x \in \mathbb{R}$ on a posé $C_+(x,r) = C(x,r) \cap \{z \in \mathbb{C},\ \mathrm{Im}(z) \geq 0\}$ le demi-cercle supérieur de centre x et de rayon r.

En considérant L et P des déterminations du logarithme et de la puissance α sur $\mathbb{C} \setminus i\mathbb{R}_-$ et en intégrant $z \mapsto \frac{P(z)L(z)}{z^2-1}$, montrer

$$\int_0^{+\infty} \frac{t^\alpha \ln(t)}{t^2-1} dt = \frac{\pi^2}{4\cos^2(\frac{\alpha\pi}{2})}.$$

4.5 *Théorème de l'application ouverte et conséquences

Théorème de l'application ouverte : Dans ce paragraphe, on analyse la question de l'inverse d'une fonction holomorphe, un peu comme on l'a fait pour le théorème de la bijection 3.14.

Commençons par montrer qu'une application holomorphe non constante est ouverte, c'est-à-dire que l'image d'un ouvert de l'espace de départ est un ouvert de l'espace d'arrivée :

Théorème 4.16. *Soit U un ouvert connexe de $\mathbb{C}$, et $f : U \to \mathbb{C}$ holomorphe et non constante. Alors $f(U)$ est ouvert dans $\mathbb{C}$.*

Démonstration. Soit $b = f(a) \in f(U)$. Comme f est non constante, par le théorème des zéros isolés 2.37 appliqué à la fonction $f - b$, il existe un disque D centré en a tel que $f - b$ ne s'annule pas sur $\overline{D} \setminus \{a\}$. Ainsi, en notant $C = \partial D$, on a

$$\rho := d(b, C) = \inf_{z\in C} |f(z) - b| > 0.$$

On va montrer que $D(b, \frac{\rho}{2}) \subset f(U)$, ce qui permet de conclure. Supposons par l'absurde qu'il existe $w \in D(b, \frac{\rho}{2}) \setminus f(U)$. On peut alors regarder la fonction $g : z \in U \mapsto \frac{1}{f(z)-w}$ qui est holomorphe sur U. D'après le principe du maximum 3.40 appliqué à g dans le disque $\overline{D}$, on obtient en passant à l'inverse :

$$|f(a) - w| \geq \min_{z\in C} |f(z) - w|$$

ce qui est une contradiction car

$$|f(a) - w| = |b - w| < \frac{\rho}{2}, \qquad \text{et} \quad \forall z \in C,\ |f(z) - w| \geq |f(z) - b| - |b - w| \geq \rho - \frac{\rho}{2} = \frac{\rho}{2}. \tag{4.2}$$

□

Remarque 4.17. On peut modifier un peu la fin de la preuve si on a fait l'exercice 3.66 (qu'on a en fait reprouvé ici). En effet avec les notations de la preuve, si $w \in D(b, \frac{\rho}{2})$ (on ne raisonne plus par l'absurde), alors on peut considérer $h : z \mapsto |f(z) - w|$ qui par continuité admet un minimum sur le compact $\overline{D}$. Par les calculs de (4.2), on obtient que ce minimum n'est pas sur C, et donc qu'il est un minimum local de g, donc par l'exercice 3.66, g s'annule en ce minimum, et donc w a bien un antécédent par f.

Exercice 4.18 (*Preuve du théorème de d'Alembert-Gauss)**.** Soit P un polynôme non constant. Montrer que $P(\mathbb{C})$ est ouvert et fermé dans $\mathbb{C}$, et en déduire le théorème de d'Alembert-Gauss.

****Comportement locale d'une fonction holomorphe :** on peut en fait affiner la compréhension du comportement local d'une fonction holomorphe, qui sera à changement de variable bijectif (et holomorphe) près comme le comportement des fonctions $z \mapsto \alpha + z^k$ pour $\alpha \in \mathbb{C}$ et $k \in \mathbb{N}^*$. On peut par exemple montrer :

Théorème 4.19. *Soit f une fonction holomorphe sur un voisinage de z_0. Alors il existe $m \in \mathbb{N}^*$ et $\varphi : V \to W$ une fonction biholomorphe*[85] *où V, W sont des voisinages ouverts de z_0 et 0 respectivement, tels que*

$$\forall z \in V,\ \ f(z) = f(z_0) + \varphi(z)^m.$$

On peut trouver une preuve dans [Min, Théorème 4.6.3], assez simple en utilisant un fonction logarithme, et dans [AM04, Théorème 8.3.6] par une méthode très différente. Voir aussi [Tau06, Théorème 8.7.1].

Remarque 4.20. En fait, m est la multiplicitié de z_0 comme racine de $z \mapsto f(z) - f(z_0)$ (voir exercice 2.58). Le théorème 4.19 implique que l'équation $f(z) = b$ aura m racines simples dans un voisinage de z_0 si b est proche mais différents de $f(z_0)$. Cela implique donc le théorème 4.16 (voir aussi [AM04, Corollaire 8.3.8].

85. C'est-à-dire holomorphe, bijective et d'inverse holomorphe.

****Théorème d'inversion locale pour les fonctions holomorphes :** Le théorème 4.19 a deux conséquences, qui sont une adaptation au cadre holomorphe des théorèmes d'inversions locale et globale :

Théorème 4.21 (Théorème d'inversion locale holomorphe)**.** *Soit $f : U \to \mathbb{C}$ holomorphe sur un ouvert connexe U. Soit $z_0 \in U$ tel que $f'(z_0) \neq 0$. Alors il existe V voisinage ouvert de z_0 et V voisinage ouvert de $f(z_0)$ tels que f est bijective de V dans W, et f^{-1} est holomorphe sur W.*

Ce résultat est une conséquence du théorème 4.19, voir [Tau06, Théorème 8.7.3], sinon on peut aussi regarder [Rud98, Théorème 10.30].

Enfin, on peut déduire (voir [Tau06, Corollaire 8.7.4] ou [Rud98, Théorème 10.33]).

Théorème 4.22 (Théorème d'inversion globale holomorphe)**.** *Soit f une fonction holomorphe et injective sur U un ouvert connexe de $\mathbb{C}$. Alors f' ne s'annule pas, et $f^{-1} : f(U) \to U$ est une fonction holomorphe.*

Remarque 4.23. Il est intéressant de comparer avec le théorème de la bijection, voir aussi l'exercice 3.24. On montre ici que l'inverse d'une fonction holomorphe bijective est elle-même holomorphe.

Il est également intéressant de comparer avec le théorème d'inversion global que nous verrons au Chapitre VII. En effet, ici on n'a pas demandé que la différentielle soit inversible, c'était une conséquence directe du caractère holomorphe et injectif de f.

4.6 *Automorphismes de $\mathbb{D}$ et $\mathbb{C}$

Voir par exemple [AM04, 10.3].

Etant donné un ouvert U de $\mathbb{C}$, on appelle un automorphisme de U une fonction $f : U \to U$ qui est holomorphe, bijective, et d'inverse holomorphe.

Exercice 4.24. Montrer que l'ensemble des automorphismes de U est un groupe pour la composition.

On va identifier ces groupes dans deux cas particuliers d'ensemble U :

Exercice 4.25 (Automorphismes de $\mathbb{C}$)**.**
1. Pour $\alpha \in \mathbb{C}^*$ et $\beta \in \mathbb{C}$, montrer que $f : z \mapsto \alpha z + \beta$ est un automorphisme de $\mathbb{C}$.
2. On va montrer la réciproque de la question précédente. Soit donc $f : \mathbb{C} \to \mathbb{C}$ un automorphisme.
   (a) Considérons la fonction
   $$g : z \in \mathbb{C}^* \longmapsto f\left(\frac{1}{z}\right).$$
   Montrer que l'ensemble $g(\{0 < |z| < 1\})$ n'est pas dense dans $\mathbb{C}$, et en déduire que f est un polynôme.
   (b) Montrer que f est de degré 1, et donc de la forme attendue.

Proposition 4.26 (Automorphismes de $\mathbb{D} = D(0,1)$)**.** *On pose pour $\alpha \in \mathbb{D}$, $\varphi_\alpha : z \in \mathbb{D} \mapsto \frac{z-\alpha}{1-\bar{\alpha}z}$. Alors l'ensemble des automorphismes de $\mathbb{D}$ est exactement*

$$\left\{\lambda\varphi_\alpha, \quad \alpha \in \mathbb{D}, \lambda \in \mathbb{C} \ \textit{ tel que } \ |\lambda| = 1\right\}$$

La preuve utilise le lemme de Schwarz donné ci-dessous, et le principe du maximum.

Lemme 4.27 (Lemme de Schwarz)**.** *Soit $f : D(0,1) \to \mathbb{C}$ holomorphe telle que $f(0) = 0$ et $\forall z \in D(0,1), |f(z)| < 1$. Alors*

$$\forall z \in D(0,1), \quad |f(z)| \leq |z|, \qquad \textit{et} \qquad |f'(0)| \leq 1$$

De plus, si $|f'(0)| = 1$ ou s'il existe $z \in D(0,1)$ tel que $|f(z)| = |z|$, alors il existe $\lambda \in \mathcal{U}$ tel que $\forall z \in D(0,1), \quad f(z) = \lambda z$.

Voir [Tau06, Théorème 7.3.1] ou [AM04, Théorème 4.9.1] pour une preuve.

4.7 **Preuves du théorème de Cauchy

Il existe plusieurs preuves du théorème de Cauchy et de la formule de Cauchy. Notamment, il y a des preuves assez simples (en tout cas beaucoup plus simples) si on suppose f holomorphe et de dérivée continue. D'ailleurs, certains auteurs ne s'encombrent pas, en imposant le caractère continu de la dérivée dans la définition d'une fonction holomorphe (voir par exemple [God03, page 311]). Ça n'est pas absurde si cela suffit pour les applications, ce qui est souvent le cas [86]. Certains auteurs se contentent aussi de montrer la formule de Cauchy dans le cas particulier des cercles, puisque cela suffit pour les principales applications.

On propose ici une de ces preuves sous forme d'exercice, et on donne ensuite la démonstration du cas d'un domaine convexe, mais sans supposer la continuité de la dérivée.

Exercice 4.28 (Formule de Cauchy sur des cercles avec dérivée continue)**.** Cet exercice est tiré de [Gou08b, 254] [87], on peut également trouver cette preuve dans [AM04, page 96]. Soit $f : D(0,R) \to \mathbb{C}$ une fonction $\mathbb{C}$-dérivable et de dérivée continue sur $D(0,R)$, avec $R \in]0,+\infty]$. On fixe $r \in]0,R[$, $z \in D(0,r)$, et on pose

$$\forall \lambda \in [0,1], \quad \varphi(\lambda) := \frac{1}{2i\pi} \int_{C(0,r)} \frac{f((1-\lambda)z + \lambda\zeta)}{\zeta - z} d\zeta.$$

La formule de Cauchy consiste à montrer $\varphi(1) = f(z)$, c'est donc l'objectif de l'exercice.

1. Montrer que $\varphi(0) = f(z)$.
2. Montrer que φ est dérivable et que
$$\forall \lambda \in [0,1], \quad \varphi'(\lambda) = \frac{1}{2i\pi} \int_{C(0,r)} f'((1-\lambda)z + \lambda\zeta) d\zeta. \tag{4.3}$$
3. En identifiant une primitive explicite pour le calcul de l'intégrale qui apparait dans la formule (4.3), montrer que pour tout $\lambda \in]0,1]$, $\varphi'(\lambda) = 0$ et conclure.

Pour la preuve du cas énoncé dans le théorème 3.24, on a recours au lemme suivant :

Lemme 4.29 (Lemme de Goursat)**.** *Soit U un ouvert de $\mathbb{C}$ et Δ un triangle plein* [88] *inclus dans U. Soit $f : U \to \mathbb{C}$ holomorphe sur U. Alors* [89]

$$\int_{\partial\Delta} f(z)dz = 0.$$

De plus, le résultat reste valable si f est continue sur U et holomorphe sur U privé d'un point. [90]

Démonstration. Commençons par le cas où on suppose f holomorphe sur tout l'ouvert U. On note (a,b,c) les sommets du triangle. Alors

$$\int_{\partial\Delta} f(z)dz = \int_{[a,b]} f(z)dz + \int_{[b,c]} f(z)dz + \int_{[c,a]} f(z)dz.$$

86. Par exemple pour la démonstration du théorème de d'Alembert-Gauss, ça ne coûte pas grand chose de vérifier que la fonction $\frac{1}{P}$ (où P est un polynôme qui ne s'annule pas) a une dérivée continue

87. Attention tout de même à cette référence : tout y est écrit sans vraiment avoir recours à la notion d'intégrale sur un chemin, ce qui n'est pas forcément naturel au niveau de l'agrégation. Aussi, au milieu de la preuve il est écrit "D'après la question a), φ est dérivable", ce que je trouve un peu étrange. Il faut à mon avis surtout citer un théorème de dérivabilité des intégrales à paramètre ; un théorème suffit largement, voir par exemple [Gou98, page 158] ou [Gou08b, page 157]. Mais c'est justement sur ce point que le caractère continu de f' intervient, ça vaut donc le coup de ne pas laisser trainer un flou à cet endroit.

88. On appelle triangle l'ensemble constitué de 3 segments $[a,b]$, $[b,c]$ et $[c,a]$ (où $(a,b,c) \in \mathbb{C}^3$), et triangle plein le triangle auquel on ajoute tous les points qui sont à l'intérieur. Autrement dit $\Delta = \text{Conv}(a,b,c)$.

89. Rigoureusement on n'a pas précisé la paramétrisation du triangle, en particulier pas son orientation ; en fait comme le résultat est nul, on peut choisir l'orientation qu'on veut. Néanmoins, dans la preuve ci-dessous, on fera attention à l'orientation des triangles considérés.

90. On a mis ce raffinement à part, car ça peut paraître artificiel et technique ; en fait ça ne coûte pas cher du point de vue de la preuve, et comme on l'a vu précédemment, cela a été utile pour montrer la formule de Cauchy.

On définit (a', b', c') les milieux respectifs des segments $[b, c]$, $[c, a]$ et $[a, b]$. Cela permet de définir 4 triangles pleins $(\Delta_i)_{1\leq i\leq 4}$ dont les bords sont orientés et respectivement délimités par les sommets $(a, c', b'), (b, a', c'), (c, b', a'), (a', b', c')$ [91]. En décomposant comme ci-dessus et en utilisant les exemples 3.11 et 3.12, on obtient :

$$\int_{\partial\Delta} f(z)dz = \sum_{i=1}^{4} \int_{\partial\Delta_i} f(z)dz.$$

Notons $M = |\int_{\partial\Delta} f(z)dz|$, dont on veut montrer la nullité, et L le périmètre du triangle Δ. Faisons maintenant deux remarques :

— chaque triangle Δ_i a un périmètre $\frac{L}{2}$ d'après le théorème de Thalès.

— comme $|\int_{\partial\Delta} f(z)dz| \leq \sum_{i=1}^{4} \left|\int_{\partial\Delta_i} f(z)dz\right|$, il existe $i_0 \in [\![1, 4]\!]$ tel que $\left|\int_{\partial\Delta_{i_0}} f(z)dz\right| \geq \frac{M}{4}$.

On va réitérer l'opération précédente pour ce triangle plein Δ_{i_0}, c'est-à-dire qu'on le découpe en 4 triangles $(\Delta_{i_0,j})_{1\leq j\leq 4}$, qui seront de périmètre $\frac{L}{2^2}$ et tel que l'un d'eux (indexé par $j_0 \in [\![1, 4]\!]$) satisfera $\left|\int_{\partial\Delta_{i_0,j_0}} f(z)dz\right| \geq \frac{M}{4^2}$, et ainsi de suite. On obtient donc une suite $(T_n)_{n\in\mathbb{N}}$ de triangles tel que pour tout $n \in \mathbb{N}$, T_n est de périmètre $\frac{L}{2^n}$ et tels que

$$\left|\int_{\partial T_n} f(z)dz\right| \geq \frac{M}{4^n}.$$

Par la propriété des fermés emboîtés [92] (issue de la complétude de $\mathbb{C}$), on a $\bigcap_{n\in\mathbb{N}} T_n$ est un singleton, qu'on note $\{z_0\}$. En effet, on remarque que le diamètre de T_n est inférieur à son périmètre.

Commençons par supposer que f est holomorphe sur tout U. Bien sûr $z_0 \in \Delta \subset U$, donc on peut utiliser la $\mathbb{C}$-dérivabilité de f en z_0 : étant donné $\varepsilon > 0$, il existe $\alpha > 0$ tel que $D(z_0, \alpha) \subset \mathbb{C}$ et

$$\forall z \in D(z_0, \alpha), \quad |f(z) - f(z_0) - f'(z_0).(z - z_0)| \leq \varepsilon|z - z_0|.$$

Posons $g : z \in \mathbb{C} \mapsto f(z_0) + f'(z_0).(z - z_0)$. Alors

$$\forall n \geq n_0, \quad \int_{\partial T_n} f(z)dz = \int_{\partial T_n} g(z)dz + \int_{\partial T_n} (f(z) - g(z))dz.$$

Clairement g a une primitive, donc par la proposition 3.21, $\int_{\partial T_n} g(z)dz = 0$. D'un autre côté, il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, $T_n \subset D(z_0, \alpha)$. Ainsi on peut écrire :

$$\frac{M}{4^{n_0}} \leq \left|\int_{\partial T_{n_0}} f(z)dz\right| = \left|\int_{\partial T_{n_0}} (f(z) - g(z))dz\right| \leq \text{Long}(\partial T_{n_0}) \sup_{\partial T_{n_0}} |f - g| \leq \varepsilon.\frac{L}{2^{n_0}} \sup_{z\in\partial T_{n_0}} |z - z_0| \leq \varepsilon\frac{L^2}{4^{n_0}}$$

ce qui donne $M \leq \varepsilon L^2$, et ceci étant vrai pour tout $\varepsilon > 0$ [93], on obtient bien $M = 0$.

Il nous reste à voir comment on doit adapter la preuve au cas où f est continue mais seulement holomorphe sur $U \setminus \{p\}$ pour un point p. Trois cas se présentent :

1. Si $p \notin \Delta$, la preuve précédente s'applique sans changement.
2. Si p est un sommet de Δ, disons $p = a$ avec les notations précédentes. On introduit alors deux suites $(a_n^{(1)})_{n\in\mathbb{N}}$ et $(a_n^{(2)})_{n\in\mathbb{N}}$ telle que

$$\forall n \in \mathbb{N}, \quad a_n^{(1)} \in]a, b], \quad a_n^{(2)} \in]a, c], \quad \text{et} \quad a_n^{(1)} \xrightarrow[n\to+\infty]{} a, \ a_n^{(2)} \xrightarrow[n\to+\infty]{} a.$$

Etant donné $n \in \mathbb{N}$, on découpe alors Δ en 3 triangles, à savoir ceux respectivement formés par les sommets orientés $(a_n^{(1)}, b, c), (a_n^{(1)}, b, a_n^{(2)})$ et $(a, a_n^{(1)}, a_n^{(2)})$. Alors $\int_{\partial\Delta} f(z)dz$ est la somme des intégrales sur chacun de ces triangles. On applique le cas 1. aux deux premiers triangles, puisque f y est partout holomorphe, les deux intégrales sont donc nulles.. Enfin, par continuité de f, l'intégrale sur le triangle $(a, a_n^{(1)}, a_n^{(2)})$ tend vers 0 quand $n \to +\infty$, donc au final $\int_{\partial\Delta} f(z)dz = 0$.

91. Faire un dessin !!!
92. Voir le paragraphe 1.4 du Chapitre III.
93. Attention : ici n_0 dépend de α qui lui même dépend de ε. La raison pour laquelle on a pu faire tendre ε vers 0 est que n_0 disparu de l'estimation quand on a écrit $M \leq \varepsilon L^2$.

3. Si p est à l'intérieur de Δ ou sur une arête privée de ses sommets, on subdivise Δ en triangles dont p sera un sommet, ce qui ramène au point 2.

□

Preuve de la première partie du Thérorème 3.24 : La preuve est désormais très similaire à celle du théorème 3.23 : par hypothèse, il existe $z_0 \in U$ tel que U est étoilé par rapport à z_0, et donc on peut poser

$$\forall z \in U, \quad F(z) = \int_{[z_0,z]} f(\zeta)d\zeta. \tag{4.4}$$

Étant donné h tel que $z + h \in U$, par convexité de U le triangle plein Δ formé par les sommets $(z_0, z, z+h)$ est dans U, donc $\int_{\partial\Delta} f(\zeta)d\zeta = 0$. Comme dans la preuve du théorème 3.23, on en déduit

$$F(z+h) - F(z) = \int_{[z,z+h]} f(\zeta)d\zeta,$$

et par suite que F est dérivable en z et $F'(z) = f(z)$. □

Remarque 4.30. La preuve précédente fonctionne si U est seulement étoilé par rapport à un point : en effet au lieu de prendre z_0 quelconque comme on l'a fait, on choisit z_0 le point qui rend U étoilé. Alors pour tout $z \in U$, le segment $[z_0, z]$ est inclus dans U, donc on peut bien définir (4.4). Ensuite, $z \in U$ étant donné, comme U est ouvert, il existe $\alpha > 0$ tel que $D(z, \alpha) \subset U$. Alors pour tout $h \in D(0, \alpha)$, le triangle formé par les sommets $(z_0, z, z+h)$ est bien dans U, car ce triangle est la réunion des segments

$$[z_0, (1-t)z + t(z+h)] \text{ pour } t \in [0,1]$$

qui sont bien dans U par hypothèse.

Remarque 4.31. À l'image du théorème 3.23, on a en fait montré que si $f : U \to \mathbb{C}$ est continue sur un ouvert étoilé U, alors f admet une primitive si et seulement si son intégrale sur tout triangle de U est nulle.

On conclut ce paragraphe par une réciproque du lemme de Goursat : même si la preuve va paraître expéditive, ce résultat n'est pas simple du tout, il utilise plusieurs des résultats précédents.

Théorème 4.32 (Théorème de Morera)**.** *Soit U un ouvert et $f : U \to \mathbb{C}$ continue sur U. On suppose que pour tout triangle plein $\Delta \subset U$, $\int_{\partial\Delta} f(\zeta)d\zeta = 0$. Alors f est holomorphe sur U.*

Démonstration. Soit $z_0 \in U$ et $r > 0$ tel que $D(z_0, r) \subset U$. Pour $z \in D(z_0, r)$, on peut poser

$$F(z) = \int_{[z_0,z]} f(\zeta)d\zeta.$$

En répétant la preuve de la première partie du théorème 3.24 ci-dessus, on obtient que F est holomorphe sur $D(z_0, r)$, et $F' = f$. Mais d'après le théorème 3.28, F est analytique, et d'après la proposition 2.13 sa dérivée aussi, et elle est donc holomorphe. □

4.8 **Homotopie et Théorie de Cauchy

4.8.1 Définition de l'homotopie

Commençons par remarquer qu'un chemin peut toujours être reparamétré sur $[0,1]$ en considérant $\varphi : t \in [0,1] \mapsto (1-t)a + tb \in [a,b]$. Pour cette raison, tous les chemins considérés ici seront paramétrés sur $[0,1]$.

Définition 4.33. Soit U un ouvert de $\mathbb{C}$, γ_0 et γ_1 deux lacets à valeurs dans U. Une homotopie entre γ_0 et γ_1 est une application continue $H : [0,1] \times [0,1] \to U$ telle que

— $\forall t \in [0,1], \quad H(0,t) = \gamma_0(t), \quad \text{et} \quad H(1,t) = \gamma_1(t)$

— $\forall s \in [0,1], \quad H(s,0) = H(s,1).$

On dit alors que γ_0 et γ_1 sont homotopes dans U [94].

Il peut être commode de noter $\gamma_s = H(s, \cdot)$. Ainsi, les deux conditions permettent de voir $s \mapsto \gamma_s$ comme une famille continue de lacets qui va de γ_0 vers γ_1. Autrement dit l'homotopie H est une déformation continue (et à valeurs dans U) du lacet γ_0 en le lacet γ_1.

Exercice 4.34. Soit γ un lacet dans un ouvert $U \subset \mathbb{C}$, et $\varphi : [0,1] \to [0,1]$ une application continue telle que $\varphi(0) = 0$ et $\varphi(1) = 1$. Montrer que les lacets γ et $\gamma \circ \varphi$ sont homotopes dans U.

On voit ainsi que la notion d'homotopie n'est pas affectée par les reparamétrisations.

Définition 4.35. Un ouvert U de $\mathbb{C}$ est dit simplement connexe s'il est connexe et si tout lacet de U est homotope à un point dans U [95].

Remarque 4.36. Intuitivement, cette notion est très claire. Elle signifie que l'ensemble U est en un morceau (connexité) et surtout, n'a pas de trou. Mais mathématiquement, elle n'est pas si simple à manipuler. Par exemple, l'ensemble $\mathbb{C}^*$ n'est pas simplement connexe : autant il est évident qu'il a un "trou". Mais il n'est pas forcément évident de le prouver avec la définition. On le prouvera avec le théorème de Cauchy ci-dessous.

Exercice 4.37. Donnons quelques exemples d'ensembles simplement connexes.

1. Montrer qu'un ouvert étoilé est simplement connexe (et donc un ensemble convexe l'est également).
2. Montrer que l'ensemble $\mathbb{C} \setminus \mathbb{R}_-$ est simplement connexe.

Exercice 4.38.
1. Montrer que si U est un ouvert de $\mathbb{C}$, alors l'homotopie dans U est une relation d'équivalence sur l'ensemble des lacets à valeurs dans U.
2. Montrer que si U et V sont deux ouverts homéomorphes de $\mathbb{C}$, et si U est simplement connexe, alors V est simplement connexe.

4.8.2 Théorème de Cauchy homotope

Nous donnons désormais un résultat qui va remplacer le théorème 3.24 :

Théorème 4.39. *Soit U un ouvert de $\mathbb{C}$, et $f : U \to \mathbb{C}$ holomorphe. Soit γ_0 et γ_1 deux lacets C^1 par morceaux qui sont homotopes dans U. Alors*

$$\int_{\gamma_0} f(z)dz = \int_{\gamma_1} f(z)dz.$$

En particulier, si γ est un lacet C^1 par morceaux homotope à un point dans U, alors

$$\int_{\gamma} f(z)dz = 0.$$

On a besoin d'un lemme préparatoire, affirmant qu'un lacet peut être approché de façon uniforme par un lacet affine par morceaux. L'idée est semblable à l'approximation d'une fonction continue sur un segment qu'on a approchée par une fonction constante par morceaux dans la construction de l'intégrale de Riemann, voir la Définition-Proposition 5.5.

Lemme 4.40. *Soit U un ouvert. Pour tout chemin $\gamma : [0,1] \to U$ et $\varepsilon > 0$, il existe un chemin affine par morceaux $\widetilde{\gamma} : [0,1] \to U$ tel que*

$$\gamma(0) = \widetilde{\gamma}(0), \quad \gamma(1) = \widetilde{\gamma}(1), \quad \sup_{[0,1]} |\widetilde{\gamma} - \gamma| \leq \varepsilon, \quad \textit{et} \quad \Big[\, \widetilde{\gamma} \textit{ est homotope à } \gamma \textit{ dans } U \,\Big].$$

94. On oublie souvent de dire "dans U", mais bien sûr le caractère homotope ou non de deux lacets dépend complètement de l'ouvert dans lequel on travaille. On s'efforcera, même si cela alourdit un peu le discours, de toujours le préciser.
95. Un lacet homotope à un point est parfois appelé contractile.

Démonstration. Par continuité de γ, l'ensemble $\mathrm{Im}(\gamma)$ est un compact (voir 1.5 au Chapitre III), donc sa distance au fermé $\mathbb{C} \setminus U$ est strictement positive : supposons $0 < \varepsilon < d(\mathrm{Im}(\gamma), \mathbb{C} \setminus U)$ (la propriété recherchée étant plus difficile pour ε petit, on peut faire cette hypothèse sans perte de généralité). Par le théorème de Heine, la fonction γ est uniformément continue, donc il existe $\alpha > 0$ tel que

$$\forall (t,t') \in [0,1]^2, \ \ |t-t'| \leq \alpha \ \Rightarrow \ |\gamma(t) - \gamma(t')| \leq \frac{\varepsilon}{2}.$$

On considère[96] une subdivision $0 = t_0 < t_1 < \ldots < t_N = 1$ telle que pour tout $i \in [\![0, N-1]\!]$, $|t_{i+1} - t_i| < \alpha$. On considère alors $\widetilde{\gamma}$ la concaténation des chemins $[\gamma(t_i), \gamma(t_{i+1})]$ pour $i \in [\![0, N-1]\!]$.

Vérifions que $\widetilde{\gamma}$ convient. Les valeurs en $0 = t_0$ et $1 = t_N$ sont bien les mêmes. Si $t \in [0,1]$, il existe $i \in [\![0, N-1]\!]$ tel que $t \in [t_i, t_{i+1}]$ et alors

$$|\widetilde{\gamma}(t) - \gamma(t)| \leq |\widetilde{\gamma}(t) - \gamma(t_i)| + |\gamma(t_i) - \gamma(t)| \leq |\gamma(t_{i+1}) - \gamma(t_i)| + \frac{\varepsilon}{2} \leq \varepsilon.$$

Pour le caractère homotope, on pose $\forall (s,t) \in [0,1]^2, H(s,t) = (1-s)\gamma(t) + s\widetilde{\gamma}(t)$ qui est bien une homotopie entre γ et $\widetilde{\gamma}$, il faut vérifier qu'elle reste à valeurs dans U. Etant donné $(s,t) \in [0,1]^2$, on a $|H(s,t) - \gamma(t)| = s|\widetilde{\gamma}(t) - \gamma(t)| \leq \varepsilon$, mais comme $d(\mathrm{Im}(\gamma), \mathbb{C} \setminus U) > \varepsilon$, on a

$$\forall x \in \mathrm{Im}(\gamma), \forall y \in \mathbb{C} \setminus U, d(x,y) > \varepsilon,$$

ce qui montre que $H(s,t) \in U$. On en déduit en particulier pour $s = 1$ que $\widetilde{\gamma}$ est bien à valeurs dans U, et donc ceci conclut la preuve. □

Preuve du théorème 4.39 :

Étape 1 : Supposons d'abord γ_0 et γ_1 suffisamment proches. Plus précisément, on pose $r_{\gamma_0} = \frac{1}{2} d(\mathrm{Im}(\gamma_0), \mathbb{C} \setminus U)$ et on suppose $\sup_{[0,1]} |\gamma_1 - \gamma_0| < r_{\gamma_0}$. Comme dans la preuve du lemme 4.40, par uniforme continuité de γ_0, il existe $t_0 = 0 < t_1 < \ldots < t_N = 1$ tels que

$$\forall i \in [\![0, N-1]\!], \forall t \in [t_i, t_{i+1}], \ \ |\gamma_0(t) - \gamma_0(t_i)| < r_{\gamma_0}.$$

On pose alors, pour $i \in [\![0, N-1]\!]$, ℓ_i est le lacet obtenu en concaténant

$$\gamma_{0|[t_i, t_{i+1}]}, \quad [\gamma_0(t_{i+1}), \gamma_1(t_{i+1})], \ \text{l'opposé de } \gamma_{1|[t_i, t_{i+1}]}, \quad [\gamma_1(t_i), \gamma_0(t_i)].$$

On constate rapidement que $\mathrm{Im}(\ell_i) \subset D(\gamma_0(t_i), 2r_{\gamma_0}) \subset U$. Donc on peut appliquer le théorème 3.24 dans le convexe $D(\gamma_0(t_i), 2r_{\gamma_0})$, ce qui donne $\int_{\ell_i} f(z)dz = 0$, et ce pour tout $i \in [\![0, N-1]\!]$. Or par la relation de Chasles,

$$\sum_{i=0}^{N-1} \int_{\ell_i} f(z)dz = \int_{\gamma_0} f(z)dz - \int_{\gamma_1} f(z)dz$$

car les segments qui interviennent dans les chemins ℓ_i se composent car ils sont apparaissent deux fois avec des orientations opposées. D'où le résultat dans le cas où γ_0 et γ_1 sont proches.

Étape 2 : Pour le cas général, on va procéder en utilisant successivement l'étape 1 le long de $s \mapsto H(s, \cdot)$ où H est l'homotopie entre γ_0 et γ_1. C'est là que le lemme 4.40 va entrer en jeu : en effet, comme H est seulement continue, on ne pourrait pas intégrer sur $H(s, \cdot)$ car celui-ci n'est pas un chemin C^1 par morceaux, on le remplacera donc par une approximation affine par morceaux.

Par continuité de H et compacité de $[0,1] \times [0,1]$, on a $r := \frac{1}{2} d(\mathrm{Im}(H), \mathbb{C} \setminus U) > 0$. Par uniforme continuité de H, on trouve une subdivision $s_0 = 0 < s_1 < \ldots < s_p = 1$ telle que

$$\forall t \in [0,1], \ \ \forall i \in [\![0, p-1]\!], \ \ |H(s_{i+1}, t) - H(s_i, t)| \leq \frac{r}{4}.$$

Pour chaque $i \in [\![1, p-1]\!]$, on note $\ell_i = \widetilde{H(s_i, \cdot)}$ donné par le lemme 4.40 avec $\varepsilon = \frac{r}{4}$, ainsi que $\ell_0 = \gamma_0$ et $\ell_p = \gamma_p$. Alors pour chaque $i \in [\![0, p-1]\!]$,

$$|\ell_{i+1} - \ell_i| \leq |\ell_{i+1} - H(s_{i+1}, \cdot)| + |H(s_{i+1}, \cdot) - H(s_i, \cdot)| + |H(s_i, \cdot) - \ell_i| \leq \frac{3r}{4} < \frac{1}{2} d(\mathrm{Im}(\ell_i), \mathbb{C} \setminus U)$$

qui est la condition pour appliquer l'étape 1, d'où $\int_{\ell_{i+1}} f(z)dz = \int_{\ell_i} f(z)dz$, pour tout $i \in [\![0, p-1]\!]$. Par récurrence immadiate, on obtient le résultat. □

96. On devrait dire, “il existe”, ou mieux encore, en exhiber une.

4.8.3 Conséquences

On donnera certaines conséquences du théorème de Cauchy homotopique sous la forme d'exercices, mais chaque résultat est utile en soi.

Corollaire 4.41. *Si U est un ouvert simplement connexe, alors toute fonction holomorphe sur U admet une primitive sur U.*

Démonstration. C'est une conséquence directe du théorème 4.39 et du théorème 3.23. □

Exercice 4.42. Soit U un ouvert simplement connexe de $\mathbb{C}$, et $f : U \to \mathbb{C}$ une fonction holomorphe qui ne s'annule pas. En considérant une primitive de la fonction $\frac{f'}{f}$, montrer qu'il existe $\phi : U \to \mathbb{C}$ holomorphe telle que $f = \exp \circ \phi$.

Exercice 4.43.
1. Montrer que si U est un ouvert simplement connexe, alors pour toute fonction $f : U \to \mathbb{C}$ holomorphe et tout lacet dans U, $\int_\gamma f(z)dz = 0$.
2. En déduire que $\mathbb{C}^*$ n'est pas simplement connexe.

Avec l'exercice 4.38, on en déduit que $\mathbb{R}^2$ et $\mathbb{R}^2 \setminus \{(0, 0)\}$ ne sont pas homéomorphes.

Exercice 4.44. Soit $z \in \mathbb{C}$, et γ_0, γ_1 deux lacets C^1 par morceaux et homotopes dans $\mathbb{C} \setminus \{z\}$. Montrer

$$\mathrm{ind}_{\gamma_0}(z) = \mathrm{ind}_{\gamma_1}(z).$$

Remarque 4.45. En rapport à l'exercice précédent, on peut montrer que l'indice par rapport à 0 caractérise les lacets dans $\mathbb{C}^*$, à homotopie près. Plus précisément, tout lacet γ de $\mathbb{C}^*$ est homotope au lacet $t \in [0, 1] \mapsto e^{2i\pi nt}$ (qui consiste à tracer $|n|$ cercles orientés selon le signe de n) où $n = \mathrm{ind}_\gamma(0)$.

Théorème 4.46 (Formule de Cauchy bis)**.** *Soit U un ouvert de $\mathbb{C}$, $f : U \to \mathbb{C}$ holomorphe, γ un lacet C^1 par morceaux et homotope à un point dans U. Alors*

$$\forall z \notin \mathrm{Im}(\gamma), \quad f(z) ind_\gamma(z) = \frac{1}{2i\pi} \int_\gamma \frac{f(\zeta)}{\zeta - z} d\zeta.$$

Démonstration. Comme dans cette partie de la preuve du théorème 3.24, on définit g comme à l'équation (3.3). Par la proposition 3.32, la fonction g ainsi définie est holomorphe sur U, donc par le théorème 4.39, $\int_\gamma g(z)dz = 0$, d'où le résultat. □

4.8.4 Développement en série de Laurent et théorème des résidus améliorés

À l'image du théorème 3.28, le résultat suivant donne une réciproque au fait qu'une fonction définie par une série de Laurent est holomorphe dans une couronne.

Théorème 4.47. *Soit f holomorphe sur un anneau $A(z_0, r, R)$. Alors f admet un unique développement en série de Laurent : il existe une unique suite $(a_n)_{n\in\mathbb{Z}}$ telle que*

$$\forall z \in A(z_0, r, R), \quad f(z) = \sum_{n\in\mathbb{Z}} a_n (z - z_0)^n.$$

De plus, les coefficients sont donnés par :

$$\forall n \in \mathbb{Z}, a_n = \frac{1}{2i\pi} \int_{C(z_0,r')} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} d\zeta \tag{4.5}$$

pour n'importe quel $r' \in]r, R[$.

Démonstration. Commençons par prouver une variante de la formule de Cauchy pour des anneaux : si $r < r_1 < r_2 < R$, et $z \in A(z_0, r_1, r_2)$, alors

$$f(z) = \frac{1}{2i\pi} \int_{C(z_0,r_2)} \frac{f(\zeta)}{\zeta - z} d\zeta - \frac{1}{2i\pi} \int_{C(z_0,r_1)} \frac{f(\zeta)}{\zeta - z} d\zeta. \tag{4.6}$$

En effet, on introduit la fonction g définie par (3.3) dans le cas $U = A(z_0, r, R)$, qui est holomorphe. Les cercles $C(z_0, r_1)$ et $C(z_0, r_2)$ sont homotopes, donc

$$\int_{C(z_0,r_1)} g(\zeta)d\zeta = \int_{C(z_0,r_2)} g(\zeta)d\zeta$$

ce qui donne bien (4.6) puisque $\mathrm{ind}_{C(z_0,r_2)}(z) = 1$ et $\mathrm{ind}_{C(z_0,r_1)}(z) = 0$.

À partir de (4.6), en procédant comme dans la preuve du théorème 3.28 avec la formule (3.5), on obtient l'expression souhaitée avec

$$a_n = \frac{1}{2i\pi}\int_{C(z_0,r_2)} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} \text{ si } n \geq 0, \quad a_n = \frac{1}{2i\pi}\int_{C(z_0,r_1)} \frac{f(\zeta)}{(\zeta - z_0)^{n+1}} \text{ si } n < 0.$$

Du fait que pour tout $r' \in]r, R[$, les cercles $C(z_0, r')$ sont homotopes entre eux dans l'anneau $A(z_0, r, R)$, on peut prendre $r_1 = r'$ et $r_2 = r'$ dans les formules précédentes.

Pour l'unicité, on observe que si $f(z) = \sum_{n\in\mathbb{Z}} a_n(z-z_0)^n$, alors les coefficients a_n sont nécessairement donnés par (4.5) : en effet, par convergence normale de la série, si $r' \in]r, R[$, alors

$$\frac{1}{2i\pi}\int_{C(z_0,r')} \frac{f(\zeta)}{(\zeta - z_0)^{m+1}}d\zeta = \sum_{n\in\mathbb{Z}} \frac{a_n}{2i\pi}\int_{C(z_0,r')} (\zeta - z_0)^{n-m-1}d\zeta,$$

avec [97]

$$\forall k \in \mathbb{Z}, \quad \frac{1}{2i\pi}\int_{C(z_0,r')} (\zeta - z_0)^k d\zeta = \delta_{k,-1},$$

d'où le résultat. □

Théorème 4.48 (Théorème des résidus). *Soit U un ouvert de $\mathbb{C}$, $(z_1, z_2, \ldots, z_p)$ un nombre fini de points de U deux à deux distincts. Soit $f : U \setminus \{z_1, z_2, \ldots, z_p\}$ holomorphe, et γ un lacet C^1 par morceaux, à valeurs dans $U \setminus \{z_1, z_2, \ldots, z_p\}$, et homotope à un point dans U. Alors*

$$\frac{1}{2i\pi}\int_\gamma f(z)dz = \sum_{k=1}^{p} \mathrm{Res}(f, z_k)\mathrm{ind}_\gamma(z_k). \tag{4.7}$$

Démonstration. La preuve est la même que pour le théorème 3.58, mais en utilisant le Théorème 4.46. □

4.9 Etude de fonctions classiques

Comme souvent à l'agrégation, il faut savoir manipuler des exemples concrets de fonctions holomorphes. Ici les exemples ne manquent pas. On a déjà traité les fonctions exponentielles et logarithme, et beaucoup de choses peuvent aussi être dites sur la fonction Γ ou sur la fonction ζ introduites à l'exercice 3.68.

4.10 *Détection de zéros d'une fonction holomorphe

La formule des résidus fournit de moyens de compter le nombre de zéros d'une fonction holomorphe :

Lemme 4.49 (Principe de l'argument). *Soit U un ouvert de $\mathbb{C}$ contenant le disque fermé $\overline{D}(z_0, r)$, et $f : U \to \mathbb{R}$ holomorphe, ne s'annulant pas sur le cerle $C(z_0, r)$. Alors*

$$\frac{1}{2i\pi}\int_{C(z_0,r)} \frac{f'(z)}{f(z)}dz = \mathrm{ind}_{f(C(z_0,r))}(0) \tag{4.8}$$

est égal au nombre de zéros de f dans le disque $D(z_0, r)$, comptées avec multiplicités.

97. On utilise le symbole de Kronecker,

$$\delta_{x,y} = \begin{cases} 1 & \text{si } x = y \\ 0 & \text{sinon} \end{cases}$$

Démonstration. La formule (4.8) se vérifie en considérant $\gamma : t \in [0,1] \mapsto e^{2i\pi t}$ et $f \circ \gamma$ qui paramétrisent respectivement $C(z_0, r)$ et $f(C(z_0, r))$ respectivement.

Par le principe des zéros isolés, f a un nombre fini de zéros dans $\overline{D}(z_0, r)$, qu'on note $z_1, \dots, z_N$, et $m_1, \dots, m_N$ leurs multiplicités. Pour $k \in [\![1, N]\!]$, on peut écrire pour z dans un voisinage de z_k :

$$f(z) = (z - z_k)^{m_k} g_k(z), \quad \text{où } g_k(z_k) \neq 0, \qquad \text{et} \quad \frac{f'(z)}{f(z)} = \frac{g_k'(z)}{g_k(z)} + \frac{m_k}{z - z_k}.$$

On peut appliquer le théorème 3.58 des résidus, ce qui donne

$$\frac{1}{2i\pi} \int_{C(z_0,r)} \frac{f'(z)}{f(z)} dz = \sum_{k=1}^{N} \operatorname{Res}\left(\frac{f'}{f}, z_k\right) = \sum_{k=1}^{N} m_k$$

car l'indice du cercle $C(z_0, r)$ est 1 pour chaque z_k. □

Remarque 4.50. Ce lemme permet de donner une deuxième approche au résultat mentionné dans la remarque 4.20 qui décrit le nombre de solutions de l'équation $f(z) = w$ localement, ainsi qu'au fait que si f est injective, alors sa dérivée ne s'annule pas, voir par exemple page 86 de ce lien.

Exercice 4.51. Soit U un ouvert de $\mathbb{C}$ contenant le disque fermé $\overline{D}(z_0, r)$, et $f : U \to \mathbb{R}$ méromorphe, n'ayant aucun zéro ni pôle sur le cerle $C(z_0, r)$. Calculer $\frac{1}{2i\pi} \int_{C(z_0,r)} \frac{f'(z)}{f(z)} dz$ en fonction du nombre de zéros et de pôles de f.

Citons deux autres conséquences classiques du principe de l'argument. Attention, la démonstration du premier résultat (et le second en est une conséquence) utilise les résultats du paragraphe 4.8.

Théorème 4.52 (**Théorème de Rouché)**.** *Soit U un ouvert de $\mathbb{C}$ contenant le disque $\overline{D}(z_0, r)$, et f, g deux fonctions holomorphes sur U telles que*

$$\forall z \in C(z_0, r), \quad |f(z) - g(z)| < |f(z)|.$$

Alors f et g ont le même nombre de zéros, comptés avec multiplicité, dans le disque $D(z_0, r)$.

Démonstration. D'après le lemme 4.49, il s'agit de montrer que l'indice des lacets $f(C(z_0, r))$ et $g(C(z_0, r))$ sont égaux. Par l'exercice 4.44 lui-même conséquence du théorème 4.39, il suffit de montrer que les deux chemins sont homotopes dans $\mathbb{C}^*$. On considère

$$\forall (s,t) \in [0,1]^2, \quad H(s,t) = (1-s) f \circ \gamma(t) + s g \circ \gamma(t)$$

où $\gamma(t) = e^{2i\pi t}$. C'est bien une homotopie dans $\mathbb{C}^*$ car avec l'hypothèse,

$$\forall (s,t) \in [0,1]^2, \quad |H(s,t)| \geq |f(\gamma(t))| - |(g - f)(\gamma(t))| > 0.$$

□

Corollaire 4.53 (**Théorème de Hurwitz)**.** *Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions holomorphes sur un ouvert connexe U de $\mathbb{C}$. On suppose que la suite $(f_n)_{n \in \mathbb{N}}$ converge uniformément sur tout compact vers une fonction $f : U \to \mathbb{C}$ (qui sera donc holomorphe).*

- *Si pour tout $n \in \mathbb{N}$ la fonction f_n ne s'annule pas sur U, alors soit f est identiquement nulle, soit f ne s'annule pas.*
- *Si pour tout $n \in \mathbb{N}$ la fonction f_n est injective sur U, alors soit f est constante, soit f est injective.*

Démonstration. Pour le premier point, on raisonne par contraposée en supposant que f n'est pas identiquement nulle, mais s'annule en $z_0 \in U$. Par le principe des zéros isolés, il existe $r_0 > 0$ tel que f ne s'annule pas sur $C(z_0, r_0)$. Par compacité du cercle et continuité de f, il existe $\varepsilon_0 > 0$ tel que $|f| \geq \varepsilon_0$ sur $C(z_0, r_0)$. Par convergence de $(f_n)_{n \in \mathbb{N}}$, il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, $|f_n - f| < \varepsilon_0 \leq |f|$. On peut appliquer le théorème de Rouché à f_{n_0} et f dans le disque $D(z_0, r_0)$. Comme f s'annule en z_0, f_{n_0} doit s'annuler dans le disque $D(z_0, r_0)$, et donc on a prouvé la contraposée de la première assertion.

Pour le second point, on suppose f non constante, et $z_1 \neq z_2$ tels que $f(z_1) = f(z_2) = w$. Quitte à soustraire w à toutes les fonctions, on peut supposer $w = 0$. Le point précédent, appliqué près de z_1 et z_2 montre que pour tout r suffisamment petit, il existe n_0 tel que f_{n_0} doit s'annuler dans $D(z_1, r)$ et dans $D(z_2, r)$. On choisit r tel que ces deux disques soit disjoints, et alors f_{n_0} ne sera pas injective. □

4.11 *Fonctions méromorphes

Préalablement, le titre de la leçon **245** était “Fonctions holomorphes et méromorphes sur un ouvert de $\mathbb{C}$. Exemples et applications”, et la référence explicite aux fonctions méromorphe a été retirée vers 2014. Néanmoins, il faut connaître la définition d'une fonction méromorphe (qui apparaît dans le programme), comme le suggère le rapport de jury sur cette leçon :

Définition 4.54. Soit U un ouvert de $\mathbb{C}$. On dit que f est une fonction méromorphe sur U s'il existe $A \subset U$ un ensemble fermé de U constitué de points isolés tel que

— f est holomorphe sur $U \setminus A$,

— tout point de A est un pôle de f.

Remarque 4.55. Attention à la terminologie, on parle bien d'une fonction méromorphe sur U alors que la fonction ne sera pas définie en tout point de U.

La seconde condition interdit à f d'avoir des singularités essentielles en les points de A. Quant aux possibles singularités effaçables de f, il n'est pas naturel de les inclure dans les singularités de f puisqu'on peut la prolonger de façon holomorphe en ces points.

Enfin, remarquons que la condition faite sur A est équivalente au fait d'être localement fini dans U, c'est-à-dire que pour tout compact $K \subset U$, l'ensemble $K \cap A$ doit être fini.

Exemple 4.56. Bien sûr, les fonctions holomorphes sur U sont méromorphes sur U (on prend $A = \emptyset$). Mais les exemples typiques de fonctions méromorphes sont les fractions rationnelles, qui sont méromorphes sur $\mathbb{C}$: l'ensemble A est constitué de l'ensemble des racines du dénominateur, qui est fini.

Exemple 4.57. La fonction $z \mapsto \frac{1}{\sin \frac{1}{z}}$ est une fonction méromorphe sur $\mathbb{C}^*$ car ses singularités sont les $\left\{\frac{1}{n\pi}, n \in \mathbb{Z}^*\right\}$ qui sont bien des pôles et qui sont isolées. Par contre, elle n'est pas méromorphe sur $\mathbb{C}$ car 0 est une singularité essentielle, qui de surcroît n'est plus isolée.

Exercice 4.58. Soit f le quotient de deux fonctions holomorphes non nulles sur U. Montrer que f est une fonction méromorphe.

Exercice 4.59. Montrer que l'ensemble des fonctions méromorphes sur U un ouvert connexe de $\mathbb{C}$ est un corps.

Remarque 4.60. L'exercice 4.58 admet une réciproque ([Tau06, Corollaire 11.2.5], pas du tout évidente, qui est conséquence d'un théorème dit de Weierstrass [98]. En conséquence, l'ensemble des fonctions méromorphes s'avère être le corps des fractions de l'ensemble des fonctions holomorphes, dont on a vu à l'exercice 2.59 qu'il s'agit d'un anneau intègre.

Il existe un théorème de convergence uniforme pour une série de fonctions méromorphes (attention, il y a une condition pour contrôler les pôles, voir [BMP05, page 71] ou [Tau06, Définition 8.8.2 et théorème 8.8.4]), qui permettent de montrer assez élégamment les formules

$$\forall z \in \mathbb{C} \setminus \mathbb{Z}, \ \left(\frac{\pi}{\sin(\pi z)}\right)^2 = \sum_{n \in \mathbb{Z}} \frac{1}{(z-n)^2}, \quad \text{et} \quad \pi^2 \frac{\cos(\pi z)}{\sin^2(\pi z)} = \sum_{n \in \mathbb{Z}} \frac{(-1)^n}{(z-n)^2}$$

voir [Tau06, 8.8.6], [BMP05, Exercice 2.14],

4.12 **Théorème de représentation conforme de Riemann

Ce théorème assez mythique affirme que tout ouvert simplement connexe distinct de $\mathbb{C}$ est en bijection holomorphe avec le disque unité de $\mathbb{C}$. C'est un théorème difficile qui requiert beaucoup de prérequis hors-programme sur les fonctions holomorphes, incluant plusieurs des paragraphes précédents, ainsi que des éléments d'analyse fonctionnelle qu'on abordera au Chapitre III.

98. Rien à voir avec la densité des polynômes dans les fonctions continues...

Théorème 4.61. *Si U est un ouvert simplement connexe de $\mathbb{C}$, non vide et distinct de $\mathbb{C}$, alors il existe une application f biholomorphe*[99] *entre U et $\mathbb{D} = D(0,1)$.*

---

5 Extraits d'annales

Exemple 5.1 (2012, Partie 1)**.** Soit $n \in \mathbb{N}^*$ et $(w_1, \ldots, w_n) \in \mathbb{C}^n$ tel que $|w_1 + \ldots + w_n| = |w_1| + \ldots + |w_n|$.

1. Montrer que pour j, l dans $\{1, \ldots n\}$ distincts, on a $\mathrm{Re}(\overline{w_j} w_l) = |w_j||w_l|$.
2. En déduire qu'il existe $\theta \in [0, 2\pi[$ tel que pour tout $j \in \{1, \ldots n\}$, $w_j = e^{i\theta}|w_j|$.

Exemple 5.2 (2020, Partie I)**.**

1. Montrer que la série de fonctions $\displaystyle\sum_{n \geq 1} \frac{(-1)^n}{n} x^n$ converge uniformément sur $[0,1]$ et que sa somme est une fonction continue sur $[0,1]$.
2. Déterminer le rayon de convergence r de la série entière $\displaystyle\sum_{n \geq 1} \frac{(-1)^n}{n} x^n$ et rappeler la valeur de sa somme sur $]-r, r[$.
3. En déduire la valeur de $\displaystyle\sum_{n=1}^{+\infty} \frac{(-1)^n}{n}$.

Exemple 5.3 (Docteur 2019)**.** Pour une suite $a = (a_n)_{n \in \mathbb{N}}$ de nombres complexes, on pose :

$$A(a) = \{r \geq 0 \quad \text{tel que } (|a_n| r^n)_{n \in \mathbb{N}} \text{ est majorée}\},$$

$$B(a) = \left\{r \geq 0 \quad \text{tel que } \lim_{n \to +\infty} a_n r^n = 0\right\},$$

$$C(a) = \left\{r \geq 0 \quad \text{tel que } \sum_{n \in \mathbb{N}} a_n r^n \text{ est convergente}\right\}.$$

1. Justifier les inclusions $C(a) \subset B(a) \subset A(a)$, et montrer que ces inclusions peuvent être strictes.
2. Montrer que dans $\mathbb{R}_+ \cup \{+\infty\}$, on a
$$\sup A(a) = \sup B(a) = \sup C(a).$$
On rappelle que le rayon de convergence de la série entière $\displaystyle\sum_{n \in \mathbb{N}} a_n z^n$ correspond à la borne supérieure de l'ensemble $A(a)$ dans $\mathbb{R}_+ \cup \{+\infty\}$.
3. Déterminer les rayons de convergence des séries entières :
   (a) $\displaystyle\sum_{n=1}^{\infty} \frac{1+i}{n} z^{n^2}$ (avec $i^2 = -1$),
   (b) $\displaystyle\sum_{n=0}^{\infty} 2^{(-1)^n n} z^n$,
   (c) $\displaystyle\sum_{n=0}^{\infty} \cos(2^n) z^n$.

99. C'est-à-dire holomorphe, bijective et d'inverse holomorphe, même si cette 3ième condition est conséquence des deux premières d'après le théorème 4.22.

4. On suppose que la série entière $\sum_{n=0}^{\infty} a_n z^n$ a un rayon de convergence $R \in]0, +\infty[$. Montrer que la série entière $\sum_{n=0}^{\infty} \frac{a_n}{n!} z^n$ a un rayon de convergence infini.
5. Si la série entière $\sum_{n=0}^{\infty} \frac{a_n}{n!} z^n$ a un rayon de convergence égal à 2, que peut-on dire du rayon de convergence de $\sum_{n=0}^{\infty} a_n z^n$? Donner un tel exemple de série entière.

Exemple 5.4 (Docteur sujet 0)**.** 1. Pour $x \in \mathbb{R}$, justifier l'existence de

$$f(x) = \sum_{n=0}^{\infty} \frac{e^{i2^n x}}{n!}.$$

2. Montrer que la fonction de $\mathbb{R}$ dans $\mathbb{C}$ ainsi définie est de classe C^∞ sur $\mathbb{R}$. Pour $k \in \mathbb{N}$, donner une expression simple de $f^{(k)}(0)$.
3. Montrer que le rayon de convergence de la série entière

$$\sum_{k \geq 0} \frac{f^{(k)}(0)}{k!} x^k$$

est nul.

Exemple 5.5 (2015, Partie I)**.** On considère les fonctions φ et H_k ($k \in \mathbb{N}$) définies sur $\mathbb{R}$ par :

$$\forall x \in \mathbb{R}, \quad \varphi(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}, \qquad H_k(x) = (-1)^k \frac{\varphi^{(k)}(x)}{\varphi(x)}.$$

1. Soit $x \in \mathbb{R}$. On considère la fonction $\psi : \mathbb{R} \to \mathbb{R}$ définie par :

$$\forall t \in \mathbb{R}, \quad \psi(t) = \varphi(x - t).$$

   (a) Rappeler les développements en série entière des fonctions $t \mapsto e^{2xt}$ et $t \mapsto e^{-t^2}$ en $t = 0$, ainsi que leurs rayons de convergence.
   (b) Etablir que la fonction ψ est développable en série entière en 0.
   (c) En déduire que :

$$\forall (x,t) \in \mathbb{R}^2, \quad \sum_{k=0}^{+\infty} \frac{H_k(x)}{k!} t^k = e^{2xt - t^2}. \tag{5.1}$$

2. Soit $x \in \mathbb{R}$. Justifier que l'on peut dériver terme à terme par rapport à t l'équation (5.1) et en déduire que :

$$\forall k \geq 1, \ \forall x \in \mathbb{R}, \ H_{k+1}(x) = 2xH_k(x) - 2kH_{k-1}(x).$$

3. Soit $k \in \mathbb{N}$. Prouver que H_k est une fonction polynomiale à coefficients réels. Préciser son degré et son coefficient dominant.

Exemple 5.6 (2020, Partie III)**.** Pour $s \in \mathbb{C}$ tel que la série converge, on note

$$\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}, \qquad G(s) = \sum_{n=1}^{+\infty} \frac{(-1)^n}{n^s}.$$

1. Montrer que les séries de fonctions $\sum_{n=1}^{+\infty} \frac{1}{n^s}$ et $\sum_{n=1}^{+\infty} \frac{(-1)^n}{n^s}$ convergent simplement dans le demi-plan $\{s \in \mathbb{C} \text{ tel que } \mathrm{Re}(s) > 1\}$.

2. Montrer que les fonctions ζ et G sont holomorphes dans le demi-plan $\{s \in \mathbb{C}$ tel que $\mathrm{Re}(s) > 1\}$.
3. Montrer que $\zeta(s) = \frac{2^{s-1}}{1-2^{s-1}} G(s)$ pour $s \in \mathbb{C}$ tel que $\mathrm{Re}(s) > 1$.

Exemple 5.7 (2022,Partie I)**.** Soit $\boldsymbol{a} = (a_n)_{n\in\mathbb{N}^*}$ une suite de nombres complexes. On suppose qu'il existe un réel σ_0 tel que la série de terme général $(|a_n|n^{-\sigma_0})_{n\in\mathbb{N}^*}$ est convergente dans $\mathbb{R}_+$.

1. Montrer que pour tout s dans $\boldsymbol{H}_{\sigma_0} := \{z \in \mathbb{C},\ \mathrm{Re}(z) > \sigma_0\}$, la série de terme général $(a_n n^{-s})_{n\in\mathbb{N}^*}$ est convergente.

On note $\varphi_{\boldsymbol{a}}(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ sa somme qui est appelée *somme de Dirichlet associée à* $\boldsymbol{a}$.

2. Montrer que $\varphi_{\boldsymbol{a}}$ est holomorphe sur l'ouvert $\boldsymbol{H}_{\sigma_0}$.
3. Montrer que $\lim_{\sigma\to+\infty} \varphi_{\boldsymbol{a}}(\sigma) = a_1$.
   Soit un entier $n_0 \geq 2$. On suppose que $a_1 = \ldots = a_{n_0-1} = 0$. Montrer que $\lim_{\sigma\to+\infty} n_0^{\sigma}\varphi_{\boldsymbol{a}}(\sigma) = a_{n_0}$.
4. Soit $\boldsymbol{b} = (b_n)_{n\in\mathbb{N}^*}$ telle que la série de terme général $(|b_n|n^{-\sigma_0})_{n\in\mathbb{N}^*}$ est convergente dans $\mathbb{R}_+$. On suppose que les séries de Dirichlet $\varphi_{\boldsymbol{a}}$ et $\varphi_{\boldsymbol{b}}$ coïncident sur $\boldsymbol{H}_{\sigma_0}$. Montrer que $a_n = b_n$ pour tout $n \in \mathbb{N}^*$.

Exemple 5.8 (2013, Partie II)**.** 1. Soit $(z_n)_{n\in\mathbb{N}}$ une suite de nombres complexes.

(a) Montrer que pour tout $N \in \mathbb{N}$, on a

$$\left| \left[\prod_{j=0}^{N} (1+z_j) \right] - 1 \right| \leq \left[\prod_{j=0}^{N} (1+|z_j|) \right] - 1.$$

(b) Montrer que pour tout $N \in \mathbb{N}$, on a

$$\prod_{j=0}^{N} (1+|z_j|) \leq \exp\left(\sum_{j=0}^{N} |z_j| \right).$$

2. Soit $(g_j)_{j\in\mathbb{N}}$ une suite de fonctions holomorphes sur un ouvert $U \subset \mathbb{C}$ telle que la série de terme général g_j converge normalement sur tout compact de U. Montrer que la suite de fonctions $(G_N)_{N\in\mathbb{N}}$ définie, pour $z \in U$, par

$$G_N(z) = \prod_{j=0}^{N} (1+g_j(z))$$

converge uniformément sur tout compact de U vers une fonction qui est holomorphe sur U.

Chapitre III

Topologie

Dans les deux chapitres précédents, nous avons rencontré à plusieurs reprises la notion de convergence : pour des réels, pour des complexes, mais aussi pour des suites de fonctions, pour lesquelles on a même défini deux formes de convergence (la convergence simple et la convergence uniforme).

Dans ce chapitre, nous allons généraliser le concept de convergence, ainsi que toutes les notions qui s'en déduisent, comme le caractère fermé d'un ensemble ou la continuité d'une application. Ce chapitre est absolument central dans le programme ; au moins 7 leçons (**201-203-204-205-207-208-213**) se basent quasi-exclusivement sur les notions abordées dans ce chapitre et dans son extension, à savoir le Chapitre IX. Mais bien sûr, les outils de ce chapitre interviennent dans absolument tous les aspects du programme. On a d'ailleurs déjà évoqué quelques résultats (avec ou sans démonstration selon les cas) dans des cas particuliers dans les deux chapitres précédents, et nous allons revenir sur ceux-ci avec plus de généralité ici.

Précisons que la définition d'une topologie (au sens de la topologie générale) n'est pas au programme. Il n'est pas pour autant farfelu dans le cadre de l'agrégation de citer des espaces ou des convergences qui ne rentrent pas dans le cadre des espaces métriques ; mais on se contentera pour ces cas-là de tout faire à la main et on se passera d'un cadre abstrait. On évoquera quand même brièvement la question à la remarque 1.23 pour le lecteur curieux et ambitieux, d'autant qu'il peut paraître mystérieux de parler de topologie des espaces métriques, sans savoir ce qu'est une topologie. Le lecteur doit cependant avoir conscience que la plupart des notions qui suivent se traitent de façon différente dans le cadre des espaces topologiques, et que de nombreuses complications arrivent dans ce contexte ; la plupart des énoncés deviennent faux si on ne rajoute pas les hypothèses ad hoc. Il ne s'agit donc pas d'une simple extension de se confronter à ce cadre plus général. On conseille donc le lecteur à beaucoup de prudence s'il décide de s'investir dans cette direction ; notamment, le jury pourra naturellement demander à un candidat faisant ce choix d'agrémenter ces considérations par des exemples pertinents et intéressants qui justifient un tel investissement. Si tel n'est pas le cas, l'effet sera clairement négatif au regard du jury [1].

On abordera au paragraphe 3 quelques éléments d'analyse fonctionnelle, en revenant notamment à la notion de convergence uniforme et en étudiant plus précisément l'espace des fonctions continues. Nous aurons l'occasion de revenir plus en détails sur l'analyse fonctionnelle au Chapitre IX.

Je ne précise pas de référence globale : les notions abordées ici sont assez élémentaires et traitées dans de nombreux livres.

1 Espaces métriques

Dans le cadre du programme, nous allons étudier les espaces métriques. La plupart des exemples pertinents au niveau de l'agrégation rentreront dans un cas particulier de ce cadre, à savoir les espaces

1. Au cas où le message n'était pas clair, sauf situation exceptionnelle, je déconseille personnellement de faire cet investissement, ou en tout cas de l'inclure dans vos leçons ; un très bon candidat pourra tirer un bénéfice certain à se renseigner sur ce cadre plus général, mais il se rendra rapidement compte que les complications sont multiples, et qu'il est difficile de valoriser cela de manière cohérente avec le programme. Il y a bien d'autres choses pertinentes et intéressantes à développer et qui sont à mon avis plus logiques et "rentables" dans le cadre du concours.

vectoriels normés, mais nous repoussons au paragrape 2 les détails concernant ce cas particulier, afin que le lecteur puisse distinguer ce qui est général à tout les espaces métriques, et ce qui est propre aux espaces vectoriels normés. Ainsi, pour le moment, nous ne supposons aucune structure algébrique sur l'ensemble de travail, qu'on notera souvent X.

1.1 Définitions et notions topologiques

1.1.1 Distance : définition et exemples

Définition 1.1 (Métrique/Distance)**.** Soit X un ensemble. On appelle distance sur X une application $d : X \times X \to \mathbb{R}_+$ à valeurs dans $\mathbb{R}_+$ et telle que

1. pour tout $(x, y) \in X^2$, $d(x, y) = 0 \;\Leftrightarrow\; x = y$, (séparation)
2. pour tout $(x, y) \in X^2$, $d(x, y) = d(y, x)$, (symétrie)
3. pour tout $(x, y, z) \in X^3$, $d(x, z) \leq d(x, y) + d(y, z)$ (inégalité triangulaire).

On appelle espace métrique la donnée d'un couple (X, d) où d est une distance sur X.

Remarque 1.2. Comme son nom l'indique, cette notion permet de mesurer par un réel positif ou nul la distance entre deux éléments quelconques d'un ensemble. Assez naturellement on a imposé 4 propriétés (n'oublions pas qu'on a demandé que d soit à valeurs dans $\mathbb{R}_+$ [2]) qui correspondent à l'intuition naturelle d'une notion de distance.

Exercice 1.3 (Partie gauche de l'inégalité triangulaire)**.** Montrer que si (X, d) est un espace métrique, alors

$$\forall (x, y, z) \in X^3, \;\; |d(x, y) - d(y, z)| \leq d(x, z).$$

Exemple 1.4. Si $X = \mathbb{R}$, la fonction $\forall (x, y) \in \mathbb{R}^2, d(x, y) = |y - x|$ est une distance sur $\mathbb{R}$. Le même exemple vaut sur $\mathbb{C}$ avec le module des nombres complexes à la place de la valeur absolue. Sauf mention contraire, $\mathbb{R}$ ou $\mathbb{C}$ seront toujours munis de cette métrique, qu'on appellera distance usuelle sur $\mathbb{R}$ ou $\mathbb{C}$.

Exercice 1.5. Soit X un ensemble. On pose

$$\forall (x, y) \in X^2, \quad d(x, y) = \begin{cases} 1 & \text{si } x \neq y \\ 0 & \text{sinon.} \end{cases}$$

Montrer que d est une métrique sur X. On l'appellera distance discrète sur X.

Exemple 1.6. Notons $X := \mathbb{R} \cup \{-\infty, +\infty\}$. On peut poser

$$\forall (x, y) \in X^2, \;\; d(x, y) = |\arctan(x) - \arctan(y)|$$

où on a convenu que $\arctan(\frac{\pi}{2}) = +\infty$ et $\arctan(-\frac{\pi}{2}) = -\infty$; d définit bien une métrique sur $\mathbb{R} \cup \{\pm\infty\}$.

Exercice 1.7 (Distance SNCF)**.** On pose

$$\forall (z, z') \in \mathbb{C}^2, \quad d(z, z') = \begin{cases} |z - z'| & \text{si } z \text{ et } z' \text{ sont colinéaires,} \\ |z| + |z'| & \text{sinon.} \end{cases}$$

Montrer que d est une distance sur $\mathbb{C}$, et interprétez son nom.

Exemple 1.8 (Distance induite)**.** Si (X, d) est un espace métrique, et si $Y \subset X$, on peut considérer la restriction de d à $Y \times Y$, qu'on note d_Y : il s'agit d'une distance sur Y, qu'on appelle distance induite, et alors (Y, d_Y) est un espace métrique. Attention, par la suite, quand il y aura plusieurs "espaces ambiants" possibles, il faudra toujours préciser auquel on fait référence. La métrique est peut-être inchangée (ou en tout cas on ne change que son espace de départ), mais le fait que l'ensemble considéré change peut avoir une incidence sur les notions que nous allons manipuler.

2. En fait, il se trouve que le fait que d est à valeurs dans $\mathbb{R}_+$ est une conséquence des propriétés 1-2-3 puisque pour $(x, y) \in X^2$, on a $0 = d(x, x) \leq d(x, y) + d(y, x) = 2d(x, y)$. Mais bon, je conseille plutôt de ne pas en tenir compte, ça ne coûte rien en général, de vérifier que d est à valeurs positives. Attention par contre, il est interdit que d prenne la valeur $+\infty$.

Exercice 1.9 (Métrique produit)**.** Soit $N \in \mathbb{N}^*$ et $(X_1, d_1), \dots, (X_N, d_N)$ des espaces métriques. On pose $X = \prod_{i=1}^N X_i$ et

$$\forall x = (x_1, \dots, x_N) \in X, \ \ \forall y = (y_1, \dots, y_N) \in X, \qquad d_X^1(x,y) = \sum_{i=1}^N d_i(x_i, y_i),$$
$$d_X^\infty(x,y) = \max_{i \in [\![1,N]\!]} d_i(x_i, y_i).$$

1. Montrer que d_X^1 et d_X^∞ sont des métriques sur X.
2. Montrer que ces deux métriques sont Lipschitz-équivalentes (voir exercice 1.65).

L'une ou l'autre de ces métriques sera appelée métrique produit, et sauf mention contraire, un espace de la forme X sera toujours muni d'une telle métrique (ou d'une métrique Lipschitz-équivalente).

Exercice 1.10. On pose $X = [0,1]^{\mathbb{N}}$. Pour $x = (x_n)_{n \in \mathbb{N}}$ et $y = (y_n)_{n \in \mathbb{N}}$ dans X, on pose

$$d(x,y) := \sum_{n \in \mathbb{N}} \frac{|x_n - y_n|}{2^n}.$$

Montrer que d est une métrique sur X.

Définition 1.11. Soit (X_1, d_1) et (X_2, d_2) deux espaces métriques. On dit que $f : X_1 \to X_2$ est une isométrie si

$$\forall (x,y) \in X_1, \ \ d_2(f(x), f(y)) = d_1(x,y).$$

On dit que (X_1, d_1) et (X_2, d_2) sont isométriques s'il existe une isométrie bijective entre eux.

Remarque 1.12. On laisse le lecteur se convaincre qu'une isométrie est toujours injective. C'est donc la surjectivité qui peut manquer à une isométrie pour être bijective.

Exercice 1.13. Soit X un ensemble, (Y, d) un espace métrique et $f : X \to Y$ une application injective. Montrer que

$$\forall (x,y) \in X^2, \ \ \widetilde{d}(x,y) := d(f(x), f(y))$$

est une métrique sur X, et qu'en munissant X de cette métrique, f devient une isométrie.

1.1.2 Boules, ouverts, fermés

Définition 1.14 (Boules)**.** Soit (X, d) un espace métrique.

— On pose pour $x \in X$ et $r \geq 0$,

$$B_X(x,r) = \left\{ y \in X, \ \ d(y,x) < r \right\}, \qquad B_X^f(x,r) = \left\{ y \in X, \ \ d(x,y) \leq r \right\}$$

respectivement la boule ouverte et la boule fermée de centre x et de rayon r. La référence à l'espace X en indice ne sera indiquée que s'il y a ambiguïté.

— On appelle aussi sphère de centre x et de rayon r l'ensemble

$$S_X(x,r) = \left\{ y \in X, \ \ d(y,x) = r \right\}.$$

— On dit que $V \subset X$ est un voisinage de x (dans X) s'il existe $r > 0$ tel que $B_X(x,r) \subset V$.

Exercice 1.15. Montrer que $V \subset X$ est un voisinage de x si et seulement s'il existe $r > 0$ tel que $B^f(x,r) \subset V$.

Remarque 1.16 (*Ensemble borné)**.** Dans un espace métrique (X, d), on est assez tenté de dire qu'un ensemble $A \subset X$ est borné s'il existe $M > 0$ tel que

$$\forall (x,y) \in A^2, \ d(x,y) \leq M. \tag{1.1}$$

On peut montrer que ceci est équivalent au fait qu'il existe $x_0 \in X$ et $r_0 > 0$ tels que $A \subset B^f(x_0, r_0)$.

Néanmoins, on se rend compte assez vite que cette définition est assez peu naturelle, car elle dépend fortement de la métrique choisie, et non de la convergence sous-jacente. Autrement dit, on pourrait définir une métrique topologiquement équivalente[3] à d (par exemple $\min\{1,d\}$) pour laquelle tous les ensembles seraient bornés pour cette définition. On dit que la définition (1.1) n'est pas topologique.

Même si la littérature utilise fréquemment la définition précédente, on préfèrera dire "de diamètre fini" pour une partie qui satisfait (1.1) : en effet si A est non vide, on peut définir le diamètre de A comme

$$\operatorname{diam}(A) = \sup_{(x,y)\in A^2} d(x,y).$$

Nous reviendrons sur cette subtilité dans le cadre des espaces vectoriels. Nous donnerons une autre définition d'un ensemble borné, qui diffère de (1.1), voir l'exercice 4.23. On verra aussi que quand l'espace vectoriel est normé, il n'y a pas d'ambiguïté car les deux définitions sont équivalentes si on ne considère que des métriques issues de normes.

Définition 1.17 (Ensembles ouverts, fermés)**.** Soit $O \subset X$. On dit que O est un ouvert de X si O est un voisinage de tous ses points, c'est-à-dire

$$\forall x \in O, \quad \exists r > 0,\ B(x,r) \subset O. \tag{1.2}$$

Soit $F \subset X$. On dit que F est un fermé de X si $F^c = X \setminus F$ est un ouvert de X.

Exercice 1.18. Soit (X,d) un espace métrique. Montrer qu'une boule ouverte de X est un ouvert de X, et qu'une boule fermée de X est un fermé de X !

Remarque 1.19. On dira autant que possible "ouvert/fermé de X" et pas seulement "ouvert/fermé". En effet, notamment quand on est dans le cadre d'une topologie induite avec $Y \subset X$, le caractère ouvert d'un ensemble $O \subset Y$ dépend de l'espace ambiant dans lequel on se place. En effet, la définition dépend des boules qui elles-mêmes changent si on les voit dans Y ou dans X. Prenons un exemple dans $X = \mathbb{R}$, avec $Y =]0,2]$. L'ensemble $O =]1,2]$ est un sous-ensemble de Y et donc aussi de X :

— O n'est pas un ouvert de $\mathbb{R}$, car 2 est dans O mais n'admet pas de voisinage inclus dans O : en effet, pour tout $r > 0$,

$$2 + \frac{r}{2} \in]2-r, 2+r[\setminus O, \qquad \text{où }]2-r,2+r[= B_{\mathbb{R}}(2,r)\,.$$

— Par contre, O est un ouvert de Y : concentrons-nous à vérifier que 2 admet bien un voisinage qui reste dans Y (les autres points étant clairs) :

$$B_Y\left(2,\frac{1}{2}\right) = \left]2-\frac{1}{2}, 2\right] \subset O.$$

Voir aussi l'exercice 1.27.

Remarque 1.20. Attention, on n'a pas dit qu'un ensemble était fermé s'il n'était pas ouvert ! ! ! Il existe en général des ensembles qui ne sont ni ouverts, ni fermés. Il arrive aussi qu'un ensemble soit à la fois ouvert et fermé !

Exercice 1.21. Considérons $(\mathbb{R}, |\cdot|)$.

1. Reprendre la liste des intervalles de $\mathbb{R}$ (voir paragraphe 1.2 au Chapitre I) et identifiez les intervalles qui sont ouverts, ceux qui sont fermés, ceux qui ne sont ni l'un ni l'autre, ceux qui sont les deux.
2. [*Ouverts de $\mathbb{R}$] Montrer qu'un ouvert de $\mathbb{R}$ peut s'écrire comme une union dénombrable d'intervalles ouverts deux à deux disjoints.

Proposition 1.22 (Structure de l'ensemble des ouverts[4])**.** *Soit (X,d) un espace métrique. Alors*

3. Voir l'exercice 1.65.

4. Notamment, on comparera en temps voulu avec les propriétés des ensembles mesurables dans le cadre des tribus au Chapitre IV.

1. *Les ensembles $\emptyset$ et X sont ouverts dans X.*
2. *Une union quelconque*[5] *d'ouverts de X est un ouvert de X :*

$$\forall I \text{ ensemble}, \quad \Big[\forall i \in I, \ \ O_i \text{ ouvert de } X\Big] \implies \bigcup_{i\in I} O_i \text{ est un ouvert de } X.$$

3. *Une intersection finie d'ouverts de X est un ouvert de X :*

$$\forall N \in \mathbb{N}, \Big[\forall i \in [\![1,N]\!], \ \ O_i \text{ ouvert de } X\Big] \implies \bigcap_{i\in[\![1,N]\!]} O_i \text{ est un ouvert de } X.$$

Remarque 1.23. Étant donné (X,d) un espace métrique, on peut appeler topologie sur X associée à d l'ensemble des ouverts de X.

En fait, le cadre de la topologie générale est le suivant : si X est un ensemble, on appelle topologie sur X une collection $\mathscr{T}$ de sous-ensembles de X qui satisfait :

1. $\emptyset \in \mathscr{T}$, $X \in \mathscr{T}$,
2. $\mathscr{T}$ est stable par union quelconque,
3. $\mathscr{T}$ est stable par intersection finie.

On dit que $(X, \mathscr{T})$ est un espace topologique, et les ouverts de X sont alors simplement les éléments de $\mathscr{T}$.

On vient donc de montrer que si X est muni d'une métrique, alors la collection des ouverts définis par (1.2) est effectivement une topologie.

Comme on l'a évoqué en introduction de ce chapitre, ce cadre n'est pas au programme, et de nombreux notions ou énoncés ne fonctionnent plus dans ce cadre général.

Démonstration. 1. Le caractère ouvert de $\emptyset$ consiste à montrer une propriété vraie pour tout $x \in \emptyset$. C'est donc bien le cas. Le caractère ouvert de X vient du fait que si $x \in X$, alors $B(x, 42) \subset X$.

2. Soit $x \in \cup_{i\in I} O_i$. Alors il existe i_0 tel que $x \in O_{i_0}$. Ainsi comme O_{i_0} est ouvert, il existe $r > 0$ tel que $B(x, r_0) \subset O_{i_0} \subset \cup_{i\in I} O_i$, d'où le résultat.
3. Soit $x \in \cap_{i\in[\![1,N]\!]} O_i$. Alors pour tout $i \in [\![1,N]\!]$, $x \in O_i$ qui est ouvert, donc il existe $r_i > 0$ tel que $B(x, r_i) \subset O_i$. On pose alors $r_0 = \min\{r_i, i \in [\![1,N]\!]\}$. Alors

$$\forall i \in [\![1,N]\!], \ \ B(x,r_0) \subset B(x,r_i) \subset O_i, \qquad \text{d'où} \quad B(x,r_0) \subset \bigcap_{i\in[\![1,N]\!]} O_i.$$

□

Exercice 1.24. Essayez de mesurer dans la preuve précédente comment se distinguent les preuves des cas 2 et 3, et pourquoi on a eu besoin ou non du caractère fini de l'ensemble des indices. C'est une habitude à prendre de comprendre où les hypothèses ont été utilisées.

Afin de constater effectivement la nécessité de supposer l'ensemble des indices fini dans le cas 3, je vous invite à exhiber un contre-exemple dans le cas d'une intersection infinie.

Exercice 1.25 (Structure de l'ensemble des fermés de X)**.** Soit (X,d) un espace métrique. Montrer que

1. Les ensembles $\emptyset$ et X sont fermés dans X.
2. Une intersection quelconque de fermés de X est un fermé de X.
3. Une union finie de fermés de X est un fermé de X.

Exercice 1.26 (Espace discret)**.** On considère $\mathbb{N}$ muni de la distance $\forall (x,y) \in \mathbb{N}^2, d(x,y) = |x-y|$. Décrire les sous-ensembles de $\mathbb{N}$ qui sont ouverts. Décrire ceux qui sont fermés.

5. Cet adjectif signifie ici que le "nombre" d'ouverts considéré est quelconque : il peut être fini, dénombrable, ou indénombrable.

Exercice 1.27 (Ouverts pour la métrique induite)**.** Soit (X, d) un espace métrique, et $Y \subset X$ que l'on munit de la métrique induite d_Y (c'est-à-dire la restriction de d à $Y \times Y$).

1. Montrer que $A \subset Y$ est un ouvert de (Y, d_Y) si et seulement s'il existe O un ouvert de X tel que $A = O \cap Y$.
2. Montrer que

$$Y \text{ est un ouvert de } X \quad \Longrightarrow \quad \text{les ouverts de } Y \text{ sont tous des ouverts de X.}$$

1.1.3 Intérieur, adhérence

Définition 1.28. Soit (X, d) un espace métrique, $A \subset X$ et $x \in X$.

— On dit que x est intérieur à A s'il existe $r > 0$ tel que $B(x, r) \subset A$.

— On dit que x est adhérent à A si pour tout $r > 0$, $B(x, r) \cap A \neq \emptyset$.

— On dit que x est un point frontière de A si pour tout $r > 0$, $B(x, r) \cap A \neq \emptyset$ et $B(x, r) \setminus A \neq \emptyset$, c'est-à-dire que x est adhérent à A et à $A^c = X \setminus A$.

On pose

$$\mathring{A} := \Big\{ x \in X, \quad x \text{ est un point intérieur à } A \Big\},$$

$$\overline{A} := \Big\{ x \in X, \quad x \text{ est un point adhérent à } A \Big\},$$

$$\partial A := \Big\{ x \in X, \quad x \text{ est un point frontière de } A \Big\}.$$

qu'on appelle respectivement intérieur, adhérence et frontière de l'ensemble A.

Enfin, on dit que A est dense dans X si $\overline{A} = X$ [6].

Exercice 1.29. On se place dans $(\mathbb{R}, |\cdot|)$.

1. Montrer qu'un sous-ensemble A de $\mathbb{R}$ est dense dans $\mathbb{R}$ si et seulement si

$$\forall (x, y) \in \mathbb{R}^2, \quad]x, y[\cap A \neq \emptyset.$$

2. Calculer l'intérieur des intervalles de la forme $]x, y]$, $[x, y[$ ou $[x, y]$ avec $x \leq y$, et l'adhérence des intervalles de la forme $]x, y[$, $[x, y[$ ou $]x, y]$.
3. Calculer l'intérieur de $\mathbb{Q}$ et de $\mathbb{R} \setminus \mathbb{Q}$.

Exercice 1.30. Dans le cadre de la définition précédente, montrez

$$\mathring{A} \subset A \subset \overline{A},$$

et

$$\partial A = \overline{A} \setminus \mathring{A}.$$

Proposition 1.31 (Caractérisation de l'intérieur et de l'adhérence)**.** *Soit (X, d) un espace métrique, et $A \subset X$. Alors*

1. *$\mathring{A}$ est un ouvert, et c'est le plus grand ouvert [7] qui soit inclus dans A, c'est-à-dire*

$$\forall\, O \subset A, \quad O \ \text{ ouvert } \Rightarrow \ O \subset \mathring{A}.$$

6. Attention, il existe plusieurs conventions sur la notion de densité : on rencontre parfois la définition suivante : étant donnés A et B deux sous-ensembles de X (qui est muni d'une métrique), on dit que A est dense dans B si $\overline{A} \supset B$ (ce qui voudra dire, quand on aura vu les caractérisations séquentielles, que tout élément de B est limite d'éléments de A). En particulier, cette définition ne demande pas que $A \subset B$. On évitera si possible cette terminologie. À titre d'exemple, cette convention permet par exemple de dire que l'ensemble des fonctions en escaliers est dense dans l'ensemble des fonctions continues pour la norme uniforme (dans cet exemple, $A = Esc([a, b])$, $B = C^0([0, 1])$ et $(X, d) = (C^0_m([a, b]), \|\cdot\|_\infty)$ (voir plus loin pour la métrique associée à une norme)) ; il faut quand même s'assurer qu'il y a bien un espace qui contient à la fois A et B, et qu'on appellera espace ambiant.

7. "Plus grand" pour l'inclusion.

2. *$\overline{A}$ est un fermé, et c'est le plus petit fermé qui contient A, c'est-à-dire*

$$\forall\, A \subset F \subset X, \quad F \text{ fermé } \Rightarrow \quad \overline{A} \subset F.$$

Exercice 1.32. En déduire que $\Big[$ A est fermé si et seulement si $A = \overline{A}$ $\Big]$, et $\Big[$ A est ouvert si et seulement si $\mathring{A} = A$ $\Big]$.

Démonstration. Soit $x \in \mathring{A}$. Alors il existe $r > 0$ tel que $B(x,r) \subset A$. Nous allons montrer que $B(x, \frac{r}{2}) \subset \mathring{A}$. Soit $y \in B(x, \frac{r}{2})$: alors

$$\forall z \in B\left(y, \frac{r}{2}\right), \quad d(z,x) \le d(z,y) + d(y,x) < \frac{r}{2} + \frac{r}{2} = r, \quad \text{d'où} \quad B\left(y, \frac{r}{2}\right) \subset B(x,r) \subset A \tag{1.3}$$

donc par définition, $y \in \mathring{A}$. Ainsi $\mathring{A}$ est un voisinage de tous ses points, et donc $\mathring{A}$ est ouvert, et on a vu à l'exercice 1.30 que $\mathring{A}$ est inclus dans A.

Soit maintenant O un ouvert inclus dans A. Si $x \in O$, il existe $r > 0$ tel que $B(x,r) \subset O \subset A$, donc $x \in \mathring{A}$, et on a montré que $O \subset \mathring{A}$, ce qui montre que $\mathring{A}$ est bien le plus grand ouvert inclus dans A.

Montrons que $X \setminus \overline{A}$ est ouvert. Soit $x \in X \setminus \overline{A}$. Comme x n'est pas dans $\overline{A}$, il existe $r_0 > 0$ tel que $B(x, r_0) \cap A = \emptyset$, c'est-à-dire $B(x, r_0) \subset X \setminus A$. Montrons que $B(x, \frac{r_0}{2}) \subset X \setminus \overline{A}$: avec le même calcul qu'en (1.3),

$$\forall y \in B\left(x, \frac{r_0}{2}\right), \quad B\left(y, \frac{r_0}{2}\right) \subset B(x, r_0) \subset X \setminus A, \quad \text{i.e. } y \notin \overline{A}$$

donc $B\left(x, \frac{r_0}{2}\right) \subset X \setminus \overline{A}$ et donc $X \setminus \overline{A}$ est ouvert. Donc $\overline{A}$ est bien un fermé qui contient A.

Soit maintenant F un fermé contenant A. Soit $x \in \overline{A}$. Alors pour $r > 0$ on a $B(x,r) \cap A \neq \emptyset$, mais comme $A \subset F$, on a aussi $B(x,r) \cap F \neq \emptyset$. Si par l'absurde on avait $x \notin F$ alors comme $X \setminus F$ est ouvert, il existerait $r_0 > 0$ tel que $B(x, r_0) \subset X \setminus F$, c'est-à-dire $B(x, r_0) \cap F = \emptyset$, ce qui constituerait une contradiction. Et donc $x \in F$ et $\overline{A} \subset F$, ce qu'il fallait démontrer. □

Exercice 1.33. Soit (X,d) un espace métrique, $A \subset X$ et $B \subset X$.

1. Montrer que
$$\overline{A} = \left(\mathring{\widehat{A^c}}\right)^c, \qquad \text{et} \qquad \mathring{A} = \left(\overline{A^c}\right)^c.$$

2. Montrer [8]
$$\overset{\circ}{\widehat{A \cap B}} = \mathring{A} \cap \mathring{B}, \quad \overline{A \cup B} = \overline{A} \cup \overline{B}.$$

3. Montrer
$$\mathring{A} \cup \mathring{B} \subset \overset{\circ}{\widehat{A \cup B}} \quad \overline{A \cap B} \subset \overline{A} \cap \overline{B},$$
et exhiber des exemples où il n'y a pas égalité.

4. [*] Étudiez les formules qui restent valable quand on considère une union ou une intersection quelconque d'ensembles (au lieu de deux comme ci-dessus).

Exercice 1.34 (Boule fermée et adhérence)**.** Soit (X,d) un espace métrique, $x \in X$ et $r > 0$. Montrer

$$\overline{B(x,r)} \subset B^f(x,r).$$

Trouver un exemple pour lequel il n'y a pas égalité (c'est-à-dire qu'il faut trouver X, d, x et r pour lesquels l'égalité n'a pas lieu).

Exercice 1.35. Étant donné (X,d) un espace métrique, $A \subset X$ et $x \in X$. On dit que

8. Une fois n'est pas coutume, je déconseille d'essayer de retenir par cœur de telles formules, on a tôt fait de confondre celles qui "marchent" de celles qui ne "marchent pas", ou alors d'oublier le sens de l'inclusion qui reste tout le temps vraie. Par contre, être capable de retrouver rapidement les preuves des inclusions qui sont vraies ou de trouver des contre-exemples quand on se pose la question de la véracité d'une formule "topologique", est un objectif naturel à avoir pour tester notre compréhension de ces concepts.

— x est un point d'accumulation de A si pour tout $\varepsilon > 0$, $A \cap B(x, \varepsilon) \setminus \{x\} \neq \emptyset$.

— x est un point isolé de A s'il existe $\varepsilon > 0$ tel que $A \cap B(x, \varepsilon) = \{x\}$.

Montrer que l'adhérence de A est égale à l'union des points isolés et des points d'accumulation de A.

Définition 1.36. Un espace métrique (X, d) est dit discret si tout point de X est isolé.

Exercice 1.37 (*Espaces discrets)**.**
1. Soit X un ensemble muni de la métrique discrète (voir exercice 1.5). Montrer que X est un espace discret.
2. Soit (X, d) un espace métrique discret. Montrer que toutes les parties de X sont ouvertes et fermées. En déduire que la distance d est topologiquement équivalente à la métrique discrète (voir exercice 1.65 pour une définition du caractère topologiquement équivalent).
3. Montrer que $\mathbb{N}$ muni de la distance induite par $\mathbb{R}$ est un espace discret, et comparez avec l'exercice 1.26.

Définition 1.38. Un espace métrique (X, d) est dit *séparable* s'il contient un sous-ensemble dénombrable dense.

Exemple 1.39. L'espace $(\mathbb{R}, |\cdot|)$ est séparable puisque $\mathbb{Q}$ est dénombrable et dense.

Exercice 1.40 (*)**.** Montrer que si (X, d) est séparable et $Y \subset X$, alors l'espace Y muni de la métrique induite (Y, d_Y) est également séparable.

1.2 Suites et caractérisations séquentielles

1.2.1 Suites convergentes

Avec la notion de métrique, on mesure la distance entre deux éléments par un réel. On est donc en mesure de calquer les définitions de convergence vues au Chapitre I.

Définition 1.41. Soit (X, d) un espace métrique, $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de X, et $\ell \in X$. On dit que la suite $(x_n)_{n\in\mathbb{N}}$ converge vers ℓ pour la métrique d [9] si

$$\Big[\,\forall \varepsilon > 0, \quad \exists n_0 \in \mathbb{N}, \quad \forall n \geq n_0, \quad d(x_n, \ell) \leq \varepsilon\,\Big], \qquad \text{c'est-à-dire} \quad d(x_n, \ell) \xrightarrow[n\to+\infty]{} 0,$$

et on peut noter

$$x_n \xrightarrow[n\to+\infty]{d} \ell.$$

On dira que la suite $(x_n)_{n\in\mathbb{N}}$ est convergente (dans X et pour la métrique d) s'il existe $\ell \in X$ tel que $(x_n)_{n\in\mathbb{N}}$ converge vers ℓ pour la métrique d. On appellera alors ℓ "une"[10] limite de $(x_n)_{n\in\mathbb{N}}$.

Proposition 1.42. *Soit (X, d) un espace métrique, et $(x_n)_{n\in\mathbb{N}}$ une suite convergente. Alors la limite de la suite $(x_n)_{n\in\mathbb{N}}$ est unique.*

Démonstration. Soit ℓ_1 et ℓ_2 deux éléments de X qui sont limites de $(x_n)_{n\in\mathbb{N}}$. Alors

$$\forall n \in \mathbb{N}, \quad 0 \leq d(\ell_1, \ell_2) \leq d(\ell_1, x_n) + d(x_n, \ell_2).$$

Par passage à la limite, on en déduit $d(\ell_1, \ell_2) = 0$, d'où $\ell_1 = \ell_2$. □

Exercice 1.43. On se place dans le cadre de l'exercice 1.9. On munit X de d_X^1 ou d_X^2 au choix. Montrer qu'une suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de X converge dans X si et seulement si les suites coordonnées $(x_{i,n})_{n\in\mathbb{N}}$ convergent dans (X_i, d_i) pour tout $i \in [\![1, N]\!]$, et la limite de $(x_n)_{n\in\mathbb{N}}$ est égale au N-uplet formé des limites des suites $(x_{i,n})_{n\in\mathbb{N}}$ pour $i \in [\![1, N]\!]$.

9. On invite encore une fois le lecteur à préciser le plus souvent possible la métrique pour laquelle il y a convergence, car bien souvent, il y a plusieurs métriques possibles, et la convergence peut avoir lieu pour une métrique et pas pour une autre.

10. Avec la proposition 1.42 on pourra dire "la".

Exercice 1.44. On travaille dans $\mathbb{R}\cup\{\pm\infty\}$ muni de la métrique d donnée à l'exemple 1.6. Montrer que pour $(x_n)_{n\in\mathbb{N}}$ une suite de réels et $\ell\in\mathbb{R}\cup\{\pm\infty\}$,

$$(x_n)_{n\in\mathbb{N}} \text{ tend vers } \ell \text{ au sens du Chapitre I} \quad\Leftrightarrow\quad x_n \xrightarrow[n\to\infty]{d} \ell.$$

Ainsi, on voit qu'une suite $(x_n)_{n\in\mathbb{N}}$ de réels qui "diverge vers $+\infty$" au sens des réels, est une suite qui converge vers $+\infty$ dans l'espace métrique $(\mathbb{R}\cup\{\pm\infty\},d)$, voir aussi la remarque 2.6 du Chapitre I.

1.2.2 Caractérisation séquentielle d'un ensemble fermé

On peut obtenir une caractérisation séquentielle de l'adhérence d'un ensemble (et donc de son caractère fermé) qui est parfois plus commode que le fait de passer au complémentaire ou de revenir à des boules :

Proposition 1.45. *Soit (X,d) un espace métrique, et $A\subset X$. On a*

- *l'élément $x\in X$ est adhérent à A si et seulement s'il existe $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de A qui converge vers x.*
- *en conséquence, A est fermé si et seulement si toute suite d'éléments de A qui converge dans X pour la métrique d, converge dans A, c'est-à-dire*

$$A \text{ est fermé dans } X \iff \Big[\,\forall (x_n)_{n\in\mathbb{N}}\in A^{\mathbb{N}},\ \forall \ell\in X,\ \Big(x_n \xrightarrow[n\to+\infty]{} \ell \Rightarrow \ell\in A\Big)\Big]$$

Remarque 1.46. On a donc plusieurs façons d'aborder les questions de topologie. On peut préférer travailler avec des boules, comme on l'a fait au paragraphe précédent : on dit alors qu'on fait un raisonnement topologique. Quand on préfère manipuler des suites, on parle d'un raisonnement séquentiel. Dans le cadre des espaces métriques, nous avons vu que ces façons d'aborder les notions sont équivalentes. Dans le cadre des espaces topologiques généraux (voir la remarque 1.23), ceci n'est plus nécessairement le cas.

Démonstration. Soit x adhérent à A. Pour $n\in\mathbb{N}$, on a $B(x,\frac{1}{n+1})\cap A\neq\emptyset$, donc il existe $x_n\in A$ tel que $d(x,x_n)<\frac{1}{n+1}$. Ainsi il existe bien une suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de A qui converge vers x.

Réciproquement, s'il existe une suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de A qui converge vers x, alors pour $r>0$, il existe $n_0\in\mathbb{N}$ tel que pour tout $n\geq n_0, d(x_n,x)<r$ donc en particulier $x_{n_0}\in B(x,r)\cap A$ donc x est adhérent à A.

Supposons A fermé. On considère alors une suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de A qui converge vers $\ell\in X$. Ainsi par le point précédent, ℓ est adhérent à A, mais comme A est fermé, $\ell\in A$ (voir l'exercice 1.32).

Réciproquement, si on suppose que pour toute suite d'éléments de A qui converge dans X, la limite de cette suite est dans A, alors par le premier point, tous les points adhérents à A sont dans A, ce qui signifie que $\overline{A}=A$ et donc A est fermé. □

Exercice 1.47 (Distance à une partie)**.** Soit (X,d) un espace métrique, et $A\subset X$ non vide. Étant donné $x\in X$, on définit la distance entre x et A par [11]

$$d(x,A) := \inf\big\{d(x,y),\ y\in A\big\}.$$

1. Montrer que $x\in\overline{A}$ si et seulement si $d(x,A)=0$.
2. Montrer
$$\forall (x,y)\in X^2,\ \ |d(x,A)-d(y,A)|\leq d(x,y).$$
On dira que l'application $x\mapsto d(x,A)$ est 1-lipschitzienne.
3. (a) Montrer que pour tout $n\in\mathbb{N}^*$, l'ensemble $A_n=\{x\in X, d(x,A)<\frac{1}{n}\}$ est ouvert.
   (b) Déterminer $\cap_{n\in\mathbb{N}}A_n$.

11. Attention, on réutilise la notation d, mais il ne faut pas confondre la métrique d, qui prend pour paramètre deux éléments de X, et la distance d'un point à une partie ; cet abus n'est pas trop dangereux car un ensemble (ici A) n'est pas un élément de X. Soyez quand même vigilant !

(c) En déduire que si A est fermé dans X, alors il peut s'écrire comme une intersection dénombrable d'ouverts.

4. Montrer également que tout ouvert de X peut s'écrire comme union dénombrable de fermés.
5. [Lemme d'Urysohn dans un espace métrique] Soit F un fermé de X et O un ouvert de X tels que $F \subset O$. Montrer qu'il existe $f : X \to [0,1]$ continue telle que $f = 1$ sur F et $f = 0$ sur O^c.

On peut aussi définir, si A et B sont deux parties non vides de X,

$$d(A,B) := \inf \left\{ d(x,y),\ x \in A,\ y \in B \right\}$$

la distance entre A et B[12].

Exercice 1.48 (*Distance de Hausdorff)**.** Soit (X,d) un espace métrique. On pose $\mathcal{P}_f(X)$ l'ensemble des parties fermées, non vides et de diamètre fini de X. On pose pour $(A,B) \in \mathcal{P}_f(X)$,

$$d_H(A,B) = \max \left\{ \sup_{x \in A} d(x,B), \sup_{x \in B} d(x,A) \right\}.$$

1. Pour $A \subset X$ et $\varepsilon > 0$, on pose $V_r(A) = \{x \in X, d(x,A) < r\}$. Montrer
$$d_H(A,B) = \inf \{\varepsilon > 0,\ A \subset V_\varepsilon(B) \ \text{ et } \ B \subset V_\varepsilon(A)\ \}.$$
2. Montrer que d_H est une métrique sur l'ensemble $\mathcal{P}_f(X)$. On l'appelle distance de Hausdorff.
3. On considère $X = [0,1]$ muni de sa métrique usuelle. On pose pour $n \in \mathbb{N}$,
$$A_n = \left\{ \frac{k}{n+1}, k \in [\![0, n+1]\!] \right\}.$$
Montrer que (A_n) converge vers X pour la distance de Hausdorff.

1.2.3 Valeurs d'adhérence et limites de suites extraites

Définition 1.49. Soit (X,d) un espace métrique, et $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de X. On dit que $x \in X$ est une valeur d'adhérence de $(x_n)_{n\in\mathbb{N}}$ si pour tout $\varepsilon > 0$, l'ensemble des $n \in \mathbb{N}$ tels que $x_n \in B(x,\varepsilon)$ est infini.

On notera $\mathrm{Adh}((x_n)_{n\in\mathbb{N}})$ l'ensemble des valeurs d'adhérence de la suite $(x_n)_{n\in\mathbb{N}}$.

Remarque 1.50. Autrement dit, une valeur d'adhérence de la suite $(x_n)_{n\in\mathbb{N}}$ est un point de X près duquel s'accumulent une infinité de termes de la suite $(x_n)_{n\in\mathbb{N}}$.

Nous allons donner deux expressions de l'ensemble des valeurs d'adhérences : la première est une manipulation simple des définitions, la seconde est un peu plus élaborée :

Lemme 1.51. *Soit (X,d) un espace métrique, et $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de X. Alors*

$$\mathrm{Adh}((x_n)_{n\in\mathbb{N}}) = \bigcap_{n\in\mathbb{N}} \overline{\{x_m, m \geq n\}}.$$

Démonstration. Soit $x \in \mathrm{Adh}((x_n)_{n\in\mathbb{N}})$. Alors par définition, si $\varepsilon > 0$, l'ensemble $\left\{k \in \mathbb{N}, u_k \in B(x,\varepsilon)\right\}$ est infini, donc si $n \in \mathbb{N}$, il existe $m \geq n$ tel que $u_m \in B(x,\varepsilon)$, et donc $x \in \overline{\{u_m, m \geq n\}}$, et ce pour tout $n \in \mathbb{N}$.

Réciproquement, si $x \in \bigcap_{n\in\mathbb{N}} \overline{\{x_m, m \geq n\}}$ et $\varepsilon > 0$, alors pour tout $n \in \mathbb{N}$, $x \in \overline{\{x_m, m \geq n\}}$ donc $B(x,\varepsilon) \cap \{x_m, m \geq n\}$ est non vide. Ainsi pour tout $n \in \mathbb{N}$, il existe $m \geq n$ tel que $x_m \in B(x,\varepsilon)$, ce qui montre bien que l'ensemble des indices $k \in \mathbb{N}$ tels que $x_k \in B(x,\varepsilon)$ est infini[13], ce qui achève la preuve. □

12. En complément de la note 11, faites attention que la fonction ainsi définie d qui prend pour paramètre deux sous-ensembles de X, n'est pas elle-même une distance sur $\mathcal{P}(X)$ l'ensembles des parties de X. Voir l'exercice 1.48 pour un exemple de distance entre les parties (fermées) d'un espace métrique.

13. En fait on a utilisé dans cette preuve qu'un ensemble d'entiers naturels est infini si et seulement s'il n'est pas majoré.

Lemme 1.52. *Soit (X,d) un espace métrique, et $(x_n)_{n\in\mathbb{N}}$ une suite de X. Alors*

$$\mathrm{Adh}((x_n)_{n\in\mathbb{N}}) = \left\{x \in X, \exists \phi \text{ extraction telle que } x_{\phi(n)} \to x\right\}.$$

Rappelons qu'une fonction d'extraction est par définition une fonction strictement croissante de $\mathbb{N}$ dans $\mathbb{N}$.

Remarque 1.53. Autrement dit, une valeur d'adhérence de la suite $(x_n)_{n\in\mathbb{N}}$ est une limite d'une suite extraite de $(x_n)_{n\in\mathbb{N}}$. Il est fréquent de lire cette définition dans la littérature. Mais dans le cadre des espaces topologiques, l'égalité du lemme 1.52 est fausse, et c'est bien la définition 1.49 qui correspond à l'ensemble des valeurs d'adhérences (et la formule du lemme 1.51 reste valable).

Démonstration. Remarquons d'abord que

$$x \in \bigcap_{n\in\mathbb{N}} \overline{\{x_m, m \geq n\}} \Leftrightarrow \Big(\forall \varepsilon > 0, \forall n \in \mathbb{N}, \exists m \geq n, x_m \in B(x,\varepsilon)\Big).$$

Nous raisonnons alors par double inclusion : soit $x \in \cap_{n\in\mathbb{N}}\overline{\{x_m, m \geq n\}}$. Nous allons construire ϕ par récurrence. Posons $\phi(0) = 0$; alors $d(x_{\phi(0)}, x) < +\infty$. Supposons construits $\phi(0) < \phi(1) < \ldots < \phi(n)$ tels que $d(x_{\phi(k)}, x) < \frac{1}{k+1}$ pour $k \in [\![0,n]\!]$. On pose alors

$$\phi(n+1) = \min\left\{m \geq \phi(n)+1, d(x_m,x) < \frac{1}{n+2}\right\}$$

où cet élément existe car par ce qui précède l'ensemble considéré est non vide et que toute partie non vide de $\mathbb{N}$ admet un plus petit élément.

Ainsi par construction, ϕ est strictement croissante, et pour tout $n \in \mathbb{N}, d(x_{\phi(n)}, x) < \frac{1}{n+1}$ ce qui montre que x est limite de la suite extraite par ϕ.

Soit maintenant $x \in X$ tel qu'il existe une extraction ϕ avec $x_{\phi(n)} \xrightarrow[n\to+\infty]{} x$ et soit $\varepsilon > 0$. Il existe donc $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0, d(x_{\phi(n)}, x) < \varepsilon$. Soit alors $n \in \mathbb{N}$. Si $n \geq n_0$, on prend $m = \phi(n)$ qui est supérieur ou égal à n[14]. Si $n < n_0$, on prend $m = \phi(n_0) \geq n_0 \geq n$. Dans les deux cas, on a trouvé $m \geq n$ tel que $x_m \in B(x,\varepsilon)$, ce qui conclut la démonstration. □

Exercice 1.54. Soit (X,d) un espace métrique.

1. Montrer que si $(x_n)_{n\in\mathbb{N}}$ est une suite d'éléments de X qui converge vers $x_\infty \in X$, alors

$$\mathrm{Adh}((x_n)_{n\in\mathbb{N}}) = \{x_\infty\}.$$

2. Montrer que la réciproque de la question précédente est fausse en général[15], en exhibant une suite de réels $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ qui a une seule valeur d'adhérence dans $\mathbb{R}$, mais qui ne converge pas.

1.3 Continuité

On va maintenant généraliser la notion de continuité introduite au paragraphe 3.2 du Chapitre I, en prenant des espaces de départ et d'arrivée plus généraux :

Définition 1.55. Soit (X_1,d_1) et (X_2,d_2) deux espaces métriques, $f : X_1 \to X_2$ et $x \in X_1$. On dit que f est continue en x si

$$\forall \varepsilon > 0, \;\; \exists \alpha > 0, \;\; \forall y \in B_{X_1}(x,\alpha), \;\; f(y) \in B^f_{X_2}(f(x),\varepsilon),$$

c'est-à-dire

$$\forall \varepsilon > 0, \;\; \exists \alpha > 0, \;\; \forall y \in X_1, \; d_1(y,x) < \alpha \;\Rightarrow\; d_2(f(y),f(x)) \leq \varepsilon.$$

On dit que f est continue sur X_1 si pour tout $x \in X_1$, f est continue en x.

14. Rappelons le lemme très simple et classique : $\forall n \in \mathbb{N}, \phi(n) \geq n$, voir exercice 2.20.
15. Elle est vraie si l'espace (X,d) est compact.

Remarque 1.56. Comparez avec la définition 3.8 du Chapitre I pour une fonction $f : D \to \mathbb{R}$ où $D \subset \mathbb{R}$ est muni de la métrique induite par la valeur absolue sur $\mathbb{R}$.

Exercice 1.57. Soit (X, d) un espace métrique.

1. Soit $x_0 \in X$. Montrer que l'application
$$\begin{array}{rccc} d(x_0,\cdot): & X & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & d(x_0,x) \end{array}$$
est continue sur X.
2. On munit $X \times X$ de la métrique produit (voir exercice 1.9) : montrer que d est continue sur $X \times X$.

Donnons maintenant des caractéristions topologiques et séquentielles de la continuité :

Proposition 1.58 (Caractérisation séquentielle ; situation ponctuelle)**.** *Soit (X_1, d_1) et (X_2, d_2) deux espaces métriques, $f : X_1 \to X_2$, et $x \in X_1$. Alors*
$$f \textit{ est continue en } x \iff \left\{ \begin{array}{c} \forall (x_n)_{n\in\mathbb{N}} \in X_1^{\mathbb{N}} \textit{ telle que } x_n \xrightarrow[n\to+\infty]{d_1} x, \\ \\ f(x_n) \xrightarrow[n\to+\infty]{d_2} f(x). \end{array} \right.$$

Démonstration. La preuve est une traduction de la preuve de la Proposition 3.4 au Chapitre I :
$\Rightarrow$: Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de X_1 qui converge vers x, et $\varepsilon > 0$. Comme f est continue en x, il existe $\alpha > 0$ tel que
$$\forall y \in B_{X_1}(x,\alpha), \quad d(f(x), f(y)) \leq \varepsilon.$$
On applique la définition de la convergence de $(x_n)_{n\in\mathbb{N}}$ avec $\frac{\alpha}{2} > 0$: il existe $n_0 \in \mathbb{N}$ tel que
$$\forall n \geq n_0, d_{X_1}(x_n, x) \leq \frac{\alpha}{2}, \quad \text{donc en particulier} \quad x_n \in B_{X_1}(x,\alpha).$$
On obtient ainsi
$$\forall n \geq n_0, \quad d_{X_2}(f(x_n), f(x)) \leq \varepsilon.$$
donc $(f(x_n))_{n\in\mathbb{N}}$ converge vers $f(x)$ dans X_2.
$\Leftarrow$: Montrons la contraposée : on suppose qu'il existe ε_0 tel que
$$\forall \alpha > 0, \exists y \in B_{X_1}(x,\alpha), \ \ d_{X_2}(f(y), f(x)) > \varepsilon_0.$$
Pour tout $n \in \mathbb{N}^*$, on applique cette propriété pour $\alpha = \frac{1}{n} > 0$, ce qui nous donne un élément $x_n \in X_1$ tel que $x_n \in B_{X_1}(x, \frac{1}{n})$ et $d(f(x_n), f(x)) > \varepsilon_0$. La suite $(x_n)_{n\in\mathbb{N}^*}$ ainsi définie converge vers x dans X_1, mais $(f(x_n))_{n\in\mathbb{N}}$ ne converge pas vers $f(x)$ dans X_2, ce qu'il fallait démontrer.

□

Proposition 1.59 (Caractérisation topologique : situation globale)**.** *Soit (X_1, d_1) et (X_2, d_2) deux espaces métriques, $f : X_1 \to X_2$. Alors les assertions suivantes sont équivalentes :*

1. *f est continue sur X_1,*
2. *l'image réciproque d'un ouvert de X_2 par f est un ouvert de X_1 :*
$$\forall O \subset X_2, \ \ O \textit{ ouvert } \Rightarrow \ f^{-1}(O) \textit{ est un ouvert de } X_1,$$
3. *l'image réciproque d'un fermé de X_2 par f est un fermé de X_1.*

Démonstration. 1. $\Rightarrow$ 2. Supposons f continue sur X_1, et O un ouvert de X_2. Soit $x \in f^{-1}(O)$. Comme $f(x) \in O$ et O ouvert, il existe $r > 0$ tel que $B_{X_2}(f(x), r) \subset O$. Par continuité de f en x pour $r > 0$, il existe $\alpha > 0$ tel que pour tout $y \in B_{X_1}(x, \alpha)$, $f(y) \in B_{X_2}(f(x), r)$. Ainsi par les deux propriétés précédentes, on a en particulier

$$\forall y \in B_{X_1}(x, \alpha), f(y) \in O, \qquad \text{i.e.} \qquad B_{X_1}(x, \alpha) \subset f^{-1}(O)$$

et donc $f^{-1}(O)$ est bien un ouvert.
2. $\Rightarrow$ 3. On suppose que l'image réciproque d'un ouvert de X_2 est un ouvert de X_1. Soit F un fermé de X_2. Alors $F^c = X_2 \setminus F$ est un ouvert de X_2, et comme $f^{-1}(F^c) = (f^{-1}(F))^c$ est un ouvert, on conclut bien que $f^{-1}(F)$ est un fermé de X_1.
3. $\Rightarrow$ 2. On raisonne exactement comme pour 2. $\Rightarrow$ 3.
2. $\Rightarrow$ 1. On suppose que l'image réciproque d'un ouvert de X_2 est un ouvert de X_1. Soit $x \in X_1$ et $\varepsilon > 0$. On applique l'hypothèse à la boule ouverte $B_{X_2}(f(x), \varepsilon)$, donc $f^{-1}(B_{X_2}(f(x), \varepsilon))$ est un ouvert, mais comme x appartient à cet ouvert, il existe $\alpha > 0$ tel que $B_{X_1}(x, \alpha) \subset f^{-1}(B_{X_2}(f(x), \varepsilon))$, c'est-à-dire que pour tout $y \in B_{X_1}(x, \alpha)$, $f(y) \in B_{X_2}(f(x), \varepsilon)$, d'où le résultat. $\square$

Exercice 1.60 (Principe de prolongement des égalités)**.**

1. Soit (X_1, d_1) et (X_2, d_2) deux espaces métriques, $f : X_1 \to X_2$ et $g : X_1 \to X_2$ deux fonctions continues. Montrer que s'il existe $D \subset X_1$ dense dans X_1 tel que $\forall x \in D,\ f(x) = g(x)$, alors $f = g$ sur X_1.
2. Soit $f : \mathbb{R} \to \mathbb{R}$ continue telle que $\forall n \in \mathbb{Z}, \forall x \in \mathbb{R}, f(nx) = nf(x)$. Montrer qu'il existe $\lambda \in \mathbb{R}$ tel que $\forall x \in \mathbb{R}, f(x) = \lambda x$.

Exercice 1.61 (Restrictions sur les espaces de départ et d'arrivée)**.** Soit (X_1, d_1) et (X_2, d_2) deux espaces métriques, et $f : X_1 \to X_2$.

1. On suppose $f : X_1 \to X_2$ continue.
   (a) Soit $A \subset X_1$ qu'on munit de la métrique induite. Montrer que $f_{|A} : A \to X_2$ est continue.
   (b) Soit $B \subset X_2$, muni de la métrique induite, et tel que $f(X_1) \subset B$. Montrer que $f : X_1 \to B$ est continue.
2. Soit $A \subset X_1$ qu'on munit de la métrique induite.
   (a) Donner un exemple tel que $f_{|A} : A \to X_2$ est continue, mais $f : X_1 \to X_2$ n'est continue en aucun point de A.
   (b) Montrer que si $f : X_1 \to X_2$ est continue en tout point de A, alors $f : A \to X_2$ est continue.

Exercice 1.62 (*Fonctions ouvertes et fermées)**.** Attention, on a bien pris les images réciproques dans la proposition 1.59. On dit d'une application $f : X_1 \to X_2$ qu'elle est ouverte si l'image directe d'un ouvert de X_1 est un ouvert de X_2, et qu'elle est fermée si l'image directe d'un fermé de X_1 est un fermé de X_2.

1. L'application $x \in \mathbb{R} \mapsto E(x) \in \mathbb{R}$ est-elle ouverte, fermée ? L'application $x \in \mathbb{R} \mapsto E(x) \in \mathbb{Z}$ est-elle ouverte, fermée ? (on munit $\mathbb{R}$ et $\mathbb{Z}$ de leur métriques usuelles).
2. Exhiber une fonction continue qui n'est pas ouverte, une fonction ouverte qui n'est pas continue, une fonction continue qui n'est pas fermée, et une fonction fermée qui n'est pas continue.
3. Montrer que si (X_1, d_1) et (X_2, d_2) sont deux espaces métriques, et si $f : X_1 \to X_2$ est bijective et continue, alors
$$f \text{ est un homéomorphisme } \Leftrightarrow\ f \text{ est ouverte } \Leftrightarrow\ f \text{ est fermée .}$$
4. Soit (X_1, d_1) et (X_2, d_2) sont deux espaces métriques. On suppose (X_1, d_1) compact, et $f : X_1 \to X_2$ continue. Montrer que f est fermée.

Proposition 1.63. *Soit $(X_1, d_1), (X_2, d_2), (X_3, d_3)$ des espaces métriques, $f : X_1 \to X_2$ et $g : X_2 \to X_3$ deux applications, et $x \in X_1$.*

— *Si f est continue en x et g est continue en $f(x)$, alors $g \circ f$ est continue en x.*

— *Si f est continue sur X_1 et g continue sur X_2, alors $g \circ f$ est continue sur X_1.*

Démonstration. On peut reproduire les deux preuves de la proposition 3.5 au Chapitre I. Nous reproduisons ici uniquement la seconde preuve, et invitons le lecteur à reproduire la première.

Soit $(x_n)_{n\in\mathbb{N}} \in X_1^{\mathbb{N}}$ une suite qui converge vers x. Comme f est continue en x, par la proposition 1.58, $f(x_n) \xrightarrow[n\to+\infty]{d_2} f(x)$. Comme g est continue en $f(x)$, on a $g \circ f(x_n) \xrightarrow[n\to+\infty]{d_3} g(f(x))$. Comme on a montré cette propriété pour toute suite $(x_n)_{n\in\mathbb{N}} \in X_1^{\mathbb{N}}$, on conclut bien que $g \circ f$ est continue en x. □

Exercice 1.64 (Période d'une fonction)**.** On considère $f : \mathbb{R} \to \mathbb{R}$. On dit que $a \in \mathbb{R}$ est appelé période de f si $\forall x \in \mathbb{R}, f(x+a) = f(x)$. On pose $P(f)$ l'ensemble des périodes de f.

1. Montrer que soit il existe $a_0 \in \mathbb{R}$ tel que $P(f) = a_0\mathbb{Z}$, soit $P(f)$ est dense dans $\mathbb{R}$.
2. Que dire de f si $P(f) = \mathbb{R}$?
3. Donner un exemple de fonction f telle que $P(f) = a_0\mathbb{Z}$ où $a_0 \in \mathbb{R}$; donner un exemple de fonction f telle que $P(f)$ est dense dans $\mathbb{R}$, mais distinct de $\mathbb{R}$.
4. On suppose f continue et non constante. Montrer que $P(f)$ est de la forme $a_0\mathbb{Z}$ pour $a_0 \in \mathbb{R}$. Si $a_0 \neq 0$, on dit que f est périodique, et on peut appeler $|a_0|$ "la" période de f, qui est le plus petit réel strictement positif que f admet pour période.

Exercice 1.65 (*Équivalences de métriques)**.** Attention, il existe plusieurs notions d'équivalence pour les métriques. Soit X un ensemble, et d_1, d_2 deux métriques sur X. On dit que d_1 et d_2 sont

— topologiquement équivalentes si elles définissent les mêmes ensembles ouverts dans X,

— uniformément équivalentes si l'application identité est uniformément continue de (X, d_1) dans (X, d_2) et aussi de (X, d_2) dans (X, d_1),

— bornologiquement équivalentes si elles sont uniformément équivalentes et si elles définissent les mêmes parties de diamètre fini,

— Lipschitz-équivalentes s'il existe $(a, b) \in]0, +\infty[^2$ telles que $ad_1 \leq d_2 \leq bd_1$.

1. Montrer que
$$\begin{aligned} &d_1 \text{ et } d_2 \text{ sont topologiquement équivalentes} \\ &\qquad\qquad \Leftrightarrow d_1 \text{ et } d_2 \text{ définissent les mêmes suites convergentes} \\ &\qquad\qquad \Leftrightarrow \text{l'application identité de } (X, d_1) \text{ vers } (X, d_2) \text{ est un homéomorphisme} \end{aligned}$$
2. Montrer que si d_1 et d_2 sont uniformément équivalentes, alors elles définissent les mêmes suites de Cauchy.
3. Montrer que
$$\begin{aligned} d_1 \text{ et } d_2 \text{ Lipschitz-équivalentes} &\Rightarrow d_1 \text{ et } d_2 \text{ bornologiquement équivalentes} \\ d_1 \text{ et } d_2 \text{ bornologiquement équivalentes} &\Rightarrow d_1 \text{ et } d_2 \text{ uniformément équivalentes} \\ d_1 \text{ et } d_2 \text{ uniformément équivalentes} &\Rightarrow d_1 \text{ et } d_2 \text{ topologiquement équivalentes} \end{aligned}$$
4. En considérant par exemple sur $\mathbb{R}$ les métriques
$$\begin{aligned} \forall (x, y) \in \mathbb{R}, \qquad & d_1(x, y) = |x - y|, \quad d_2(x, y) = |x^3 - y^3|, \\ & d_3(x, y) = \min\{1, |x - y|\}, \quad d_4(x, y) = \frac{|x - y|}{1 + |x - y|}, \end{aligned}$$
montrer que les réciproques des énoncés de la question précédentes ne sont pas vraies en général.

Comme dans le cas réel, on peut également définir des notions plus fortes que la continuité :

Définition 1.66. Soit (X_1, d_1) et (X_2, d_2) deux espaces métriques et $f : X_1 \to X_2$.

— On dit que f est uniformément continue sur X_1 si
$$\forall \varepsilon > 0, \ \exists \alpha > 0, \ \forall (x, y) \in X_1^2, \quad d_1(x, y) \leq \alpha \ \Rightarrow \ d_2(f(y), f(x)) \leq \varepsilon.$$

— On dit que f est lipschitzienne sur X_1 s'il existe $k \in \mathbb{R}$ tel que
$$\forall (x, y) \in X_1^2, \ d_2(f(y), f(x)) \leq kd_1(x, y).$$

Exercice 1.67. Montrer qu'une fonction lipschitzienne est uniformément continue.

1.4 Complétude

1.4.1 Définitions

Définition 1.68. Soit (X,d) un espace métrique. On dit que $(x_n)_{n\in\mathbb{N}} \in X^{\mathbb{N}}$ est de Cauchy [16] si

$$\forall \varepsilon > 0,\ \exists n_0 \in \mathbb{N},\ \forall p \geq n_0, \forall q \geq n_0,\ \ d(x_q, x_p) \leq \varepsilon. \tag{1.4}$$

Exercice 1.69. Soit (X,d) un espace métrique. Montrer que si $(x_n)_{n\in\mathbb{N}}$ est une suite convergente de l'espace (X,d), alors elle est de Cauchy.

Exercice 1.70. Soit (X,d) un espace métrique. Montrer qu'une suite de Cauchy est bornée, c'est-à-dire qu'il existe $x \in X$ et $M \in \mathbb{R}$ tels que $\forall n \in \mathbb{N},\ \ x_n \in B_X^f(x, M)$.

Exercice 1.71. Soit (X,d) un espace métrique. Montrer que si $(x_n)_{n\in\mathbb{N}}$ est une suite de Cauchy de X ayant une valeur d'adhérence, alors $(x_n)_{n\in\mathbb{N}}$ est convergente et converge vers cette valeur d'adhérence.

Définition 1.72. Soit (X,d) un espace métrique. On dit que (X,d) est complet si toute suite de X qui est de Cauchy est convergente dans X, c'est-à-dire

$$\forall (x_n)_{n\in\mathbb{N}} \text{ telle que (1.4)},\ \ \exists \ell \in X,\ \ x_n \xrightarrow[n\to+\infty]{d} \ell.$$

Exemple 1.73. L'ensemble $\mathbb{R}$ muni de sa distance naturelle est un espace complet, voir le paragraphe 1 du Chapitre I.

L'ensemble $(\mathbb{Q}, |\cdot|)$ n'est par contre pas un espace complet. Pour le voir, il suffit de considérer une suite de rationnels qui converge vers un réel qui n'est pas un rationnel. Une telle suite existe, par densité de $\mathbb{Q}$ dans $\mathbb{R}$. En tant que suite convergente dans $\mathbb{R}$, elle est de Cauchy. Elle l'est donc aussi vue comme une suite de $\mathbb{Q}$. Mais elle n'est pas convergente dans $\mathbb{Q}$.

Pour un exemple explicite de telle suite, on peut par exemple considérer :

$$\forall n \in \mathbb{N},\quad x_n = \sum_{k=0}^{n} \frac{1}{k!}.$$

Par rapport à l'exemple précédent, on a plus généralement la caractérisation suivante de la complétude des sous-ensembles d'un espace complet :

Proposition 1.74. *Soit (X,d) un espace métrique complet, et $A \subset X$. Alors en notant d_A la restriction de d à $A \times A$, on a*

$$(A, d_A) \text{ est complet } \ \Leftrightarrow\ A \text{ est fermé dans } X.$$

Remarque 1.75. Comme on le verra dans la preuve, le sens $\Rightarrow$ dans la proposition précédente n'utilise pas la complétude de (X,d).

Démonstration. $\Rightarrow$ Supposons (A, d_A) complet. Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de A qui converge vers $x \in X$. D'après l'exercice 1.69, la suite $(x_n)_{n\in\mathbb{N}}$ est donc de Cauchy comme suite de X, et comme c'est une suite d'éléments de A, elle est donc de Cauchy dans (A, d_A) qui est complet, et donc $(x_n)_{n\in\mathbb{N}}$ est convergente dans A. D'après l'unicité de la limite, $x \in A$ et donc par la proposition 1.45, A est fermé.
$\Leftarrow$ Supposons A fermé dans X. Soit $(x_n)_{n\in\mathbb{N}}$ une suite de Cauchy d'éléments de A. Vue comme une suite d'éléments de X, elle est de Cauchy également, et par complétude de (X,d), il existe $x \in X$ telle que $(x_n)_{n\in\mathbb{N}}$ converge vers X. Mais comme A est fermé, $x \in A$, d'où le résultat. □

Exercice 1.76. Soit (X, d_X) un espace métrique complet.

1. (*) Soit d'_X une autre métrique sur X, équivalente à d_X en l'un des quatre sens donnés à l'exercice 1.65. Déterminer si cette équivalence permet de dire que (X, d'_X) est encore complet, dans le cas négatif, exhiber un contre-exemple.
2. Montrer que si (Y, d_Y) est isométrique à (X, d_X), alors (Y, d_Y) est complet.

Exercice 1.77. Montrer qu'un produit fini d'espaces métriques complets est complet pour la métrique produit (voir exercice 1.9).

16. Encore une fois, s'il y a ambiguïté, il vaut mieux dire "de Cauchy dans (X,d)", ou au moins "de Cauchy pour la métrique d".

1.4.2 Propriétés des espaces complets

Intéressons-nous maintenant aux propriétés des espaces métriques complets :

Proposition 1.78 (Propriété des fermés emboîtés)**.** *Soit (X, d) un espace métrique complet, et $(F_n)_{n\in\mathbb{N}}$ une suite de parties fermées non vides de X, décroissante pour l'inclusion, et dont le diamètre tend vers 0. Alors l'intersection des $(F_n)_{n\in\mathbb{N}}$ est un singleton.*

Remarque 1.79. Avec les notations de l'énoncé, le fait que l'intersection des $(F_n)_{n\in\mathbb{N}}$ contient au plus un élément suit des hypothèses sur les ensembles F_n pour $n \in \mathbb{N}$: en effet le diamètre de cette intersection est inférieur au diamètre de tous les F_n pour $n \in \mathbb{N}$, et donc ne peut être que 0. La complétude de (X, d) permet de voir que cette intersection n'est pas vide, et donc est au final exactement un singleton.

Exercice 1.80. On dit d'un espace métrique qu'il satisfait la propriété des segments emboîtés si pour toute suite de fermés non vides, décroissante pour l'inclusion et dont le diamètre tend vers 0, l'intersection de ces fermés est non vide.

1. Montrer que $(]0, 1[, |\cdot|)$ et $(\mathbb{Q}, |\cdot|)$ ne satisfont pas la propriété des segments emboîtés.
2. Montrer que si (X, d) satisfait la propriété des segments emboîtés, alors (X, d) est complet.
   Ainsi, la propriété des segments emboîtés caractérise la complétude.

Démonstration. Pour tout $n \in \mathbb{N}$, comme F_n est non vide, il existe $x_n \in F_n$. Nous allons montrer que $(x_n)_{n\in\mathbb{N}}$ est de Cauchy : si $\varepsilon > 0$, alors par convergence du diamètre de F_n vers 0, il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, $\operatorname{diam}(F_n) \leq \varepsilon$. Mais alors pour $p \geq n_0$ et $q \geq n_0$, $x_p \in F_p \subset F_{n_0}$ et $x_q \in F_q \subset F_{n_0}$ donc par définition du diamètre, $d(x_p, x_q) \leq \operatorname{diam}(F_{n_0}) \leq \varepsilon$.

Par complétude de (X, d), il existe $x_\infty \in X$ telle que $(x_n)_{n\in\mathbb{N}}$ converge vers x_∞. Étant donné $n \in \mathbb{N}$, on a $\forall k \geq n, x_k \in F_k \subset F_n$, et donc par la caractère fermé de F_n, on a $x_\infty \in F_n$. Comme on a montré cela pour tout $n \in \mathbb{N}$, on a bien montré que $\cap_{n\in\mathbb{N}} F_n$ était non vide. □

Théorème 1.81 (Théorème de point fixe de Picard)**.** *Soit (X, d) un espace métrique complet non vide, et $f : X \to X$ une application contractante, c'est-à-dire telle qu'il existe $k \in [0, 1[$ telle que*

$$\forall (x, y) \in X^2, \quad d(f(y), f(x)) \leq kd(y, x).$$

Alors f admet un unique point fixe dans X.

De plus, si on définit

$$\begin{cases} x_0 \in X \\ \forall n \in \mathbb{N}, \quad x_{n+1} = f(x_n) \end{cases} \tag{1.5}$$

alors la suite $(x_n)_{n\in\mathbb{N}}$ converge vers l'unique point fixe ℓ de f, et de plus

$$\forall n \in \mathbb{N}, \quad d(x_n, \ell) \leq k^n d(x_0, \ell).$$

Remarque 1.82. Même si la première partie de l'énoncé est la plus importante, nous conseillons de ne pas oublier la partie sur la suite récurrente, qui sert souvent, et suggère aussi la preuve de ce théorème central du programme. La preuve étant relativement simple, il arrive assez souvent, en fait, qu'on redémontre ce théorème dans le contexte qui nous intéresse. Remarquons aussi que la récurrence (2.2) a été définie pour un choix arbitraire de $x_0 \in X$, et la suite récurrente $(x_n)_{n\in\mathbb{N}}$ dépend bien sûr de ce x_0, mais l'énoncé affirme entre autre que sa limite n'en dépendra pas. La fin de l'énoncé permet de voir que la convergence de la suite vers sa limite est rapide.

Remarque 1.83. Attention, il est fréquent de passer à côté de l'hypothèse que la fonction f doit aller de X dans X. Si en effet on veut appliquer le théorème à une restriction de f sur un sous-ensemble A de X, il faut que (A, d_A) soit complet (c'est-à-dire fermé si (X, d) est lui-même complet) et stable par f, c'est-à-dire $f(A) \subset A$.

Démonstration. Commençons par prouver l'unicité du point fixe de f. Si $(x_1, x_2) \in X^2$ sont tous les deux points fixes de f, alors on a

$$d(x_1, x_2) = d(f(x_1), f(x_2)) \leq kd(x_1, x_2),$$

et comme $k < 1$, nécessairement $d(x_1, x_2) = 0$ et donc $x_1 = x_2$.

Pour l'existence du point fixe, on commence par utiliser que X est non vide, donc contient $x_0 \in X$. On définit alors $(x_n)_{n\in\mathbb{N}}$ par (2.2). On peut alors montrer d'une part par une récurrence élémentaire

$$\forall n \in \mathbb{N},\ \ d(x_{n+1}, x_n) \leq k^n d(x_1, x_0),$$

et d'autre part

$$\forall q \geq p \geq 0,\ \ d(x_q, x_p) \leq \sum_{i=p}^{q-1} d(x_{i+1}, x_i) \leq \left(\sum_{i=p}^{q-1} k^i\right) d(x_1, x_0) \leq \frac{k^p}{1-k} d(x_1, x_0).$$

Comme $k \in [0, 1[$, on sait que si $\varepsilon > 0$, il existe $n_0 \in \mathbb{N}$ tel que pour tout $p \geq n_0$, $\frac{k^p}{1-k} d(x_1, x_0) \leq \varepsilon$, et donc la suite $(x_n)_{n\in\mathbb{N}}$ est de Cauchy dans (X, d), elle converge donc vers $\ell \in X$. En passant à la limite dans l'égalité $\forall n \in \mathbb{N}, x_{n+1} = f(x_n)$, et en utilisant la continuité de f (qui vient du caractère lipschitzien), on récupère $\ell = f(\ell)$, ainsi ℓ est un point fixe de f, ce qui prouve l'existence d'un tel point fixe, et aussi que toute suite définie par (2.2) converge vers ℓ, puisque ce point fixe est unique.

Pour obtenir la dernière estimation, on écrit

$$\forall n \geq 1, d(x_n, \ell) = d(f(x_{n-1}), f(\ell)) \leq kd(x_{n-1}, \ell)$$

d'où le résultat par récurrence immédiate. □

Exercice 1.84.
1. Montrer qu'il n'y a en général ni existence, ni unicité si f est k-lipschitzienne avec $k \geq 1$.
2. Montrer qu'il n'y a en général plus existence si f est telle que

$$\forall (x, y) \in X^2,\ \ x \neq y\ \Rightarrow d(f(x), f(y)) < d(x, y),$$

mais qu'il y a quand même unicité.
3. Montrer qu'il n'y a en général plus existence si (X, d) n'est pas supposé complet.

Exercice 1.85. Soit (X, d) un espace métrique complet, et $f : X \to X$. On suppose qu'il existe $n \in \mathbb{N}^*$ telle que f^n (la composée n-ième de f) est contractante, alors f admet un unique point fixe.

Exercice 1.86. *[Cas particulier du théorème de Cauchy-Lipschitz ; on discute dans cet exercice une des applications les plus importantes du théorème de point fixe, en montrant sur un exemple très particulier comment en déduire l'existence et l'unicité d'une solution à une équation différentielle].* On pose $E = C^0([0, 1])$, que l'on munit de la norme uniforme $\|\cdot\|_\infty$. On définit

$$\begin{array}{llll} T : & E & \longrightarrow & E \\ & f & \longmapsto & T(f) \text{ définie par : } \ \forall\, x \in [0,1],\ \ T(f)(x) = 1 + \dfrac{1}{2}\displaystyle\int_0^x \sin(f(t))dt \end{array}$$

1. Montrer T est bien définie, et qu'il existe un unique élément $y \in E$ tel que $T(y) = y$.
2. Montrer que y est de classe C^1 et vérifie l'équation différentielle $y' = \frac{1}{2}\sin(y)$, et $y(0) = 1$.
3. Montrer que si z est une fonction C^1 satisfaisant $z' = \frac{1}{2}\sin(z)$ et $z(0) = 1$, alors $T(z) = z$. En déduire $z = y$.

Théorème 1.87 (Théorème de prolongement des applications uniformément continues)**.** *Soit (X_1, d_1) et (X_2, d_2) deux espaces métriques et $A \subset X_1$. On suppose (X_2, d_2) complet. Si $f : A \to X_2$ est uniformément continue sur A et si A est dense dans X_1, alors il existe une unique fonction continue qui prolonge f sur X_1, c'est-à-dire qu'il existe une unique fonction $\widetilde{f} : X_1 \to X_2$ continue et telle que $\forall x \in A, \widetilde{f}(x) = f(x)$.*

De plus, $\widetilde{f}$ est uniformément continue sur X_1.

Démonstration. L'unicité du prolongement découle de l'exercice 1.60.

Pour l'existence, on se donne $x \in X_1$. Par densité de A dans X_1, il existe $(a_n)_{n\in\mathbb{N}}$ une suite d'éléments de A qui converge vers x. Pour $\varepsilon > 0$, par uniforme continuité de f, il existe $\alpha > 0$ tel que

$$\forall (y,z) \in X_1^2, \;\; d_1(y,z) \leq \alpha \Rightarrow d_2(f(y), f(z)) \leq \varepsilon. \tag{1.6}$$

Comme $(a_n)_{n\in\mathbb{N}}$ est de Cauchy (car convergente), il existe $n_0 \in \mathbb{N}$ telle que pour tout $p \geq n_0$ et $q \geq n_0, d_1(a_q, a_p) \leq \alpha$. Ainsi avec (1.6) on obtient

$$\forall p \geq n_0, \forall q \geq n_0, \quad d_2(f(a_q), f(a_p)) \leq \varepsilon.$$

On a donc montré que la suite $(f(a_n))_{n\in\mathbb{N}}$ est de Cauchy dans (X_2, d_2). Par complétude, il existe $\ell \in X_2$ limite de $(f(a_n))_{n\in\mathbb{N}}$.

Il faut vérifier que cette limite ne dépend que de x et non du choix de la suite $(a_n)_{n\in\mathbb{N}} \in A^{\mathbb{N}}$ qui converge vers x. Soit donc $(b_n)_{n\in\mathbb{N}}$ une suite d'éléments de A qui converge aussi vers x. Alors par ce qui précède $(f(b_n))_{n\in\mathbb{N}}$ converge vers $\ell' \in X_2$. Comme

$$d_1(a_n, b_n) \leq d_1(a_n, x) + d_1(x, b_n) \xrightarrow[n\to+\infty]{} 0,$$

il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, $d_1(a_n, b_n) \leq \alpha$, et donc

$$\forall n \geq n_0, \quad d_2(f(a_n), f(b_n)) \leq \varepsilon. \tag{1.7}$$

Cela montre que $\ell = \ell'$ [17], et donc on peut noter $\widetilde{f}(x)$ cette limite.

Dans le cas où $x \in A$, on peut choisir pour $(a_n)_{n\in\mathbb{N}}$ la suite constante égale à x, et donc $\widetilde{f}(x) = f(x)$.

Il reste à montrer l'uniforme continuité de $\widetilde{f}$. Pour cela on travaille encore avec (1.6) et on se donne $(x_1, x_2) \in X_1^2$ tels que $d_1(x_1, x_2) \leq \frac{\alpha}{3}$. Il existe $(a_n)_{n\in\mathbb{N}} \in A^{\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}} \in A^{\mathbb{N}}$ qui convergent respectivement vers x_1 et x_2. Ainsi il existe $n_0 \in \mathbb{N}$ tel que

$$\forall n \geq n_0, \;\; d_1(a_n, x_1) \leq \frac{\alpha}{3}, \quad \text{et} \quad d_1(b_n, x_2) \leq \frac{\alpha}{3}.$$

Par l'inégalité triangulaire, on a pour tout $n \geq n_0, d_1(a_n, b_n) \leq \alpha$, et donc par (1.6),

$$\forall n \geq n_0, \;\; d_2(f(a_n), f(b_n)) \leq \varepsilon.$$

Mais par construction les suites $(f(a_n))_{n\in\mathbb{N}}$ et $(f(b_n))_{n\in\mathbb{N}}$ convergent vers $\widetilde{f}(x_1)$ et $\widetilde{f}(x_2)$ respectivement, donc il existe $n_1 \in \mathbb{N}$ tel que

$$\forall n \geq n_1, \;\; d_2(f(a_n), \widetilde{f}(x_1)) \leq \varepsilon \quad \text{et} \quad d_2(f(b_n), \widetilde{f}(x_2)) \leq \varepsilon.$$

Au final, on a avec $N = \max\{n_0, n_1\}$:

$$d_2(\widetilde{f}(x_1), \widetilde{f}(x_2)) \leq d_2(\widetilde{f}(x_1), f(a_N)) + d_2(f(a_N), f(b_N)) + d_2(f(b_N), \widetilde{f}(x_2)) \leq 3\varepsilon,$$

d'où le résultat. □

Remarque 1.88. On a montré au cours de la preuve un petit résultat qui vaut le coup d'être retenu : si une fonction est uniformément continue, alors l'image d'une suite de Cauchy par cette fonction, est une suite de Cauchy. On invite le lecteur à isoler (ou mieux encore, faire par lui-même) la preuve de ce fait, et à trouver un contre-exemple si on suppose la fonction seulement continue.

Exemple 1.89. Un cas très particulier du théorème précédent est le suivant : soit I un intervalle de $\mathbb{R}$ et $a \in I$. On suppose $f : I \setminus \{a\} \to \mathbb{R}$ uniformément continue. Alors comme $I \setminus \{a\}$ est dense dans I, il existe une unique fonction $\widetilde{f} : I \to \mathbb{R}$ continue qui prolonge f.

Exercice 1.90. Soit $a < b$ deux réels, et $f :]a, b] \to \mathbb{R}$ une fonction dérivable sur $]a, b]$ et telle que f' a une limite réelle en a.

1. Montrer que f est uniformément continue sur $]a, b]$ et en déduire qu'on peut la prolonger par continuité sur $[a, b]$.
2. Montrer que ce prolongement est dérivable en a.
3. Les résultats restent-ils valables sur on remplace $]a, b]$ par $I \setminus \{a\}$ où a est à l'intérieur de I ?

17. On peut invoquer la continuité de d_2 (voir exercice 1.57) en faisant $n \to +\infty$ puis $\varepsilon \to 0$ dans (1.7), ou simplement écrire :

$$d_2(\ell, \ell') \leq d_2(\ell, f(a_n)) + d_2(f(a_n), f(b_n)) + d_2(f(b_n), \ell') \xrightarrow[n\to+\infty]{} 0.$$

1.5 Compacité

1.5.1 Définition topologique

Définition 1.91 (Propriété de Borel-Lebesgue). — Un espace métrique (X, d) est dit *compact* si de tout recouvrement de X par une famille d'ouverts de X, on peut extraire un sous-recouvrement fini, c'est-à-dire que si $\{U_i\}_{i\in I}$ est une famille d'ouverts de X telle que $X = \bigcup_{i\in I} U_i$, alors il existe $m \in \mathbb{N}^*$ et $i_1, \ldots, i_m \in I$ tels que $X = \bigcup_{k=1}^{m} U_{i_k}$.

— Si (X, d) est un espace métrique et $A \subset X$, on dit que A est une partie compacte de X si en le munissant de sa métrique induite, (A, d_A) est un espace compact.

Remarque 1.92. De façon équivalente en passant aux complémentaires, X est compact si et seulement si on peut extraire d'une famille de fermés de X d'intersection vide une sous-famille finie d'intersection également vide. Si X est compact, on utilise cette version plus souvent par sa contraposée : étant donnée une famille $(F_i)_{i\in I}$ de fermés de X, si toute intersection d'une sous-famille finie des $(F_i)_{i\in I}$ est non vide, alors l'intersection de tous les $(F_i)_{i\in I}$ est également non vide.

Remarque 1.93. La définition 1.91 étant formulée uniquement à partir des ouverts de X, on dit que la compacité est une notion topologique. On en déduit par exemple que si $f : X \to Y$ est un homéomorphisme entre (X, d_X) et (Y, d_Y) et que (X, d_X) est compact, alors (Y, d_Y) est aussi compact.

Exemple 1.94. Pour voir que $\mathbb{R}$ n'est pas compact, on peut écrire

$$\mathbb{R} = \bigcup_{n\in\mathbb{N}^*}]-n, n[, \qquad \text{ou encore} \qquad \bigcap_{n\in\mathbb{N}} [n, +\infty[= \emptyset.$$

Remarque 1.95. Si (X, d) est un espace métrique et $A \subset X$, la compacité de A signifie que si $(V_i)_{i\in I}$ est une famille d'ouverts <u>de A</u> telle que $A = \bigcup_{i\in I} V_i$, alors on peut extraire de ce recouvrement un sous-recouvrement fini.

Avec l'exercice 1.27, on peut réexprimer cela avec des ouverts de X : A est compact si et seulement si pour tout famille d'ouverts $\{U_i\}_{i\in I}$ <u>de X</u> telle que

$$A \subset \bigcup_{i\in I} U_i, \quad \text{il existe } J \subset I \text{ fini tel que} \quad A \subset \bigcup_{i\in J} U_i.$$

Exercice 1.96. Voir [Mar02, page 96 et 340]. Soit (X, d) un espace métrique.

1. Soit $(x_n)_{n\in\mathbb{N}}$ une suite de X qui converge vers $a \in X$. Montrer que l'ensemble $\{x_n, n \in \mathbb{N}\} \cup \{a\}$ est compact.
2. (*) Soit (Y, d') un autre espace métrique. Montrer que si $f : X \to Y$ est telle que ses restrictions à tout compact est continue, alors f elle-même est continue.

Exercice 1.97 (*Une partie compacte est fermée et de diamètre fini[18] ; preuve topologique)**.** Soit (X, d) un espace métrique non vide, et $A \subset X$ une partie compacte.

1. Montrer que A est de diamètre fini.
2. Montrons maintenant que A est fermée. Si $X = A$, il n'y a rien à faire. On suppose donc $A \neq X$ et on prend $x \in A^c$.
   (a) Montrer que pour tout $y \in A$, il existe $r_y > 0$ tel que
   $$B(y, r_y) \cap B(x, r_y) = \emptyset.$$

18. Noter que la réciproque est fausse en général. Par exemple, dans les espaces vectoriels normés de dimension finie, les parties compactes sont exactement les parties fermées et bornées, mais cette caractérisation est toujours fausse dans les espace vectoriels normés de dimension infinie, c'est-à-dire qu'il y existe des fermés-bornés non compacts, voir le paragraphe 2.4.

(b) Montrer qu'il existe $m \in \mathbb{N}^*$ et $(y_1, \dots y_m) \in A^m$ tels que

$$A \subset \bigcup_{i=1}^{m} B(y_i, r_{y_i}).$$

(c) Montrer qu'il existe $r > 0$ tel que $B(x, r) \subset A^c$, et conclure que A est fermé dans X.

Exercice 1.98 (Un fermé dans un compact est compact ; preuve topologique)**.** Soit (X, d) un espace métrique compact et $A \subset X$. Montrer que si A est fermé, alors A est compact [19].

1.5.2 Caractérisation séquentielle

Définition 1.99 (Propriété de Bolzano-Weierstrass). Un espace métrique (X, d) est dit *séquentiellement compact* si de toute suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de X, on peut extraire une sous-suite convergente dans X.

Il s'avère que ces deux notions coïncident dans les espaces métriques (mais pas dans un cadre topologique plus général).

Théorème 1.100 (Théorème de Borel-Lebesgue ou de Heine-Borel). *Soit (X, d) un espace métrique. Alors X est compact si et seulement s'il est séquentiellement compact.*

Démonstration. **Étape 1 : compact implique séquentiellement compact.**

Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de X. La famille de fermés $F_n = \overline{\{x_m, m \geq n\}}$ indexée par $n \in \mathbb{N}$ est décroissante pour l'inclusion, donc si l'intersection d'une sous-famille finie était vide, cela signfierait qu'il existe $n_0 \in \mathbb{N}$ tel que F_{n_0} est vide, ce qui n'est pas le cas. Par compacité de (X, d), cela implique que $\bigcap_{n\in\mathbb{N}} F_n$ est non vide, et donc par le lemme 1.52, il existe une sous-suite de $(x_n)_{n\in\mathbb{N}}$ qui est convergente.

Étape 2 : [20] On suppose désormais et jusqu'à la fin de la preuve que (X, d) est séquentiellement compact, et on considère $\{U_i\}_{i\in I}$ un recouvrement ouvert de X.

Montrons qu'il existe un $\rho > 0$ tel que pour tout $x \in X$, il existe $i \in I$ tel que $B(x, \rho) \subset U_i$. On raisonne par l'absurde en supposant que pour tout $\rho > 0$, il existe un $x \in X$ tel que pour tout $i \in I$, $B(x, \rho) \not\subset U_i$. En particulier avec $\rho = \frac{1}{n+1}$ pour chaque $n \in \mathbb{N}$, on construit une suite $(x_n)_{n\in\mathbb{N}}$ telle que $B(x_n, \frac{1}{n+1}) \not\subset U_i$ pour tout $i \in I$. Par hypothèse, on peut en extraire une sous-suite $(x_{\phi(n)})_{n\in\mathbb{N}}$, qui converge vers un certain $x \in X$. Comme X est recouvert par $\cup_{i\in I} U_i$, il existe $i_0 \in I$ tel que $x \in U_{i_0}$, et comme U_{i_0} est ouvert, il existe $r > 0$ tel que $B(x, r) \subset U_{i_0}$. Or

$$\forall y \in B\left(x_{\phi(n)}, \frac{1}{\phi(n)+1}\right), \quad d(x, y) \leq d(x, x_{\phi(n)}) + d(x_{\phi(n)}, y) \leq d(x, x_{\phi(n)}) + \frac{1}{\phi(n)+1},$$

mais il existe $n_0 \in \mathbb{N}$ tel que pour tout $n \geq n_0$, $d(x, x_{\phi(n)}) + \frac{1}{\phi(n)+1} < r$, et donc en particulier pour $n = n_0$ on a $B\left(x_{\phi(n_0)}, \frac{1}{\phi(n_0)+1}\right) \subset B(x, r) \subset U_{i_0}$, ce qui constitue une contradiction.

Étape 3 : Montrons à présent que X peut être recouvert par un nombre fini de boules de rayon ρ. Supposons par l'absurde que ce n'est pas le cas. Nécessairement X est alors non vide, donc il existe $x_0 \in X$. Si la propriété de recouvrement finie par des boules de rayon ρ n'est pas vraie, alors en particulier la boule $B(x_0, \rho)$ ne recouvre pas X. Il existe donc $x_1 \in X$ tel que $d(x_0, x_1) \geq \rho$. Pour $p \in \mathbb{N}$, supposons construit un $(p+1)$-uplet $(x_0, x_1, \dots, x_p)$ tel que $d(x_m, x_n) \geq \rho$ pour tous $m \neq n$. Comme $\bigcup_{i=0}^{p} B(x_i, \rho)$ ne recouvre pas X, on peut trouver un x_{p+1} dans son complémentaire. On a en particulier $d(x_i, x_{p+1}) \geq \rho$ pour tout $i \leq p$, d'où la construction au rang suivant. La suite $(x_n)_{n\in\mathbb{N}}$ ainsi construite ne possède aucune sous-suite qui soit de Cauchy, donc ne possède aucune sous-suite convergente, ce qui constitue une contradiction.

19. Ainsi avec l'exercice 1.97, on a que si X est compact, alors ses sous-parties sont compactes si et seulement si elles sont fermées. Comparer avec la proposition 1.74.

20. On appelle parfois lemme de Lebesgue cette étape.

Étape 4 : conclusion. On a donc montré l'existence de $x_1, \dots, x_m \in X$ tels que $X = \bigcup_{k=1}^m B(x_k, \rho)$. Pour chaque $k \in [\![1, m]\!]$, il existe un indice $i_k \in I$ tel que $B(x_k, \rho) \subset U_{i_k}$, donc

$$X \subset \bigcup_{k=1}^{m} U_{i_k},$$

ce qui conclut la preuve vu que l'inclusion réciproque est triviale. □

On a donc maintenant deux façons d'étudier les ensembles compacts, soit de façon topologique comme au paragraphe 1.5.1, soit de façon séquentielle. Par exemple, reprenons l'exercice 1.97 avec une approche séquentielle :

Proposition 1.101. *Soit (X, d) un espace métrique, et A une partie séquentiellement compacte et non vide de X. Alors A est de diamètre fini et fermé dans X.*

Démonstration. Soit $(x_n)_{n\in\mathbb{N}}$ une suite de A qui converge vers $x \in X$. Comme A est séquentiellement compacte, on peut extraire une sous-suite qui converge vers $x_\infty \in A$. Par unicité de la limite, on en déduit que $x = x_\infty \in A$, ce qui montre que A est fermé dans X.

Comme A est non vide, il existe $x_0 \in A$. Si par l'absurde pour tout $n \in \mathbb{N}$, A n'est pas inclus dans $B(x_0, n)$, alors il existe $x_n \in A \setminus B(x_0, n)$. Par compacité séquentielle de A, on peut extraire une sous-suite $(x_{\phi(n)})_{n\in\mathbb{N}}$ qui est convergente. Cette suite est donc bornée (par l'exercice 1.70) donc il existe $n_0 \in \mathbb{N}$ tel que $\forall n \in \mathbb{N}, x_{\phi(n)} \in B(x_0, n_0)$. Mais alors $x_{\phi(n_0+1)} \notin B(x_0, \phi(n_0+1))$ et $x_{\phi(n_0+1)} \in B(x_0, n_0)$ ce qui est une contradiction puisque $\phi(n_0) + 1 > n_0$. Donc A est de diamètre fini. □

Remarque 1.102. On déduit également du théorème précédent le résultat : soit (X, d) un espace métrique. Alors

$$X \text{ compact} \implies X \text{ complet.}$$

En effet, si $(x_n)_{n\in\mathbb{N}}$ est une suite de Cauchy de X, alors par compacité séquentielle elle admet une valeur d'adhérence dans X, et par l'exercice 1.71 on conclut que la suite $(x_n)_{n\in\mathbb{N}}$ est convergente dans X.

1.5.3 Premiers exemples

On commence par la description des compacts de $\mathbb{R}$:

Théorème 1.103. *Les parties compactes de $(\mathbb{R}, |\cdot|)$ sont exactement les parties fermées et bornées de $\mathbb{R}$.*

Remarque 1.104. Rappelons qu'on a choisi de différencier les propriétés “être borné” et “être de diamètre fini”, mais dans le cadre présent où $(\mathbb{R}, |\cdot|)$ est un espace vectoriel normé, les deux notions coïncident.

Démonstration. On a vu à la proposition 1.101 que si $A \subset \mathbb{R}$ est compacte, alors A est fermée et bornée dans $\mathbb{R}$.

Réciproquement, soit $A \subset \mathbb{R}$ un ensemble fermé et borné. Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de A. Alors cette suite est bornée, donc par le théorème de Bolzano-Weierstrass 2.13 du Chapitre I, il existe φ une extraction et $x_\infty \in \mathbb{R}$ tels que $x_{\varphi(n)} \xrightarrow[n\to+\infty]{} x_\infty$. Mais comme A est fermée, $x_\infty \in A$, d'où le fait que A est séquentiellement compacte, donc compacte par le théorème 1.100. □

Exercice 1.105. On munit $\mathbb{R} \cup \{\pm\infty\}$ de la métrique de l'exercice 1.6. Montrer que $\mathbb{R} \cup \{\pm\infty\}$ est compact.

Un autre outil pour construire des ensembles compacts est le produit :

Proposition 1.106. *Tout produit fini d'espace métriques compacts est un compact pour la métrique produit.*

Démonstration. Exposons la preuve pour le produit de deux espaces (X, d_X) et (Y, d_Y). On pourra déduire le cas général soit en généralisant la preuve qui suit, soit en faisant une récurrence, qui utilisera le cas décrit ici.

Soit $(x_n, y_n)_{n\in\mathbb{N}}$ une suite d'éléments de $X \times Y$. On a vu à l'exercice 1.43 que la convergence d'une suite dans $X \times Y$ est équivalente à la convergence des suites coordonnées.

Appliquer la compacité de X et Y indépendamment ne fonctionne pas car cela donne des suites extraites $(x_{\varphi(n)})_{n\in\mathbb{N}}$ et $(y_{\psi(n)})_{n\in\mathbb{N}}$ dont on ne peut pas tirer grand chose, car il peut très bien n'y avoir aucun indice en commun dans ces extractions. On procède donc par extractions successives : par compacité de (X, d_X), il existe φ une extraction et $x_\infty \in X$ telle que $x_{\varphi(n)}$ converge vers x_∞ dans X. On considère désormais la suite $(y_{\varphi(n)})_{n\in\mathbb{N}}$ de Y, et par compacité de Y il existe ψ une extraction et $y_\infty \in Y$ telle que $y_{\varphi(\psi(n))}$ [21] converge vers y_∞.

Alors l'extraction $\varphi \circ \psi$ convient pour les deux coordonnées. C'est le cas pour $(y_{\varphi(\psi(n))})_{n\in\mathbb{N}}$ par construction, mais aussi pour $(x_{\varphi(\psi(n))})_{n\in\mathbb{N}}$ qui est une suite extraite de $(x_{\varphi(n)})_{n\in\mathbb{N}}$ donc converge également vers x_∞. □

1.5.4 Utilisation de la compacité

Nous allons reprendre des résultats importants du Chapitre I qui ont été énoncés pour des fonctions définies sur un segment $[a, b]$ de $\mathbb{R}$. On voit que ces résultats restent valables si on remplace ce segment par un espace métrique compact.

Théorème 1.107 (Généralisation du théorème des bornes)**.** *Soit (X_1, d_1) et (X_2, d_2) deux espaces métriques, et $f : X_1 \to X_2$. Si X_1 est compact et f est continue, alors $f(X_1)$ est un compact de X_2.*

En particulier si $(X_2, d_2) = (\mathbb{R}, |\cdot|)$ alors f est bornée et atteint ses bornes.

Démonstration. Soit $(y_n)_{n\in\mathbb{N}}$ une suite de $f(X_1)$. Par définition il existe une suite $(x_n)_{n\in\mathbb{N}}$ telle que $\forall n \in \mathbb{N}, y_n = f(x_n)$. Par compacité de X_1 on peut extraire une sous-suite $(x_{\varphi(n)})_{n\in\mathbb{N}}$ qui converge vers $x_\infty \in X_1$. Par continuité de f en x_∞, on a

$$y_{\varphi(n)} = f(x_{\varphi(n)}) \xrightarrow[n\to+\infty]{d_2} f(x_\infty) \in f(X_1)$$

donc $f(X_1)$ est séquentiellement compact.

Dans le cas où X_2 est l'espace des réels, on a $f(X_1) \subset \mathbb{R}$ compact donc borné (proposition 1.101), donc f est bien bornée[22]. Par propriété de la borne supérieure, il existe $(y_n)_{n\in\mathbb{N}} \in X_2^{\mathbb{N}}$ telle que $y_n \xrightarrow[n\to+\infty]{} \sup_{X_1} f$. Mais comme $f(X_1)$ est fermé (à nouveau par la proposition 1.101), on a bien $\sup_{X_1} f \in f(X_1)$, donc le supremum de f est atteint. Le même résultat vaut pour l'infimum. □

Exercice 1.108 (Distance à une partie)**.** Soit (X, d) un espace métrique, et A une partie compacte non vide de X.

1. Étant donné $x \in X$, montrer qu'il existe $a_0 \in A$ tel que
$$d(x, a_0) = d(x, A) \left(= \inf \left\{d(x, a),\ a \in A\right\}\right).$$
2. Montrer qu'il existe $(x_0, y_0) \in A^2$ tels que
$$\operatorname{diam}(A) = d(x_0, y_0).$$
3. Exhiber des contre-exemples aux questions précédentes quand A n'est pas compact.

Théorème 1.109 (Théorème de Heine)**.** *Soit (X_1, d_1) et (X_2, d_2) deux espaces métriques, $f : X_1 \to X_2$ une application continue. Si X_1 est compact, alors f est uniformément continue.*

---

21. C'est toujours un moment difficile que d'expliquer dans quel ordre les extractions doivent s'écrire. Je suggère le lecteur de prendre la règle suivante : "si je prends une suite extraite de $(u_n)_{n\in\mathbb{N}}$ via l'extraction φ, je dois remplacer n par $\varphi(n)$". Ainsi, si on prend une suite extraite de $(u_{\varphi(n)})_{n\in\mathbb{N}}$ via l'extraction ψ, il s'agit de remplacer n par $\psi(n)$ et donc on obtient bien $(u_{\varphi(\psi(n))})_{n\in\mathbb{N}}$.

22. Rappelons que cette étape peut être considérée comme redondante avec la suite car si les bornes sont atteintes, elles ne peuvent qu'être finies.

Démonstration. Les deux preuves du théorème 3.18 au Chapitre I peuvent être traduites sans grand changement dans ce cadre général. Reproduisons la preuve "topologique" :

Soit $\varepsilon > 0$. Par continuité de f, on a

$$\forall x \in X_1, \exists \alpha_x > 0, \quad \forall y \in B_{X_1}(x, \alpha_x), \quad d_2(f(y), f(x)) \leq \frac{\varepsilon}{2}. \tag{1.8}$$

On considère

$$X_1 \subset \bigcup_{x \in X_1} B_{X_1}\left(x, \frac{\alpha_x}{2}\right),$$

qui est un recouvrement de X_1 par des ouverts, donc par compacité, on peut en extraire un sous-recouvrement fini

$$X_1 \subset \bigcup_{i=1}^{N} B_{X_1}\left(x_i, \frac{\alpha_{x_i}}{2}\right) \tag{1.9}$$

où $N \in \mathbb{N}^*$ et $(x_1, \ldots, x_N) \in X_1^N$. On pose

$$\alpha = \frac{1}{2} \min\{\alpha_{x_i}\}_{i \in [\![1,N]\!]} > 0.$$

Soit $(x, y) \in X_1$ tels que $d_1(x, y) < \alpha$. Par (1.9), il existe $i \in [\![1, N]\!]$ tel que $x \in B_{X_1}\left(x_i, \frac{\alpha_{x_i}}{2}\right)$, et on constate que

$$d_1(y, x_i) \leq d(y, x) + d(x, x_i) < \alpha + \frac{\alpha_{x_i}}{2} \leq \alpha_{x_i}$$

et donc on peut appliquer (1.8) à la fois à x et à y qui sont dans $B_{X_1}(x_i, \alpha_{x_i})$, ce qui donne

$$d_2(f(y), f(x)) \leq d_2(f(y), f(x_i)) + d(f(x), f(x_i)) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

et achève la preuve. □

Remarque 1.110. Cette preuve est caractéristique de l'usage de la compacité d'un point de vue topologique. En fait, on dit parfois que la compacité est une généralisation du caractère fini d'un ensemble : ici par exemple, on a voulu échanger les quantificateurs "$\forall x \in X_1$" et "$\exists \alpha_x > 0$". Si X_1 avait été un ensemble fini, on aurait simplement posé $\alpha = \min\{\alpha_x, x \in X_1\}$, et on aurait pu voir que ce α est strictement positif et vaut pour tous les $x \in X_1$. Mais ici X_1 peut être infini, et une telle définition peut donner $\alpha = 0$ ce qui ne convient pas. Néanmoins grâce à la compacité de X_1, on a pu se ramener à un sous ensemble fini de X_1 pour faire l'échange de quantificateur.

1.6 Connexité

1.6.1 Définition et premières propriétés

Définition 1.111. Soit (X, d) un espace métrique.

— L'espace (X, d) est dit connexe si les seuls sous-ensembles de X qui sont à la fois fermés et ouverts sont l'ensemble vide $\emptyset$ et l'ensemble X lui-même[23].

— Si $Y \subset X$, on dira que Y est une partie connexe si (Y, d_Y) (muni de la métrique induite) est connexe.

Remarque 1.112. Avec l'exercice 1.27 on peut réexprimer la connexité d'une partie $A \subset X$ par :

$$\Big[A \subset O_1 \cup O_2 \text{ avec } O_1, O_2 \text{ des ouverts de } X \text{ tels que } O_1 \cap O_2 \cap A = \emptyset\Big] \quad \Rightarrow \quad O_1 \cap A = \emptyset \text{ ou } O_2 \cap A = \emptyset.$$

Remarque 1.113. La définition de la connexité étant formulée uniquement à partir des ouverts de X, c'est une notion topologique. Si $f : X \to Y$ est un homéomorphisme entre (X, d_X) et (Y, d_Y) et que (X, d_X) est connexe, alors (Y, d_Y) est connexe.

23. Par définition ces deux ensembles sont toujours ouverts et fermés, on demande donc ici que ce soient les seuls.

La définition précédente est sans doute celle qu'on utilise le plus fréquemment, mais elle ne correspond pas trop à l'intuition qu'on se fait de la connexité, qui consiste à être “en un seul morceau”. Il y a en fait plusieurs autres définitions équivalentes de la connexité :

Proposition 1.114. *Soit (X, d) métrique. Alors (X, d) est connexe si et seulement si l'une des propriétés suivantes est vérifiée :*

1. *Si X peut s'écrire comme réunion de deux ouverts disjoints, alors l'un de ces ouverts est vide*[24]*,*
2. *Si X peut s'écrire comme réunion de deux fermés disjoints, alors l'un de ces fermés est vide,*
3. *Toute fonction continue de X dans $\{0, 1\}$ est constante*[25]*.*

Démonstration. Supposons (X, d) connexe, et $X = O_1 \cup O_2$ où O_1, O_2 sont des ouverts disjoints de X. Comme $O_2 = X \setminus O_1$, O_2 est également fermé, donc $O_2 = X$ ou $O_2 = \emptyset$. Dans les deux cas, l'un des ouverts O_1, O_2 est vide.

Supposons 1. Alors si $X = F_1 \cup F_2$ où F_1, F_2 sont des fermés disjoints de X, alors $F_1 = X \setminus F_2$ et $F_2 = X \setminus F_1$ sont également des ouverts, donc par 1., $X \setminus F_1$ ou $X \setminus F_2$ est vide et alors l'autre est égal à X, et donc par passage au complémentaire, F_1 ou F_2 est vide.

Supposons 2. Soit $f : X \to \{0, 1\}$ continue. Alors $X = f^{-1}(\{0\}) \cup f^{-1}(\{1\})$ est une décomposition de X en deux fermés (par continuité de f et la proposition 1.59) disjoints, donc $f^{-1}(\{0\})$ ou $f^{-1}(\{1\})$ est vide, ce qui montre que $f^{-1}(\{0\})$ ou $f^{-1}(\{1\})$ est égal à X et donc que f est constante.

Supposons 3. Soit A un ensemble ouvert et fermé de X. On considère l'application $f = \mathbb{1}_A$ qui va de X dans $\{0, 1\}$. Pour montrer que f est continue, on va montrer que l'image réciproque d'un fermé de $\{0, 1\}$ est fermé. Comme $f^{-1}(\{1\}) = A$ est fermé et $f^{-1}(\{0\}) = X \setminus A$ est aussi fermé, f est en effet continue[26]. Donc f est constante, c'est-à-dire $A = X$ ou $A = \emptyset$. □

Exercice 1.115. On peut généraliser les premiers énoncés avec plus de deux ensembles : par exemple, montrer que si X est connexe et $(U_i)_{i \in I}$ est une partiton de X en ouverts (i.e. les ouverts sont deux à deux disjoints et leur union est X) alors tous les termes de la partition sauf un sont vides.

Exercice 1.116. Soit (X, d) un espace métrique.

1. Montrer que si $A \subset X$ est une partie connexe, alors $\overline{A}$ est connexe.
2. Plus généralement, montrer que si $A \subset X$ est connexe et B est telle que $A \subset B \subset \overline{A}$, alors B est connexe.
3. Montrer qu'on peut avoir $A \subset X$ connexe mais $\mathring{A}$ non connexe.
4. Si $(A_i)_{i \in I}$ sont des sous-ensembles connexes de X (avec I quelconque) et si $\bigcap_{i \in I} A_i \neq \emptyset$, alors $\bigcup_{i \in I} A_i$ est connexe.

1.6.2 Cas de $\mathbb{R}$

Théorème 1.117. *On munit $\mathbb{R}$ de sa métrique naturelle. Soit $A \subset \mathbb{R}$. Alors*

$$A \text{ est connexe} \iff A \text{ est un intervalle.}$$

Remarque 1.118. Rappelons qu'on a défini un intervalle comme un convexe de $\mathbb{R}$, et qu'on a vu au Chapitre I la liste des intervalles possibles.

Démonstration. **Une partie connexe est un intervalle :** Supposons que $A \subset \mathbb{R}$ n'est pas un intervalle : alors il existe $x \leq y$ deux éléments de A tels que $]x, y[\not\subset A$, c'est-à-dire qu'il existe $t \in]x, y[\setminus A$. Alors $A \subset]-\infty, t[\cup]t, +\infty[$ montre que A n'est pas connexe, puisque A rencontre chacun de ces ouverts disjoints de $\mathbb{R}$.

Un intervalle est connexe : Preuve 1 : Soit $a \leq b$ deux réels. Montrons que $[a, b]$ est connexe. On utilise la caractérsation 3. de la proposition 1.114 : soit donc $f : [a, b] \to \{0, 1\}$ continue. Supposons

24. Ainsi la seule écriture possible de X comme union de deux ouverts disjoints et l'écriture triviale $X = X \cup \emptyset$.

25. Ici $\{0, 1\}$ est muni de la métrique induite par $\mathbb{R}$; il s'agit d'un espace discret. D'ailleurs, on peut remplacer $\{0, 1\}$ par n'importe quel espace discret.

26. Il y a 4 sous-ensembles (fermés) de $\{0, 1\}$, mais $\emptyset$ et $\{0, 1\}$ ont forcément une image réciproque qui est fermée.

par exemple $f(b) = 1$, le lecteur pourra faire l'autre cas possible. Par l'absurde, si f n'est pas constante, l'ensemble des points ou f est nulle est non vide, et donc on peut poser

$$c = \sup\{x \in [a, b], \ \ f(x) = 0\}.$$

Par définition de la borne supérieure (voir exercice 2.18 au Chapitre I) il existe $(x_n)_{n\in\mathbb{N}}$ une suite de $[a, b]$ telle que $\forall n \in \mathbb{N}, f(x_n) = 0$ et $(x_n)_{n\in\mathbb{N}}$ converge vers c. Par continuité de f, on a $f(c) = 0$.

Par continuité de f en b et $f(b) = 1$, on voit que f n'est pas nulle sur un voisinage de b et donc $c < b$, mais alors

$$\forall x \in]c, b], f(x) = 1$$

ce qui donne en prenant la limite à droite en c que $f(c) = 1$, ce qui constitue une contradiction. Ainsi $[a, b]$ est connexe.

Pour conclure le cas général, on affirme que tout intervalle est une union de segments qui ont tous un point en commun, ce qui permettra d'appliquer la question 4 de l'exercice 1.116. En effet

$$\forall (a, b) \in \mathbb{R}^2 \text{ tels que } a < b, \qquad]a, b] = \bigcup_{n\in\mathbb{N}^*} \left[a + \frac{1}{n}, b\right], \ \ [a, b[= \bigcup_{n\in\mathbb{N}^*} \left[a, b - \frac{1}{n}\right],$$
$$]a, b[= \bigcup_{n\in\mathbb{N}^*} \left[a + \frac{1}{n}, b - \frac{1}{n}\right]$$

où dans la troisième écriture il faut considérer que les premiers termes peuvent être vides (si $b-a < 2$), mais que tous les termes non vides vont contenir le milieu $\frac{a+b}{2}$.

On laisse le lecteur écrire le cas des intervalles non bornés.

Un intervalle est connexe : Preuve 2 : * Soit $a < b$ avec $(a, b) \in \mathbb{R} \cup \{\pm\infty\}$. Soit $A \subset]a, b[$ qui est ouverte, fermée et non vide de $]a, b[$. Il existe $x \in A$: on pose

$$C = \Big\{y \in [x, b[, \ [x, y] \subset A\Big\}, \quad \beta = \sup C$$

ce qui est possible car C est non vide (il contient x). Supposons par l'absurde $\beta < b$. Alors β en tant que borne supérieure de C est limite d'éléments de C, qui sont nécessairement dans A, et comme A est fermée de $]a, b[$ [27], $\beta \in A$. De plus, comme A est ouvert dans l'ouvert $]a, b[$ il existe $\varepsilon > 0$ tel que $]\beta-\varepsilon, \beta+\varepsilon[\subset A$ ce qui contredit la définition de β. Ainsi $\beta = b$ et donc $[x_0, b[\subset A$. De la même manière on montre $]a, x_0] \subset A$, ce qui conclut à la connexité de $]a, b[$.

Pour conclure dans le cas général, on considère I un intervalle semi-ouvert ou fermé : son intérieur est un intervalle ouvert $]a, b[$, donc par le cas précédent il est connexe. Mais alors $]a, b[\subset I \subset \overline{[a, b]}$ donc est aussi connexe par la question 2 de l'exercice 1.116. □

Remarque 1.119. Vous aurez sans doute remarqué que la première preuve ressemble beaucoup à la preuve du théorème des valeurs intermédiaires du Chapitre I. Et c'est bien naturel, puisqu'on peut retrouver directement le théorème des valeurs intermédiaires comme cas particulier du théorème 1.124 ci-après : en effet, si $f : I \to \mathbb{R}$ est continue et I est un intervalle, alors par le théorème 1.117, I est un connexe, par le théorème 1.124 $f(I)$ est un connexe, et à nouveau par le théorème 1.117, $f(I)$ est un intervalle.

Attention par contre à ne pas créer une boucle dans vos raisonnements : si par exemple vous dites qu'un intervalle de $\mathbb{R}$ est connexe car il est connexe par arcs ou parce qu'il est convexe, vous tournez en rond, car pour montrer "connexe par arcs $\Rightarrow$ connexe" on aura besoin de la connexité des intervalles !

Exercice 1.120. Montrer que $\mathbb{R}$ n'est pas homéomorphe à $\mathbb{R}^n$ (on peut munir $\mathbb{R}^n$ de la métrique produit, ou d'une norme comme on le verra au paragraphe 2).

27. Attention on utilise ici $\beta < b$ car il faut vor que β est limite d'une suite convergente de $]a, b[$, donc si cette limite était b, ce ne serait pas le cas.

1.6.3 Composantes connexes

Définition-Proposition 1.121. Soit (X,d) un espace métrique, et $x \in X$. La réunion de tous les sous-ensembles de X qui sont connexes et qui contiennent x est un ensemble connexe de X qu'on appelle la composante connexe de x dans X. C'est la plus grande partie connexe contenant x, et c'est un fermé de X.

Démonstration. On applique la question 4 de l'exercice 1.116 pour voir que l'ensemble C union des ensembles connexes qui contiennent tous x est connexe. Mais d'après la question 1 de ce même exercice l'adhérence de C est aussi un connexe qui contient aussi x, donc finalement $\overline{C} = C$ et C est fermé. □

Exercice 1.122. Étant donné (X,d) un espace métrique.

1. Pour $(x,y) \in X^2$, on définit la relation

$$x \mathscr{R} y \quad \Leftrightarrow \quad x \text{ et } y \text{ sont dans la même composante connexe de } X.$$

   Montrer que $\mathscr{R}$ est une relation d'équivalence sur X.
2. Identifier les classes d'équivalence pour cette relation.

Exercice 1.123. Identifier les composantes connexes de $(\mathbb{Q}, |\cdot|)$.

1.6.4 Image continue d'un connexe

On trouve ici le théorème le plus célèbre lié à la connexité :

Théorème 1.124. *Soit (X_1,d_1) et (X_2,d_2) deux espaces métriques. Si X_1 est connexe et $f : X_1 \to X_2$ est continue, alors $f(X_1)$ est connexe.*

Démonstration. Soit $g : f(X_1) \to \{0,1\}$ une fonction continue. Alors on peut définir $g \circ f : X_1 \to \{0,1\}$, qui est aussi continue, donc pas connexité de X_1, $g \circ f$ est constante. Ainsi il existe $\alpha \in \{0,1\}$ tel que $\forall x \in X_1, g(f(x)) = \alpha$, ce qui montre que g est constante sur $f(X_1)$ et achève la preuve. □

1.6.5 Connexité par arcs

On définit une autre notion, également intuitive, en apparence plus complexe à définir, mais en réalité plus facile à prouver et qui va permettre de fournir simplement des exemples d'ensembles connexes :

Définition 1.125. Soit (X,d) un espace métrique.

— Étant donnés x et y deux éléments de X, on dit qu'on peut joindre x et y par un arc/chemin dans X s'il existe une application $\gamma : [0,1] \to X$ continue et telle que $\gamma(0) = x$ et $\gamma(1) = y$.

— L'espace (X,d) est dit connexe par arcs si pour tout $(x,y) \in X^2$, on peut joindre x et y dans X.

Proposition 1.126. *Soit (X,d) un espace connexe par arcs. Alors (X,d) est connexe.*

Remarque 1.127. Même s'il existe des ensembles connexes qui ne sont pas connexes par arcs, en pratique cela reste un phénomène assez rare réservé à des ensembles assez pathologiques. Donc il est très fréquent d'utiliser le résultat précédent pour prouver qu'un ensemble est connexe. Voir aussi la proposition 2.21 pour voir une classe d'ensembles pour lesquels les notions de connexité et de connexité par arcs sont équivalentes.

Démonstration. Si X est vide, le résultat est clair. Supposons X non vide : alors il existe $x_0 \in X$. Pour tout $y \in X$, il existe γ_y un chemin entre x_0 et y tracé dans X. Par les théorèmes 1.117 et 1.124, les ensembles $\gamma_y([0,1])$ sont connexes pour tout $y \in X$, mais alors on a l'égalité

$$X = \bigcup_{y \in X} \gamma_y([0,1])$$

donc X est l'union de connexes contenant tous x_0, donc en appliquant la question 4 de l'exercice 1.116, X est bien connexe. □

Exercice 1.128. Montrer que $\mathbb{R}^2 \setminus \mathbb{Q}^2$ est connexe par arcs.

1.6.6 Utilisations classiques de la connexité

Une utilisation typique de la connexité est le passage du local au global : si on sait qu'une propriété est vraie sur le voisinage de chaque point, on peut espérer par connexité que la propriété soit vraie sur l'espace entier. On donne à titre d'exemple :

Exercice 1.129. Soit (X_1, d_1), (X_2, d_2) deux espaces métriques et $f : X_1 \to X_2$. On suppose que pour chaque $x \in X_1$, il existe un voisinage de x sur lequel f est constante. On veut montrer que f est constante. Le cas $X_1 = \emptyset$ étant trivial, on suppose que $X_1 \neq \emptyset$, donc il existe $x_0 \in X_1$, et on pose

$$A = \Big\{x \in X_1,\ \ f(x) = f(x_0)\Big\}.$$

1. Montrer que A est un ouvert de X_1.
2. Montrer que A est un fermé de X_1 et conclure.

On renvoit aux Chapitres II, VII et X pour des utilisations de ce type de la connexité.

1.6.7 Exemples d'ensembles connexes

Exercice 1.130 (Des ensembles connexes mais non connexes par arcs)**.** Dans cet exercice on propose des exemples d'ensembles connexes qui ne sont pas connexes par arcs :

1. On pose
$$A = \left\{\left(x, \sin\left(\frac{1}{x}\right)\right), x > 0\right\}.$$
   (a) Montrer que A est connexe. Alors $\overline{A}$ est aussi connexe, par l'exercice 1.116.
   (b) Montrer que $\overline{A} = A \cup (\{0\} \times [-1, 1])$.
   (c) (*)Montrer que $\overline{A}$ n'est pas connexe par arcs. [28]
2. De la même manière, on peut montrer que si $f : \mathbb{R} \to \mathbb{R}$ est continue, strictement croissante et telle que $\lim_{-\infty} f = 0$, $\lim_{+\infty} f = 1$, alors l'adhérence de l'ensemble en forme de "spirale"
$$\Big\{f(\theta)e^{i\theta}, \theta \in \mathbb{R}\Big\},$$
est connexe mais non connexe par arcs, voir [Mar02, Page 121].
3. (** Voir [Gou08b, Page 44]) Montrer que l'ensemble
$$\left(\bigcup_{x\in\mathbb{Q}} \{x\} \times \mathbb{R}_+\right) \cup \left(\bigcup_{x\in\mathbb{R}\setminus\mathbb{Q}} \{x\} \times \mathbb{R}_-^*\right)$$
est connexe mais non connexe par arcs.
4. On considère
$$\Big(\{0\} \times \{0, 1\}\Big) \cup \left(\bigcup_{n\in\mathbb{N}^*} \left\{\frac{1}{n}\right\} \times [0, 1]\right) \cup \Big([0, 1] \times \{0\}\Big).$$
Montrer que cet ensemble est connexe mais non connexe par arcs.

Exercice 1.131. 1. Montrer que $GL_n(\mathbb{C})$ est connexe par arcs.

2. Le cas réel est plus élaboré :
   (a) Montrer que $GL_n(\mathbb{R})$ n'est pas connexe.
   (b) (*)Montrer que
$$GL_n^+(\mathbb{R}) = \Big\{M \in GL_n(\mathbb{R}),\ \ \det(M) > 0\Big\} \quad \text{et} \quad GL_n^-(\mathbb{R}) = \Big\{M \in GL_n(\mathbb{R}),\ \ \det(M) < 0\Big\}$$
   sont les composantes connexes de $GL_n(\mathbb{R})$.
   On trouvera ces exemples et bien d'autres dans [MT86].

28. On pourra considérer par l'absurde l'existence d'un chemin $\gamma = (\gamma_1, \gamma_2)$ entre $(0, 0)$ et $(1, \sin(1))$ et introduire $t_0 = \sup\{t > 0, \gamma_1(t) > 0\}$ et raisonner au voisinage de t_0.

Exercice 1.132 (Produit de connexes/connexes par arcs)**.** Soit (X_1, d_1) et (X_2, d_2) deux espaces métriques, et $X_1 \times X_2$ muni de la métrique produit.

1. (*)Voir par exemple [Mar09, page 45]. On suppose X_1 et X_2 connexes, et on veut montrer que $X_1 \times X_2$ est connexe. Le cas où l'un des espaces est vide étant trivial, on suppose qu'il existe $(a_1, a_2) \in X_1 \times X_2$.
   (a) Étant donné $x = (x_1, x_2) \in X_1 \times X_2$, montrer que
   $$A_x = \big(X_1 \times \{x_2\}\big) \cup \big(\{a_1 \times X_2\}\big)$$
   est connexe.
   (b) En déduire que $X_1 \times X_2$ est connexe.
2. Montrer que si X_1 et X_2 sont connexes par arcs, alors $X_1 \times X_2$ est connexe par arcs.

2 Espaces vectoriels normés (sur $\mathbb{R}$ ou $\mathbb{C}$)

Dans ce paragraphe, nous décrivons le cas important des espaces vectoriels munis d'une norme, qui sont des cas particuliers d'espace métrique. Toutes les notions du paragraphe 1 s'appliquent donc à ce cas, mais nous décrivons ici les propriétés supplémentaires satisfaites par ce type d'espace. Notez qu'au niveau de l'agrégation, la plupart des exemples et applications issues des espaces métriques rentrent en fait dans ce cadre des espaces vectoriels normés. Pour cette raison, on a donné peu d'exemples "importants" au paragraphe 1, les exemples dépassant le plus souvent le programme du concours ; il n'empêche que le vocabulaire des espaces métriques est bien au cœur du programme, et on invite le lecteur à savoir jongler sans difficulté entre les définitions du cadre métrique et du cadre des espaces vectoriels normés. On l'invite également à réfléchir aux espaces métriques qu'il connait et qui ne sont pas des espaces vectoriels normés (y compris des espaces vectoriels métriques qui ne sont pas des espaces vectoriels normés) [29].

2.1 Définitions

Définition 2.1. Soit E un espace vectoriel normé sur le corps $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. On appelle norme sur E une application $\|\cdot\| : E \to \mathbb{R}_+$ telle que

1. $\forall x \in E,\ \ \|x\| = 0 \ \Leftrightarrow\ x = 0$,
2. $\forall x \in E, \forall \lambda \in \mathbb{K},\ \|\lambda x\| = |\lambda|\|x\|$ (homogénéité),
3. $\forall (x, y) \in E^2,\ \ \|x + y\| \leq \|x\| + \|y\|$ (inégalité triangulaire/sous-additivité).

On dit que $(E, \|\cdot\|)$ est un espace vectoriel normé.

Remarque 2.2 (**)**.** On ne parlera que d'espaces vectoriels sur les corps $\mathbb{R}$ ou $\mathbb{C}$. Il faut en tout cas que ce soit un corps valué (i.e. muni d'une valeur absolue [30]), complet, non discret ; au niveau de l'agrégation, il est délicat d'aller étudier d'autres cas.

Remarque 2.3 (Métrique associée à une norme)**.** Étant donné $(E, \|\cdot\|)$ un espace vectoriel normé, on montre facilement que
$$\forall (x, y) \in E^2, \quad d(x, y) := \|x - y\|$$
définit une métrique sur E. On peut donc utiliser les définitions et résultats du paragraphe 1 aux espaces vectoriels normés.

29. On invite juste en premier lieu à se poser la question ; les exemples pertinents d'espace (vectoriel) métrique non normables sont aux limites du programme (on renvoit au paragraphe 4.4 pour des exemples intéressants et accessibles), et il peut suffire d'en avoir conscience ; il est juste malvenu de se rendre compte par une question du jury qu'on a fait un plan ou une partie de plan sans exemple pertinent.

30. Une valeur absolue sur un corps $\mathbb{K}$ est une application $|\cdot| : \mathbb{K} \to \mathbb{R}$ tel que
   1. $\forall x \in \mathbb{K},\ |x| = 0 \ \Leftrightarrow\ x = 0$,
   2. $\forall (x, y) \in \mathbb{K},\ |x + y| \leq |x| + |y|$,
   3. $\forall (x, y) \in \mathbb{K},\ |x.y| = |x|.|y|$.

À ce sujet on peut citer le théorème d'Ostrowski qui classifie les valeurs absolues sur $\mathbb{Q}$

Exemple 2.4. On peut bien sûr citer $\mathbb{R}$ muni de la valeur absolue, qui est un $\mathbb{R}$-espace vectoriel normé. De même $\mathbb{C}$ muni du module est un $\mathbb{C}$-espace vectoriel normé. Pour $n \in \mathbb{N}$, on peut également définir une norme sur $\mathbb{R}^n$, qu'on choisira par défaut et qu'on appelle norme euclidienne :

$$\forall x \in \mathbb{R}^n, \quad \|x\|_2 = \sqrt{\sum_{k=1}^{n} x_k^2}.$$

Le fait que cela définit bien une norme n'est pas tout-à-fait élémentaire. On peut le montrer en utilisant l'inégalité de Cauchy-Schwarz :

$$\forall (x,y) \in (\mathbb{R}^n)^2, \quad \sum_{k=1}^{n} |x_k y_k| \leq \sqrt{\sum_{k=1}^{n} x_k^2}\sqrt{\sum_{k=1}^{n} y_k^2}.$$

On ne donne pas les détails, car ceci sera revu au Chapitre IX dans le cadre plus général des espaces préhilbertien : en effet il existe un produit scalaire $\langle \cdot, \cdot \rangle$ sur $\mathbb{R}^n$ tel que $\|x\|_2 = \sqrt{\langle x, x \rangle}$ est la norme associée à ce produit scalaire. Aussi, on peut voir cette norme comme un cas particulier de l'exemple 2.8 ci-dessous.

Sur $\mathbb{C}^n$, on définit de même la norme

$$\forall x \in \mathbb{C}^n, \quad \|x\|_2 = \sqrt{\sum_{k=1}^{n} |x_k|^2}.$$

Exercice 2.5. Soit $(E, \|\cdot\|)$ un espace vectoriel normé, et F un s.e.v. de E.

1. Montrer que $\overline{F}$ est un s.e.v. de E.
2. Donner un exemple où $\overline{F} \neq F$ (il faudra choisir E également).
   Remarquons qu'on verra au corollaire 2.28 qu'une telle situation ne peut se produire que si F est de dimension infinie.
3. Montrer l'équivalence
$$\mathring{F} \neq \emptyset \Leftrightarrow F = E.$$

Définition 2.6. Étant donné un espace vectoriel E, et N_1, N_2 deux normes sur E, on dit que N_1 et N_2 sont équivalentes s'il existe deux réels $c > 0$ et $C > 0$ tels que

$$\forall x \in E, \quad c.N_1(x) \leq N_2(x) \leq C.N_1(x).$$

Exercice 2.7 (*Équivalence)**.** Montrer que si E est un espace vectoriel normé muni de deux normes N_1 et N_2, alors

$$N_1 \text{ et } N_2 \text{ sont équivalentes} \iff \text{l'application } \mathrm{Id} : (E, N_1) \to (E, N_2) \text{ est un homéomorphisme.}$$

Ainsi les différentes notions d'équivalence introduites à l'exercice 1.65 sont les mêmes dans le cas des espaces vectoriels normés.

Exemple 2.8 (Norme $\|\cdot\|_p$)**.** Ici $\mathbb{K}$ est $\mathbb{R}$ ou $\mathbb{C}$. Avec les résultats de l'exercice 6.34 du Chapitre I (voir aussi l'exercice 2.9 ci-dessous), on montre que si $p \in [1, \infty[$, la fonction définie par

$$\forall x \in \mathbb{K}^n, \quad \|x\|_p = \left(\sum_{k=1}^{n} |x_k|^p\right)^{\frac{1}{p}},$$

est une norme sur $\mathbb{K}^n$. La partie délicate est de prouver l'inégalité triangulaire quand $p \in]1, +\infty[$; quand $p = 1$ l'inégalité est facile.

Pour $p = +\infty$, on peut également définir

$$\forall x \in \mathbb{K}^n, \quad \|x\|_\infty = \max\left\{|x_1|, |x_2|, \dots, |x_n|\right\}$$

qui est aussi une norme sur $\mathbb{K}^n$.

Exercice 2.9. Ici $\mathbb{K}$ est $\mathbb{R}$ ou $\mathbb{C}$.

1. Dans cette question, on donne une preuve un peu plus directe de l'inégalité triangulaire pour $\|\cdot\|_p$ sans passer par l'inégalité de Hölder comme on l'a fait à l'exercice 6.34 du Chapitre I. Soit donc $(x,y) \in \mathbb{K}^n$ et $p \in [1,+\infty[$.
   (a) Montrer que
   $$\forall \lambda \in [0,1],\ \forall (u,v) \in \mathbb{R}_+^2,\ \big(\lambda u + (1-\lambda)v\big)^p \leq \lambda u^p + (1-\lambda)v^p.$$
   (b) On suppose $\|x\|_p = \|y\|_p = 1$. Montrer
   $$\forall \lambda \in [0,1], \sum_{i=1}^n \big(\lambda |x_i| + (1-\lambda)|y_i|\big)^p \leq 1.$$
   (c) Conclure que dans le cas général
   $$\|x+y\|_p \leq \|x\|_p + \|y\|_p.$$
2. Montrer que si $n \in \mathbb{N}$ et $p \in [1,\infty]$, alors
   $$\forall x \in \mathbb{K}^n, \quad \|x\|_\infty \leq \|x\|_p \leq n^{\frac{1}{p}} \|x\|_\infty$$
   et en déduire que les normes $\|\cdot\|_p$ sont toutes équivalentes sur $\mathbb{K}^n$.
3. Plus généralement, montrer
   $$\forall 1 \leq p \leq q \leq \infty,\ \ \forall x \in \mathbb{K}^n, \quad \|x\|_q \leq \|x\|_p \leq n^{\frac{1}{p}-\frac{1}{q}} \|x\|_q.$$
   En particulier à x fixé, l'application $p \mapsto \|x\|_p$ est décroissante.
4. Montrer que pour tout $x \in \mathbb{K}^n$,
   $$\lim_{p \to \infty} \|x\|_p = \|x\|_\infty.$$

Remarque 2.10. Comme dans l'exercice 1.9, si $(E_1, \|\cdot\|_1)$ et $(E_2, \|\cdot\|_2)$ sont deux espaces vectoriels normés, on peut s'intéresser au produit $E_1 \times E_2$, qu'on peut munir des normes

$$\forall (x,y) \in E_1 \times E_2, \quad \|(x,y)\|_{E_1\times E_2} = \|x\|_{E_1} + \|y\|_{E_2}, \quad \text{ou encore} \quad \|(x,y)\|_{E_1\times E_2} = \max\left\{\|x\|_{E_1}, \|y\|_{E_2}\right\}$$

qui sont des normes équivalentes sur $E_1 \times E_2$. L'une ou l'autre de ces normes peut être appelée norme produit, et on considèrera par défaut qu'un espace produit est muni d'une telle norme. On peut bien sûr itérer le procéder. Dans le cas de $\mathbb{R}^n$ vu comme le produit de n copies de $(\mathbb{R}, |\cdot|)$, on obtient en itérant la première formule la norme $\|x\| = \sum_{i=1}^n |x_i|$, et en itérant la seconde formule $\|x\| = \max\{|x_i|, i \in [\![1,n]\!]\}$ soit les normes 1 et ∞ respectivement.

Exemple 2.11. On a évoqué au Chapitre I la convergence uniforme, voir la remarque 4.4 du Chapitre I. Le cadre bien adapté aux espaces vectoriels normés est de travailler l'espace des fonctions bornées : si X est un ensemble, on pose

$$B(X,\mathbb{R}) = \{f : X \to \mathbb{R},\ \ f \text{ est bornée sur } X\} = \{f : X \to \mathbb{R},\ \exists M \in \mathbb{R},\ \forall x \in X, |f(x)| \leq M\}$$

qu'on peut munir de

$$\forall f \in B(X,\mathbb{R}), \quad \|f\|_\infty = \sup_{x\in X} |f(x)|,$$

qui est bien une norme [31] et qu'on appelle "norme uniforme" ou "norme infinie".

Ici on peut remplacer l'espace d'arrivée $(\mathbb{R}, |\cdot|)$ par un espace vectoriel $(E, \|\cdot\|_E)$, et définir

$$\forall f \in B(X,E), \quad \|f\|_\infty = \sup_{x\in X} \|f(x)\|_E.$$

31. Ce n'est pas difficile à vérifier, mais on laisse le lecteur le faire, et faire attention à l'importance d'avoir considéré des fonctions bornées.

Aussi, si (X,d) est un espace métrique, on peut considérer $C_b^0(X) = C_b^0(X,\mathbb{R})$ (ou de même $C_b^0(X,E)$) le sous-espace de $B(X,\mathbb{R})$ des fonctions continues et bornées, qu'on peut munir de la norme infinie.

Enfin, si (X,d) est un espace métrique compact, alors les fonctions continues sont bornées, donc on peut munir $C^0(X,\mathbb{R})$ de la norme infinie. Par exemple, si $a<b$ sont deux réels, $(C^0([a,b]), \|\cdot\|_\infty)$ est un espace vectoriel normé.

Exercice 2.12 (Normes sur les polynômes)**.** On définit sur $\mathbb{R}[X]$:[32]

$$\forall P = \sum_{k=0}^{\infty} a_k X^k \in \mathbb{R}[X], \qquad N_1(P) = \sup_{[0,1]} |P(x)|, \qquad N_2(P) = \int_0^1 |P(t)|dt,$$
$$N_3(P) = \max\{|a_k|, k \in \mathbb{N}\}, \qquad N_4(P) = \sum_{k\in\mathbb{N}} |a_k|$$

1. Montrer que N_i pour $i \in [\![1,4]\!]$ est une norme sur $\mathbb{R}[X]$.
2. Étudier l'équivalence de ces normes[33].

2.2 Complétude

Définition 2.13. Soit $(E, \|\cdot\|)$ un espace vectoriel normé. Si E est complet pour la métrique associée à $\|\cdot\|$, on dit que $(E,\|\cdot\|)$ est un espace de Banach.

Exemple 2.14. L'espace $(\mathbb{R}, |\cdot|)$ est un espace de Banach.

Remarque 2.15. Si un espace vectoriel est muni de deux normes équivalentes, il est complet pour l'une si et seulement s'il est complet pour l'autre.

Exercice 2.16 (Exemples d'espaces de Banach)**.** [34]

1. (a) Pour $n \in \mathbb{N}^*$, on peut munir $\mathbb{R}^n$ de la norme $\|\cdot\|_\infty$ définie à l'exercice 2.9. Montrer que $(\mathbb{R}^n, \|\cdot\|_\infty)$ est un espace de Banach.
   (b) Pour $p \in [1,+\infty[$ et $n \in \mathbb{N}^*$, montrer que $(\mathbb{R}^n, \|\cdot\|_p)$ est un espace de Banach.
   (c) Montrer que $(\mathbb{C}, |\cdot|)$ et plus généralement $(\mathbb{C}, \|\cdot\|_p)$ pour $p \in [1,+\infty]$ sont des espaces de Banach.
2. Soit X un ensemble et E un espace de Banach.
   (a) Montrer que $(B(X,E), \|\cdot\|_\infty)$ (voir l'exercice 2.11) est un espace de Banach.[35]
   (b) On suppose de plus que X est muni d'une métrique. Montrer que $(C_b^0(X,E), \|\cdot\|_\infty)$ est un espace de Banach.
3. On pose pour $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$,

$$\ell^\infty(\mathbb{N}) = \Big\{(x_n)_{n\in\mathbb{N}} \in \mathbb{K}^\mathbb{N},\ (x_n)_{n\in\mathbb{N}} \text{ est bornée}\Big\}, \quad \forall (x_n) \in \mathbb{K}^\mathbb{N},\ \|x\|_\infty = \sup_{n\in\mathbb{N}} |x_n|.$$

32. Si $P \in \mathbb{R}[X]$, on peut toujours écrire $P = \sum_{k\in\mathbb{N}} a_k X^k$ où la suite $(a_k)_{k\in\mathbb{N}}$ est supposée stationnaire à 0, ce qui fait qu'il s'agit d'une somme finie et non infinie.
33. Attention en particulier à la notation $\|\cdot\|_\infty$ quand on parle d'un polynôme ; s'agira-t-il de N_1 ou de N_3 ?
34. Le lecteur peut retenir la décomposition des exemples proposée ici en 3 catégorie :

— Espaces de dimension finie,

— Espaces de fonctions,

— Espaces de suites.

Le cas des suites est souvent négligé alors qu'il permet de fournir des exemples intéressants sans les éléments plus complexes amenés par les fonctions. Notons aussi que cette séparation est parfois artificielle ; en effet on sait qu'une suite est elle-même une fonction dont l'espace de départ est $\mathbb{N}$; d'ailleurs le lecteur attentif aura remarqué que $\ell^\infty(\mathbb{N}) = B(\mathbb{N},\mathbb{K})$. Il n'empêche que la décomposition reste commode car l'étude des espaces de fonctions suggère l'introduction de notions comme la continuité, la dérivabilité ou la mesurabilité des fonctions mises en jeu, qui sont soit triviales soit n'ont pas de sens pour les suites.

35. C'est l'occasion de retravailler la preuve de la proposition 4.7 du Chapitre I. Attention, il y a une étape de plus dans la preuve, à savoir montrer que la limite construite est bien une fonction bornée.

(a) Montrer que $(\ell^\infty(\mathbb{N}), \|\cdot\|_\infty)$ est un espace vectoriel normé, et même un espace de Banach.

(b) On considère
$$c_0(\mathbb{N}) = \left\{(x_n)_{n\in\mathbb{N}} \in \mathbb{K}^{\mathbb{N}}, \ \ x_n \xrightarrow[n\to+\infty]{} 0\right\}.$$
Montrer que $(c_0(\mathbb{N}))$ est un sous-espace vectoriel de $\ell^\infty(\mathbb{N})$ et que si on le munit de la norme $\|\cdot\|_\infty$, il s'agit également d'un espace de Banach.

Exercice 2.17. Soit $E = C^1([0,1])$ l'ensemble des fonctions $f : [0,1] \to \mathbb{R}$ de classe C^1. On définit
$$\forall f \in E, \quad \|f\|_{C^1} := \|f\|_\infty + \|f'\|_\infty.$$

1. Montrer que $\|\cdot\|_{C^1}$ est une norme.
2. Soit $(f_n)_{n\in\mathbb{N}}$ une suite d'éléments de E, et $f \in E$. Montrer que $(f_n)_{n\in\mathbb{N}}$ converge vers f au sens de la norme C^1 si et seulement si $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers f et $(f'_n)_{n\in\mathbb{N}}$ converge uniformément vers f'.
3. Montrer que $(C^1([0,1]), \|\cdot\|_{C^1})$ est un espace de Banach.

On revient ici sur une application de la complétude déjà abordée deux fois au Chapitre I et qu'on reproduit ici dans un cadre général. On voit également que la propriété citée caractérise la complétude, ce qui sera utilisé au Chapitre IV.

Proposition 2.18 (Caractérisation de la complétude par les séries absolument convergentes)**.** *Soit $(E, \|\cdot\|)$ un espace vectoriel normé. Alors*
$$E \text{ est complet} \quad \Longleftrightarrow \quad \begin{cases} \text{Pour toute suite } (x_n)_{n\in\mathbb{N}} \in E^{\mathbb{N}}, \\ \displaystyle\sum_{n\in\mathbb{N}} \|x_n\| < +\infty \ \Rightarrow \ \text{la série de terme général } (x_n)_{n\in\mathbb{N}} \text{ converge dans } E. \end{cases}$$

Remarque 2.19. Rappelons que le fait que la série de terme général $(x_n)_{n\in\mathbb{N}} \in E^{\mathbb{N}}$ converge dans E signifie qu'il existe $S \in E$ tel que
$$\left\|\sum_{k=0}^{n} x_k - S\right\| \xrightarrow[n\to+\infty]{} 0.$$
Pour qualifier $\sum_{n\in\mathbb{N}} \|x_n\| < +\infty$ on pourrait parler de convergence "absolue", mais au vu des exemples cités ci-après, cela peut prêter à confusion.

Démonstration. $\Rightarrow$ On suppose E complet. La preuve est la même que les version "complétude" des cas particuliers décrits à l'exemple 2.20 : Pour montrer que $(S_n = \sum_{k=0}^n x_k)_{n\in\mathbb{N}}$ est convergente dans E, montrons que cette suite est de Cauchy. Soit $\varepsilon > 0$: étant convergente, la suite $(\Sigma_n = \sum_{k=0}^n \|x_k\|)_{n\in\mathbb{N}}$ est également de Cauchy, donc il existe $n_0 \in \mathbb{N}$ tel que
$$\forall q \geq p \geq n_0, \quad \|\Sigma_q - \Sigma_p\| \leq \varepsilon.$$
On en déduit ainsi que
$$\forall q \geq p \geq n_0, \|S_q - S_p\| = \left\|\sum_{k=p+1}^{q} x_k\right\| \leq \sum_{k=p+1}^{q} \|u_k\| = \Sigma_q - \Sigma_p \leq \varepsilon$$
donc on a montré que la suite $(S_n)_{n\in\mathbb{N}}$ était de Cauchy dans E. Par complétude de E, la suite $(S_n)_{n\in\mathbb{N}}$ est bien convergente.

$\Leftarrow$ Soit $(x_n)_{n\in\mathbb{N}}$ une suite de Cauchy dans E. Le critère de Cauchy pour $\varepsilon = 1$ donne l'existence de $n_0 \in \mathbb{N}$ tel que
$$\forall n \geq n_0, \quad \|x_n - x_{n_0}\| \leq 1.$$
Pour $\varepsilon = \frac{1}{2}$, on obtient l'existence de n_1 tel que
$$\forall n \geq n_1, \quad \|x_n - x_{n_1}\| \leq \frac{1}{2}.$$

Remarquons qu'on peut choisir $n_1 > n_0$ puisque la propriété précédente reste valable si on remplace n_1 par un entier supérieur.

On poursuit la construction récurrente : si $k \in \mathbb{N}$ et $n_0 < n_1 < \ldots < n_k$ sont construits tels que

$$\forall i \in [\![0,k]\!],\ \ \forall n \geq n_i,\ \ \|x_n - x_{n_i}\| \leq \frac{1}{2^i},$$

on applique le critère de Cauchy pour $\varepsilon = \frac{1}{2^{k+1}}$ ce qui donne l'existence de $n_{k+1} \in \mathbb{N}$ qu'on peut choisir tel que $n_{k+1} > n_k$ et tel que

$$\forall n \geq n_{k+1},\ \ \|x_n - x_{n_{k+1}}\| \leq \frac{1}{2^{k+1}},$$

ce qui achève la construction de $(n_k)_{k\in\mathbb{N}}$. En particulier on a

$$\forall k \in \mathbb{N},\ \ \|x_{n_{k+1}} - x_{n_k}\| \leq \frac{1}{2^k}$$

donc

$$\sum_{k\in\mathbb{N}} \|x_{n_{k+1}} - x_{n_k}\| < +\infty$$

ce qui permet par hypothèse de dire qu'il existe $S \in E$ tel que

$$\sum_{k=0}^{N-1} (x_{n_{k+1}} - x_{n_k}) \xrightarrow[N\to+\infty]{} S.$$

Mais on a par somme télescopique

$$\sum_{k=0}^{N-1} (x_{n_{k+1}} - x_{n_k}) = x_{n_N} - x_{n_0}$$

donc la suite $(x_{n_k})_{k\in\mathbb{N}}$ converge vers $x_{n_0} + S \in E$. Or on sait qu'une suite de Cauchy ayant une suite extraite convergente (dans E) est elle-même convergente dans E (voir l'exercice 1.71), ce qui conclut la preuve. □

Exemple 2.20. Dans le cas où $(E, \|\cdot\|) = (\mathbb{R}, |\cdot|)$ qui est complet, on retrouve la proposition 2.39 du Chapitre I, à savoir que pour $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$

$$\sum_{n\in\mathbb{N}} |x_n| < +\infty \ \Rightarrow \quad \text{la série de terme général } (x_n)_{n\in\mathbb{N}} \text{ converge dans } \mathbb{R},$$

autrement dit que la convergence absolue implique la convergence d'une série réelle. Le même résultat vaut pour $(\mathbb{C}, |\cdot|)$ également.

Dans le cas où $(E, \|\cdot\|) = (B(X,\mathbb{R}), \|\cdot\|_\infty)$ dont on a vu qu'il est complet, on retrouve la proposition 4.17 du Chapitre I [36], à savoir que pour $(f_n)_{n\in\mathbb{N}} \in B(X,\mathbb{R})^{\mathbb{N}}$,

$$\sum_{n\in\mathbb{N}} \|f_n\|_\infty < +\infty \ \Rightarrow \quad \text{la série de terme général } (f_n)_{n\in\mathbb{N}} \text{ converge dans } B(X,\mathbb{R}),$$

à savoir que la convergence normale implique la convergence uniforme.

Attention donc à l'expression "convergence absolue" dans le cadre d'un espace vectoriel normé $(E, \|\cdot\|)$: dans le premier exemple on retrouve bien la convergence absolue des réels, mais dans le second on parle en fait de convergence normale (là on la convergence absolue de la série de fonctions a également un sens, qui n'est pas le même).

36. À ceci près qu'on suppose ici les fonctions bornées pour être dans un cadre d'espace vectoriel normé.

2.3 Connexité

Dans le cadre des espaces vectoriels normés, on a une réciproque partielle à la proposition 1.126 :

Proposition 2.21. *Soit $(E, \|\cdot\|)$ un espace vectoriel normé et $O \subset E$.*

Si O est un ouvert connexe, alors O est connexe par arcs.

En fait, nous allons montrer un résultat un peu plus fort (mais la démonstration n'est pas plus compliquée), qui a été utilisé au Chapitre II :

Proposition 2.22. *Soit $(E, \|\cdot\|)$ un espace vectoriel normé et $O \subset E$.*

Si O est un ouvert connexe, alors O est connexe par ligne brisée,

où la connexité par ligne brisée se définit comme suit :

$$\forall (x,y) \in O,\ \exists N \in \mathbb{N}^*,\ \exists (x_0 = x, x_1, \dots, x_{N-1}, x_N = y) \in O^{N+1}$$
$$\textit{tels que } \forall i \in [\![0, N-1]\!],\quad [x_i, x_{i+1}] \subset O.$$

Remarque 2.23. Pour comprendre que le caractère "connexe par ligne brisée" est plus fort que "connexe par arcs", il suffit de savoir paramétrer un segment et concaténer plusieurs chemins comme on l'a fait à l'exemple 3.11 du Chapitre II.

Démonstration. Soit O un ouvert connexe de E, et supposons O non vide. Pour $x_0 \in O$, on pose T_{x_0} l'ensemble des points de O qu'on peut joindre par une ligne brisée. Clairement $x_0 \in T_{x_0}$, le chemin $[x_0, x_0]$ étant bien une ligne brisée de x_0 à x_0. De plus T_{x_0} est ouvert, car si $x \in T_{x_0}$, alors comme O est ouvert il existe $\alpha > 0$ tel que $B(x, \alpha) \subset O$, mais alors si $y \in B(x, \alpha)$, on peut le joindre à x_0 par ligne brisée en ajoutant à une ligne brisée qui joint x_0 à x (qui existe par définition de T_{x_0}) le segment $[x, y]$ qui est dans $B(x, \alpha)$ et donc dans O. Enfin T_{x_0} est fermé dans O : pour le voir, on montre que son complémentaire (dans O) est ouvert dans O. Soit $x \in O \setminus T_{x_0}$. Comme O est ouvert, il existe $\alpha > 0$ tel que $B(x, \alpha) \subset O$. Si par l'absurde il existait $y \in B(x, \alpha) \cap T_{x_0}$, alors il existerait une ligne brisée entre x_0 et y, mais en lui ajoutant $[y, x]$ on aurait une ligne brisée entre x_0 et y, ce qui est une contradiction. Ainsi $B(x, \alpha) \subset O \setminus T_{x_0}$ et donc T_{x_0} est fermé. Par connexité, on en déduit $T_{x_0} = O$, ce qu'il fallait montrer. □

2.4 Compacité

2.4.1 Le cas des espaces vectoriels normés de dimension finie

On commence par analyser un cas particulier :

Proposition 2.24. *Les parties compactes de $(\mathbb{R}^n, \|\cdot\|_\infty)$ pour $n \in \mathbb{N}^*$ sont les fermés bornés.*

Démonstration. On a vu à la proposition 1.101 qu'un ensemble compact est toujours un ensemble fermé et borné. On s'intéresse donc à la réciproque. Soit donc $F \subset \mathbb{R}^n$ une partie fermée et bornée, et $(x_k)_{k \in \mathbb{N}}$ une suite de F. Il existe donc M tel que $F \subset \overline{B}_{\|\cdot\|_\infty}(0, M) = [-M, M]^n$. Par le théorème 1.103, $[-M, M]$ est un compact de $\mathbb{R}$, et comme la norme infinie est la norme produit sur $\mathbb{R}^n$ (voir la remarque 2.10), on peut appliquer la proposition 1.106 qui donne que $([-M, M]^n, \|\cdot\|_\infty)$ est compact. Ainsi il existe une extraction φ et $x_\infty \in [-M, M]^n$ tels que $x_{\varphi(n)} \xrightarrow[n \to +\infty]{} x_\infty$. Enfin, F étant fermé, on a bien $x_\infty \in F$. □

Ce premier exemple a pour conséquence le résultat fondamental suivant :

Théorème 2.25. *Soit E un espace vectoriel de dimension finie. Alors toutes les normes sur E sont équivalentes.*

Démonstration. On commence en supposant que le corps de base est $\mathbb{R}$.

On note $n = \dim(E)$, N une norme sur E, et $(e_i)_{i\in[\![1,n]\!]}$ une base de E. On peut définir pour $x = \sum_{i=1}^n x_i e_i$, $\|x\|_\infty = \max_{i\in[\![1,n]\!]} |x_i|$. Nous allons montrer que N est équivalente à cette norme, et par transitivité de l'équivalence des normes, on en concluera que toutes les normes sont équivalentes.

On a en premier lieu l'inégalité

$$\forall x = \sum_{i=1}^n x_i e_i \in E,\ \ N(x) = N\left(\sum_{i=1}^n x_i e_i\right) \leq \sum_{i=1}^n |x_i| N(e_i) \leq \left(\sum_{i=1}^n N(e_i)\right) \|x\|_\infty = C\|x\|_\infty \tag{2.1}$$

où on a posé $C = \sum_{i=1}^n N(e_i)$.

Une conséquence de cette première comparaison est que la fonction [37] $N : (E, \|\cdot\|_\infty) \to \mathbb{R}$ est une application continue : en effet par l'inégalité triangulaire pour N, on peut écrire :

$$\forall (x,y) \in E,\ \ |N(x) - N(y)| \leq N(x-y) \leq C\|x-y\|_\infty$$

qui implique le caractère lipschitzien et donc continu de $N : (E, \|\cdot\|_\infty) \to \mathbb{R}$.

On peut donc introduire l'ensemble $S = \{x \in E, \|x\|_\infty = 1\}$ la sphère unité de E pour la norme $\|\cdot\|_\infty$. Via la base de E choisie, cet ensemble est homéomorphe à $\Big\{(x_1,\dots,x_n) \in \mathbb{R}^n,\ \|(x_1,\dots,x_n)\|_\infty = 1\Big\}$ qui est fermé et borné, donc compact d'après la proposition 2.24.

On peut appliquer le théorème des bornes 1.107 au problème d'optimisation

$$\min\Big\{N(x),\ \ x \in S\Big\}$$

qui dit que comme $N : (E, \|\cdot\|_\infty) \to \mathbb{R}$ est continue et $(S, \|\cdot\|_\infty)$ est compact, le problème précédent admet une solution, c'est-à-dire qu'il existe $x_0 \in S$ tel que

$$\forall x \in S,\ N(x) \geq N(x_0).$$

On pose $c = N(x_0) > 0$. Par homogénéité de la norme, on peut déduire de l'inégalité précédente la propriété :

$$\forall x \in E \setminus \{0\},\ N(x) = \|x\|_\infty N\underbrace{\left(\frac{x}{\|x\|_\infty}\right)}_{\in S} \geq c\|x\|_\infty.$$

Combiné à (2.1), on conclut donc à l'équivalence entre N et $\|\cdot\|$.

Si maintenant E est un espace vectoriel sur $\mathbb{C}$ de dimension n, on peut le voir comme un espace vectoriel sur $\mathbb{R}$ de dimension $2n$, et donc par le cas précédent, le résultat est encore valable. □

Le résultat précédent est fondamental, on en donne ici quelques corollaires. Voir aussi le théorème 2.53 qui en est aussi conséquence.

Corollaire 2.26. *Soit $(E, \|\cdot\|)$ un espace vectoriel de dimension finie. Alors les compacts de E sont exactement les fermés bornés de E.*

Démonstration. En travaillant dans une base, on peut se ramener au cas où $E = \mathbb{R}^n$ avec $n \in \mathbb{N}^*$ est la dimension [38] de E. Par le théorème 2.25, la norme $\|\cdot\|$ est équivalente à la norme $\|\cdot\|_\infty$ sur $\mathbb{R}^n$, et par la proposition 2.24, les compacts pour $\|\cdot\|_\infty$ sont les fermés bornés pour $\|\cdot\|_\infty$. Comme les propriétés de fermé, borné, compact ne changent pas pour deux normes équivalentes, le résultat est donc encore valable pour caractériser les compacts de E pour la norme initiale $\|\cdot\|$. □

Remarque 2.27. Attention aux deux hypothèses :

1. on travaille dans un espace vectoriel normé,
2. on demande que l'espace vectoriel soit de dimension finie !

37. Attention, notez bien ici qu'on a muni l'espace de départ de la norme $\|\cdot\|_\infty$; quand on dit qu'une norme est continue, on sous-entend qu'on munit l'espace de départ de la norme dont il est question. Ici on a deux normes qui rentrent en jeu, et précisément on regarde ici la continuité de l'une d'elle quand l'espace est muni de l'autre.

38. On n'interdit pas la dimension $n = 0$, simplement dans ce cas il n'y a pas grand chose à dire...

Pour le second point, on verra effectivement avec le théorème de Riesz 2.29 ci-dessous que le corollaire précédent n'est jamais vrai dans un espace vectoriel de dimension infinie. Plus précisément, les boules fermées d'un espace vectoriel de dimension infinie ne sont jamais des ensembles compacts.
[*]Pour le premier point, il faut également être vigilant. On peut rencontrer des espaces vectoriels[39] muni d'une métrique, et dans lesquels les fermés bornées sont exactement les compacts de l'espace. Attention néanmoins à la définition du mot borné dans ce contexte, comme on l'a noté à la remarque 1.16 ; la notion de compacité étant topologique, un tel énoncé ne peut valoir que si on a une définition topologique du caractère borné. Or la première définition qui vient à l'esprit et qui consiste à demander aux ensembles d'être de diamètre fini n'est pas du tout une notion topologique. Voir le paragraphe 4.4.3.

Sans rapport avec la compacité, on peut néanmoins citer le corollaire fondamental suivant :

Corollaire 2.28. — *Soit $(E, \|\cdot\|)$ un espace vectoriel de dimension finie. Alors E est complet.*

— *Soit $(E, \|\cdot\|)$ un espace vectoriel normé, et $F \subset E$ un sous-espace vectoriel de E de dimension finie. Alors F est fermé.*

Démonstration. Pour la première propriété, E est isométrique à $\mathbb{R}^n$ pour un certain $n \in \mathbb{N}$, qui est complet par exemple pour la norme $\|\cdot\|_\infty$. Et donc E est complet.

Pour la seconde propriété, on dit que F (muni de la restriction à F de la norme de E) est complet par le point précédent, et donc fermé dans E. □

2.4.2 Le cas des espaces vectoriels normés de dimension infinie

Le résultat suivant ne fournit pas seulement un contre-exemple au corollaire 2.26, il affirme que n'importe quel espace vectoriel normé de dimension infinie contient des ensembles fermés et bornés qui ne sont pas des compacts, et que ces ensembles sont particulièrement simples puisqu'il suffit de considérer les boules fermés.

Théorème 2.29 (Théorème de Riesz). *Soit $(E, \|\cdot\|)$ un espace vectoriel normé. Alors la boule unité fermée est compacte si et seulement E est de dimension finie.*

Remarque 2.30. L'énoncé est donné pour la boule unité fermée, mais il est facile d'en déduire que le résultat est le même pour toute boule fermée $\overline{B}(x, r)$ de centre $x \in E$ et de rayon $r > 0$.

Exercice 2.31. Montrer "à la main", c'est-à-dire en exhibant une suite n'ayant pas de sous-suite convergente, que les boules unités fermés des espaces suivants ne sont pas compactes :

1. l'espace $\ell^\infty(\mathbb{N})$ muni de la norme $\|\cdot\|_\infty$ et plus généralement $\ell^p(\mathbb{N})$ muni de la norme $\|\cdot\|_p$ pour $p \in [1, +\infty]$.
2. l'espace des fonctions continues $C^0([0, 1])$ muni de la norme uniforme.

Démonstration. Si E est de dimension finie, le corollaire 2.26 permet de voir que tout ensemble fermé borné est compact, et c'est donc en particulier le cas de la boule unité fermée.

On propose deux preuves pour la réciproque :

Preuve 1 : Réciproquement, notons $\overline{B} = \{x \in E : \|x\| \leq 1\}$ la boule unité fermée de E, que nous supposons compacte. Par la propriété de Borel-Lebesgue, il existe un nombre fini de points $x_1, \ldots, x_N \in \overline{B}$ tels que

$$\overline{B} \subset \bigcup_{i=1}^{N} B\Big(x_i, \frac{1}{2}\Big). \tag{2.2}$$

Soit $F = \mathrm{vect}\{x_i\}_{1 \leq i \leq N}$ qui est un sous-espace vectoriel de E de dimension finie. En particulier, c'est un fermé par le corollaire 2.28. Pour tout $y \in \overline{B}$, par la propriété de recouvrement, il existe $1 \leq i_1 \leq N$, $\alpha_1 \in \mathbb{R}$ tel que $|\alpha_1| < \frac{1}{2}$ et $y_1 \in \overline{B}$, tels que

$$y = x_{i_1} + \alpha_1 y_1.$$

39. Attention, un espace métrique n'a en général aucune structure algébrique, il n'y a donc pas de sens de parler de sa dimension. On évoque donc ici les espaces vectoriels munis d'une métrique.

Mais on peut dire la même chose de $y_1 = x_{i_2} + \alpha_2 y_2$, si bien que

$$y = x_{i_1} + \alpha_1 x_{i_2} + \alpha_1\alpha_2 y_2 = z_2 + \beta_2 y_2 \text{ avec } z_2 \in F, y_2 \in \overline{B} \text{ et } |\beta_2| < \frac{1}{4}.$$

On établit ainsi par récurrence l'existence de $z_k \in F$, $y_k \in \overline{B}$ et β_k tel que $|\beta_k| < \frac{1}{2^k}$ tels que

$$y = z_k + \beta_k y_k.$$

Comme $\|\beta_k y_k\| = |\beta_k|\|y_k\| < \frac{1}{2^k} \xrightarrow[k\to+\infty]{} 0$, on voit que $z_k \xrightarrow[k\to+\infty]{} y$. Mais F est fermé, donc $y \in F$. Ceci montre que $\overline{B} \subset F$. On conclut enfin en remarquant que si $y \notin \overline{B}$, alors $\frac{y}{\|y\|} \in \overline{B}$, donc que $y = \|y\|\frac{y}{\|y\|} \in F$.

Preuve 2 : On se donne $\varepsilon \in]0,1[$ qui sera choisi en fin de preuve. Comme pour la méthode précédente, on commence par appliquer la propriété de Borel-Lebesgue, qui donne qu'il existe un nombre fini de points $x_1, \dots, x_N \in \overline{B}$ tels que

$$\overline{B} \subset \bigcup_{i=1}^{N} B(x_i, \varepsilon), \tag{2.3}$$

et on pose $F = \mathrm{Vect}\{x_i\}_{1\le i\le N}$. Si $F = E$, alors le résultat suit. Dans le cas contraire, on peut trouver un $x \in E \setminus F$. Notons $d = \mathrm{dist}(x, F) = \inf_{y\in F}\|x - y\|$. Comme F est de dimension finie, F est fermé dans E et donc $d > 0$. On peut alors trouver un $y_\varepsilon \in F$ tel que

$$d \le \|x - y_\varepsilon\| \le \frac{d}{1-\varepsilon}.$$

Posons $z_\varepsilon = \frac{x - y_\varepsilon}{\|x - y_\varepsilon\|} \in \overline{B}$. Alors pour tout $y \in F$,

$$\|z_\varepsilon - y\| = \frac{\Big\|x - y_\varepsilon - \|x - y_\varepsilon\|y\Big\|}{\|x - y_\varepsilon\|} \ge d\,\frac{1-\varepsilon}{d} = 1 - \varepsilon$$

car $y_\varepsilon + \|x - y_\varepsilon\|y \in F$, ce qui montre que

$$\mathrm{dist}(z_\varepsilon, F) \ge 1 - \varepsilon.$$

Par ailleurs, d'après (2.2), il existe un $i \in \{1, \dots, N\}$ tel que $z_\varepsilon \in B(x_i, \varepsilon)$ et donc, puisque $x_i \in F$,

$$\mathrm{dist}(z_\varepsilon, F) \le \|z_\varepsilon - x_i\| < \varepsilon.$$

On aboutit à une contradiction en choisissant $\varepsilon = 1/2$. □

Remarque 2.32. Le théorème de Riesz est en quelque sorte une mauvaise nouvelle, et suggère qu'il peut être difficile de déterminer les ensembles compacts dans un espace vectoriel de dimension infinie. C'est une question naturelle et qui n'a pas de réponse "parfaite", au sens où chaque exemple nécessitera une étude spécifique. Le théorème d'Ascoli est l'exemple le résultat le plus important en ce sens, car il donne une réponse à la question "qui est un ensemble compact" dans le cadre de l'espace des fonctions continues, voir le paragraphe 3.2.

2.4.3 Application à l'optimisation

Nous précisons ici de petites applications classiques de la classification des compacts dans les espaces vectoriels normés de dimension finie à l'optimisation ; en effet cela permet d'établir l'existence d'un minimum dans de nombreuses situations.

En premier lieu, le théorème des bornes se traduit dans ce contexte par le résultat :

Corollaire 2.33. *Soit $(E, \|\cdot\|)$ un espace vectoriel normé de dimension finie et $f : C \to \mathbb{R}$ une fonction continue sur un ensemble $C \subset E$ fermé borné. Alors il existe un $x_* \in E$ tel que $f(x_*) \le f(x)$ pour tout $x \in C$, autrement dit le problème d'optimisation*

$$\inf\Big\{f(x),\ x \in E\Big\}$$

admet une solution.

Démonstration. Il suffit d'appliquer le théorème des bornes 1.107 combiné au fait que C est compact. □

Exercice 2.34. On peut construire des contre-exemples dans le cas des espaces vectoriels de dimension infinie : on pose $E = C^0([0,1])$ muni de la norme uniforme $\|\cdot\|_\infty$. On définit

$$\begin{array}{rccl} G: & E & \to & \mathbb{R} \\ & f & \mapsto & \displaystyle\int_0^1 |f(x)|dx \end{array}$$

1. Montrer G est bien définie et continue.
2. On pose $C = \{f \in E,\ \|f\|_\infty \le 2,\ f(0) = 1\}$. Montrer que C est fermé et borné dans E.
3. Montrer à la main d'une part, et avec le théorème de Riesz d'autre part, que C n'est pas compact.
4. Etudier la question d'existence pour le problème de minimisation suivant :

$$\inf_{f\in C} G(f).$$

et en déduire à nouveau que C n'est pas compact.

Terminons par un moyen classique de traiter le cas d'ensembles C non bornés :

Définition 2.35. Soit $(E, \|\cdot\|)$ un espace vectoriel normé et $f : C \to \mathbb{R}$ une fonction définie sur un ensemble C non borné [40]. On dit que f est coercive si

$$\lim_{x\in C, \|x\|\to+\infty} f(x) = +\infty$$

c'est-à-dire

$$\forall A \in \mathbb{R},\ \exists R > 0,\ \forall x \in C \text{ tel que } \|x\| > R,\quad f(x) \ge A.$$

Proposition 2.36. *Soit $(E, \|\cdot\|)$ un espace vectoriel normé de dimension finie et $f : C \to \mathbb{R}$ une fonction continue et coercive sur un ensemble C non vide, fermé et non borné. Alors il existe un $x_* \in C$ tel que $f(x_*) \le f(x)$ pour tout $x \in C$, autrement dit le problème d'optimisation*

$$\inf \left\{f(x),\ x \in C\right\}$$

admet une solution.

Démonstration. Il existe [41] une suite minimisante $(x_n)_{n\in\mathbb{N}}$ telle que $f(x_n) \xrightarrow[n\to+\infty]{} \inf_C f$ (voir l'exercice 2.18 au Chapitre I). Mais par coercivité de f, il existe $R > 0$ tel que si $x \in C$ et $\|x\| > R$, alors $f(x) > f(0)+1$. Comme $\inf_C f < f(0)+1$, il existe $n_0 \in \mathbb{N}$ tel que $f(x_n) \le f(0)+1$ pour tout $n \ge n_0$ de sorte que $\|x_n\| \le R$ pour tout $n \ge n_0$. Donc la suite $(x_n)_{n\ge n_0}$ est bornée et, E étant de dimension finie, par le corollaire 2.26 $C \cap \overline{B}(0,R)$ est compact et donc il existe une sous-suite $(x_{\phi(n)})_{n\in\mathbb{N}}$ qui converge vers un $x_* \in C$. La fonction f étant continue en $x_* \in C$, il vient,

$$f(x_*) = \lim_{n\to+\infty} f(x_{\phi(n)}) = \inf_C f,$$

ce qui montre que x_* est un point de minimum de f. □

40. Cette hypothèse est nécessaire pour pouvoir parler de la limite de f quand $\|x\|$ tend vers $+\infty$. De toute façon, si C est borné, on est ramené au corollaire 2.33.
41. Simplement car C est supposé non vide.

2.5 Applications linéaires continues

2.5.1 Norme d'application linéaire continue

On considère deux $\mathbb{K}$-espaces vectoriels normés $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ et des applications linéaires de E dans F.

Proposition 2.37. *Une application linéaire $T\colon E \to F$ est continue si et seulement s'il existe une constante $C > 0$ telle que pour tout $x \in E$,*

$$\|T(x)\|_F \leq C\|x\|_E. \tag{2.4}$$

Démonstration. Si (2.4) est vérifié, alors pour tout $x, y \in E$, par linéarité de T, on a $\|T(y) - T(x)\|_F = \|T(y-x)\|_F \leq C\|y-x\|_E$, ce qui montre que T est continue.

Réciproquement, si T est continue, elle l'est notamment en 0. En particulier, il existe $\delta > 0$ tel que si $\|x\|_E \leq \delta$, alors $\|T(x)\|_F \leq 1$. Pour tout $x \neq 0$, $y = \delta x/\|x\|_E$ satisfait $\|y\|_E = \delta$ et donc

$$\frac{\delta}{\|x\|_E}\|T(x)\|_F = \|T(y)\|_F \leq 1,$$

soit $\|T(x)\|_F \leq \frac{1}{\delta}\|x\|_E$. Si $x = 0$, l'inégalité reste encore valable bien sûr. □

On note $\mathcal{L}(E,F)$ l'ensemble des applications linéaires continues de E dans F. Il s'agit clairement d'un espace vectoriel. Si $T \in \mathcal{L}(E,F)$, on définit la quantité[42]

$$\begin{aligned}\|T\|_{\mathcal{L}(E,F)} &= \sup_{x\in E, x\neq 0} \frac{\|T(x)\|_F}{\|x\|_E} = \sup_{\|x\|_E \leq 1} \|T(x)\|_F \\ &= \inf\left\{C \geq 0,\ \ \forall x \in E,\ \ \|T(x)\|_F \leq C\|x\|_E\right\}\end{aligned}$$

Remarque 2.38. On peut définir $\|T\|_{\mathcal{L}(E,F)}$ pour une application $T : E \to F$ seulement supposée linéaire, et alors

$$T \text{ est continue } \Leftrightarrow \|T\|_{\mathcal{L}(E,F)} < +\infty.$$

Exercice 2.39. Soit $T : E \to F$ linéaire. Montrer

$$\begin{aligned} T \text{ est continue sur } E &\Leftrightarrow T \text{ est continue en } 0 \in E, \\ &\Leftrightarrow T \text{ est bornée sur } \overline{B}(0,1), \\ &\Leftrightarrow T \text{ est lipschitzienne sur } E. \end{aligned}$$

Dans ce cas, $\|T\|_{\mathcal{L}(E,F)}$ est la borne de T sur $\overline{B}(0,1)$ et est aussi la constante de Lipschitz de T.

Proposition 2.40. *La quantité $\|\cdot\|_{\mathcal{L}(E,F)}$ définit une norme sur $\mathcal{L}(E,F)$.*

Remarque 2.41. Notez que cette norme n'est pas complètement nouvelle : il s'agit simplement de la norme uniforme de la restriction de T à la boule unité de E. La partie spécifique ici est de voir pourquoi cette restriction n'empêche pas à la "définie positivité" de la fonction : comme on le verra ci-dessous, ceci est dû au fait qu'une application linéaire est nulle si et seulement si elle est nulle sur la boule unité.

Démonstration. Si $T \in \mathcal{L}(E,F)$ et $\lambda \in \mathbb{K}$, comme $\|\cdot\|_F$ est une norme sur F, il vient

$$\|\lambda T\|_{\mathcal{L}(E,F)} = \sup_{\|u\|_E \leq 1} \|\lambda T(u)\|_F = |\lambda| \sup_{\|u\|_E \leq 1} \|T(u)\|_F = |\lambda|\|T\|_{\mathcal{L}(E,F)},$$

ce qui établit l'homogénéité. On a évidemment que $\|0\|_{\mathcal{L}(E,F)} = 0$, et si $\|T\|_{\mathcal{L}(E,F)} = 0$, alors $\|T(u)\|_F = 0$ pour tout $u \in E$, ce qui implique que $T(u) = 0$ pour tout $u \in E$, soit $T = 0$. Enfin, si T_1 et $T_2 \in \mathcal{L}(E,F)$, pour tout $u \in E$ tel que $\|u\|_E \leq 1$, on a

$$\|(T_1 + T_2)(u)\|_F = \|T_1(u) + T_2(u)\|_F \leq \|T_1(u)\|_F + \|T_2(u)\|_F \leq \|T_1\|_{\mathcal{L}(E,F)} + \|T_2\|_{\mathcal{L}(E,F)}.$$

Prenant le sup dans le membre de gauche, il vient $\|T_1 + T_2\|_{\mathcal{L}(E,F)} \leq \|T_1\|_{\mathcal{L}(E,F)} + \|T_2\|_{\mathcal{L}(E,F)}$, ce qui montre l'inégalité triangulaire. □

42. On invite le lecteur à s'assurer que les deux différentes définitions sont bien égales.

Remarquons que pour tout $x \in E$, on a

$$\|T(x)\|_F \leq \|T\|_{\mathcal{L}(E,F)} \|x\|_E, \tag{2.5}$$

et que $\|T\|_{\mathcal{L}(E,F)}$ est la plus petite constante C qui apparaît dans la Proposition 2.37.

Exercice 2.42. *[Calculs de norme d'applications linéaires ; la méthode de calcul de norme d'application linéaire est à maîtriser absolument]*

1. Soit $E = \mathbb{R}[X]$, muni de la norme $\|P\|_\infty := \sup_{x\in[0,1]} |P(x)|$.
   (a) Justifier qu'il s'agit bien d'une norme.
   (b) Pour $a \in \mathbb{R}$, on considère $\delta_a : \mathbb{R}[X] \to \mathbb{R}$ la forme linéaire définie par $\forall P \in E, \delta_a(P) = P(a)$. Déterminer pour quels $a \in \mathbb{R}$ la forme linéaire δ_a est continue, et calculer $\|\delta_a\|$ dans ce cas.
2. Soit $E = C^0([0,1])$. On note
$$\forall f \in E, \quad \|f\|_1 := \int_0^1 |f(t)| dt.$$
   (a) Montrer que $\|\cdot\|_1$ est une norme sur E.
   (b) Montrer que l'application $\delta_0 : f \in E \mapsto f(0) \in \mathbb{R}$ est continue si on munit E de la norme $\|\cdot\|_\infty$, mais n'est pas continue si on munit E de la norme $\|\cdot\|_1$.
   (c) On munit E de la norme uniforme. Soit $\phi : E \to \mathbb{R}$ la forme linéaire définie par
$$\forall f \in E, \quad \phi(f) := \int_0^{1/2} f(t)dt - \int_{1/2}^1 f(t)dt.$$
   Montrer que ϕ est continue, calculer sa norme, et montrer que cette norme n'est pas atteinte, c'est-à-dire qu'il n'existe pas de fonction $f \in E$ telle que $\|f\|_\infty = 1$ et $|\phi(f)| = \|\phi\|$.
   (d) Refaire la question précédente en munissant E de la norme $\|\cdot\|_1$.

Exercice 2.43. On travaille dans $\ell^\infty := \ell^\infty(\mathbb{N}^*)$ muni de la norme uniforme. On considère $T : \ell^\infty \to \ell^\infty$ défini par

$$\forall x \in \ell^\infty, \quad T(x) = T(x_1, x_2, x_3, \ldots) := (x_1/1, x_2/2, x_3/3, \ldots) = (x_n/n)_{n\in\mathbb{N}^*}.$$

1. Montrer que T est bien défini, et linéaire continu avec $\|T\| = 1$.
2. Montrer que T est injectif, mais non surjectif. On obtient donc un contre-exemple au résultat d'algèbre linéaire "ùn endomorphisme de E est bijectif si et seulement s'il est injectif", qui n'est pas valable si E est de dimension infinie.

Exercice 2.44. Soit E, F, G trois espaces vectoriels normés.

1. Soit $b : E \times F \to G$ une application bilinéaire. Montrer que b est continue si et seulement si il existe $M \geq 0$ tel que
$$\forall (x,y) \in E \times F, \ \|b(x,y)\|_G \leq M \|x\|_E \|y\|_F.$$
2. Montrer que
$$\|b\|_{\mathcal{L}_2(E\times F,G)} := \sup_{\|x\|_E\leq 1, \|y\|_F \leq 1} \|b(x,y)\|_G$$
définit une norme sur $\mathcal{L}_2(E \times F, G)$ l'ensemble des applications bilinéaires continues de $E \times F$ dans G.
3. (*)Construire une isométrie linéaire et bijective entre
$$\mathcal{L}(E, \mathcal{L}(F,G)) \qquad \text{et} \qquad \mathcal{L}_2(E \times F, G).$$
(Ces espaces sont munis de leurs normes naturelles). Cette identification sera utile au Chapitre VII lorqu'on définiera les différentielles secondes.

2.5.2 Propriétés

Une propriété fondamentale de ce type de norme est sa sous-additivité vis-à-vis de la composition :

Proposition 2.45. *Soit E, F, G trois espaces vectoriels normés, et $U : E \to F, V : F \to G$ deux applications linéaires (continues)*[43]. *Alors $V \circ U$ est linéaire de E dans G et*[44]

$$\|V \circ U\|_{\mathcal{L}(E,G)} \leq \|V\|_{\mathcal{L}(F,G)} \|U\|_{\mathcal{L}(E,F)}. \tag{2.6}$$

Démonstration. On utilise (2.5) deux fois :

$$\forall x \in E, \quad \|V \circ U(x)\|_G \leq \|V\|_{\mathcal{L}(F,G)} \|U(x)\|_F \leq \|V\|_{\mathcal{L}(F,G)} \|U\|_{\mathcal{L}(E,F)} \|x\|_E,$$

et donc par interprétation de la norme comme la plus petite constante dans une telle inégalité, on obtient le résultat. □

Remarque 2.46. Dans le cas des endomorphismes (applications linéaires dont l'espace de départ et d'arrivée sont les mêmes) d'un espace vectoriel normé E, on note simplement $\mathcal{L}(E)$ l'ensemble des applications linéaires continues de E dans E. L'inégalité (2.6) affirme la continuité de l'application bilinéaire $\circ : \mathcal{L}(E) \times \mathcal{L}(E) \to \mathcal{L}(E)$, voir l'exercice 2.44 : on dit que l'espace $(\mathcal{L}(E), +, \cdot, \circ, \| \cdot \|_{\mathcal{L}(E)})$ est une algèbre de Banach.

Nous allons maintenant énoncer une condition suffisante pour que $\mathcal{L}(E, F)$ soit un espace de Banach pour la norme que l'on vient de définir.

Proposition 2.47. *Si F est complet, alors $\mathcal{L}(E, F)$ l'est aussi.*

Du fait de la remarque 2.41, la preuve sera très similaire à la preuve de la proposition 4.7 du Chapitre I, voir aussi l'exercice 2.16.

Démonstration. Soit $(T_n)_{n \in \mathbb{N}}$ une suite de Cauchy dans $\mathcal{L}(E, F)$. Alors pour tout $\varepsilon > 0$, il existe un $N \in \mathbb{N}$ tel que pour tous $p \geq N$, $q \geq N$, et $x \in E$,

$$\|T_q(x) - T_p(x)\|_F \leq \|T_q - T_p\|_{\mathcal{L}(E,F)} \|x\|_E \leq \varepsilon \|x\|_E. \tag{2.7}$$

On en déduit que $(T_n(x))_{n \in \mathbb{N}}$ est une suite de Cauchy dans F, qui est complet. Il existe donc un vecteur que l'on nomme $T(x) \in F$ tel que $T_n(x) \xrightarrow[n \to +\infty]{} T(x)$. En passant à la limite dans

$$\forall \lambda \in \mathbb{K}, \ \forall (x, y) \in E^2, \ \forall n \in \mathbb{N}, \ T_n(\lambda x + y) = \lambda T_n(x + y) + T_n(y)$$

on en déduit que T est linéaire. Comme $(T_n)_{n \in \mathbb{N}}$ est une suite de Cauchy dans $\mathcal{L}(E, F)$, elle est bornée dans cet espace et donc il existe une constante $C > 0$ telle que $\|T_n\|_{\mathcal{L}(E,F)} \leq C$. Par conséquent, en fixant $x \in E$ on a $\|T_n(x)\|_F \leq \|T_n\|_{\mathcal{L}(E,F)} \|x\|_E \leq C\|x\|_E$. Comme $T_n(x) \xrightarrow[n \to +\infty]{} T(x)$ dans F, $\|T_n(x)\|_F \xrightarrow[n \to +\infty]{} \|T(x)\|_F$ et il vient par passage à la limite que $\forall x \in E$, $\|T(x)\|_F \leq C\|x\|_E$, soit $T \in \mathcal{L}(E, F)$. Enfin, pour $\|x\|_E \leq 1$ on fait tendre q vers $+\infty$ dans (2.7) et on obtient

$$\|T_p(x) - T(x)\|_F \leq \varepsilon, \text{ pour tout } p \geq N.$$

Par passage au sup en $x \in \overline{B}_E(0, 1)$ dans le membre de gauche, il s'ensuit que $\|T_n - T\|_{\mathcal{L}(E,F)} \leq \varepsilon$ pour tout $n \geq N$, ce qui montre que $T_n \xrightarrow[n \to +\infty]{} T$ dans $\mathcal{L}(E, F)$. □

Exercice 2.48 (Endomorphismes inversibles ; cet exercice est un grand classique à connaître)**.**

1. Soit $n \in \mathbb{N}^*$. Montrer que l'ensemble $GL_n(\mathbb{R})$ des matrices inversibles est un ouvert dense de $\mathcal{M}_n(\mathbb{R})$ (muni de sa topologie naturelle).

43. On a mis entre parenthèse car l'inégalité reste valable même sans continuité ; simplement l'inégalité devient alors triviale car la norme des applications non continue est $+\infty$.
44. Et en particulier l'inégalité permet de voir que $V \circ U$ est continue (de norme finie) si U et V sont continues.

2. Soit E un espace de Banach. On considère $\mathcal{L}(E)$ l'espace des endomorphismes continus de E, muni de la norme d'opérateur.
   (a) Montrer que si $u \in \mathcal{L}(E)$ vérifie $\|u\| < 1$, alors $Id - u$ est inversible, et que de plus
   $$\|(Id-u)^{-1}\| \leq (1-\|u\|)^{-1}.$$
   (b) Montrer que l'ensemble des endomorphismes inversibles de $\mathcal{L}(E)$ est un ouvert de $\mathcal{L}(E)$.
   (c) [*] Montrer que l'application $u \mapsto u^{-1}$ définie sur l'ensemble des inversibles de $\mathcal{L}(E)$ est continue.

Exemple 2.49. Si $(E, \|\cdot\|)$ est un espace vectoriel normé, on note $E' = \mathcal{L}(E, \mathbb{R})$ l'ensemble des formes linéaires continues sur E, que l'on appelle le dual topologique de E. On dira plus rapidement « dual » dans la suite, car le dual algébrique, c'est-à-dire l'espace de toutes les formes linéaires ne nous servira pas souvent. Le dual de E est un espace vectoriel normé muni la norme

$$\|L\|_{E'} = \sup_{x \in E, \|x\| \leq 1} |L(x)|,$$

dite norme duale, pour laquelle il est complet d'après la proposition 2.47. Il s'agit donc d'un espace de Banach, même si E n'en est pas un.

La caractérisation du dual d'un espace vectoriel normé est un problème difficile en analyse fonctionnelle.

Exercice 2.50 (Formes linéaires et continuité)**.** Soit E un espace vectoriel normé et $L : E \to \mathbb{K}$ une forme linéaire.

1. Montrer que
$$L \text{ est continue sur } E \quad \Leftrightarrow \quad \mathrm{Ker}(L) \text{ est fermé.}$$
2. Montrer que $\mathrm{Ker}(L)$ est soit fermé soit dense dans E.

Notons que le théorème 1.87 de prolongement des applications uniformément continues peut être réexprimé dans le cadre des applications linéaires continues :

Corollaire 2.51. *Soit $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ deux espaces vectoriels normés et $D \subset E$ un sous-espace vectoriel de E. On suppose que $(F, \|\cdot\|_F)$ est un espace de Banach. Si $T : D \to F$ est une application linéaire continue de D (muni de la norme de E) vers F et si D est dense dans E, alors il existe une unique application linéaire continue qui prolonge T sur E, c'est-à-dire qu'il existe une unique application linéaire $\widetilde{T} : E \to F$ continue et telle que $\forall x \in D, \widetilde{T}(x) = T(x)$. De plus,*

$$\|\widetilde{T}\|_{\mathcal{L}(E,F)} = \|T\|_{\mathcal{L}(D,F)}. \tag{2.8}$$

Démonstration. L'application T étant linéaire continue sur D, elle est lipschitzienne et donc en particulier uniformément continue sur D. On peut donc appliquer le théorème 1.87 (qui utilise la complétude de l'espace d'arrivée ainsi que la densité de D) qui donne l'existence d'une application $\widetilde{T} : E \to F$ qui prolonge T. Par densité de D dans E, on peut étendre la linéarité de T sur D à la linéarité de $\widetilde{T}$ sur E : en effet si $(x, y) \in E$, il existe $(x_n), (y_n)$ deux suites d'éléments de D qui convergent respectivement vers x et y. Mais alors

$$\forall \lambda \in \mathbb{K}, \quad T(\lambda x_n + y_n) = \lambda T(x_n) + T(y_n)$$

ce qui donne à la limite

$$\forall \lambda \in \mathbb{K}, \quad \widetilde{T}(\lambda x + y) = \lambda \widetilde{T}(x) + \widetilde{T}(y).$$

Il reste à montrer l'égalité des normes (2.8) : comme $\widetilde{T}$ prolonge T, on obtient la première inégalité :

$$\|T\|_{\mathcal{L}(D,F)} = \sup_{x \in D, \|x\|_E \leq 1} \|T(x)\|_F \leq \sup_{x \in E, \|x\|_E \leq 1} \|\widetilde{T}(x)\|_F = \|\widetilde{T}\|_{\mathcal{L}(E,F)}.$$

Pour l'autre inégalité, on écrit que si $x \in E$ alors il existe (x_n) une suite d'éléments de D qui converge vers x, et on peut écrire :

$$\forall n \in \mathbb{N}, \quad \|T(x_n)\|_F \leq \|T\|_{\mathcal{L}(D,F)} \|x_n\|_F, \qquad \text{d'où à la limite} \quad \|\widetilde{T}(x)\|_F \leq \|T\|_{\mathcal{L}(D,F)} \|x\|_F$$

ce qui donne bien $\|\widetilde{T}\|_{\mathcal{L}(E,F)} \leq \|T\|_{\mathcal{L}(D,F)}$ et conclut la preuve. $\square$

Exemple 2.52. Comme on l'avait évoqué au paragraphe 5.1.2 du Chapitre I, on a en fait déjà démontré un cas particulier de ce théorème quand on a construit l'intégrale de Riemann des fonctions continues par morceaux ou des fonctions réglées. En effet, si on considère

$$\begin{array}{cccc} I: & (D, \|\cdot\|_\infty) & \longrightarrow & F \\ & f & \longmapsto & \displaystyle\int_a^b f(t)dt \end{array}$$

où $D = \{$fonctions en escaliers définies de $[a,b]$ dans $\mathbb{R}\}$ et $F = \mathbb{R}$, on peut vérifier que I est une application linéaire continue sur D. On peut considérer

$$E = C^0_m([a,b]) \qquad \text{ou} \qquad E = \{\text{fonctions réglées de } [a,b] \text{ dans } \mathbb{R}\}$$

muni de la norme uniforme, dans les deux cas D est dense dans E (voir le lemme 5.4 du Chapitre I pour le premier cas ; dans le second cas, c'est par définition même des fonctions réglées que nous avons donnés), et le corollaire 2.51 montre ainsi qu'il existe un unique prolongement linéaire continu de I à E, et c'est cet objet qu'on appelle intégrale de Riemann ; voir aussi la proposition 5.8 du Chapitre I. Remarquons enfin qu'on aurait pu remplacer $\mathbb{R}$ comme espace d'arrivée par n'importe quel espace de Banach F.

2.5.3 Exemples

Le résultat suivant s'intéresse au cas où l'espace de départ est de dimension finie :

Théorème 2.53. *Soit E, F deux espaces vectoriels normés. Si E est de dimension finie, alors toutes les applications linéaires de E vers F sont continues.*

Démonstration. Soit $T : E \to F$ une application linéaire. On introduit $(e_i)_{i\in[\![1,n]\!]}$ une base de E. Alors

$$\forall x = \sum_{i=1}^n x_i e_i, \quad \|T(x)\|_F \le \sum_{i=1}^n |x_i| \|T(e_i)\|_F \le \left(\sum_{i=1}^n \|T(e_i)\|_F\right) \max_{i\in[\![1,n]\!]} |x_i|. \tag{2.9}$$

On peut munir E de la norme $\|x\|_\infty := \max_{i\in[\![1,n]\!]} |x_i|$ qui est équivalente à la norme initiale de E, puisque ce dernier est de dimension finie, voir le théorème 2.25. L'inégalité (2.9) montre alors bien que T est continue. □

Exemple 2.54. Lorsqu'on considère les endormorphismes de $\mathbb{K}^n$, on est naturellement à considérer les matrices associées. Si on munit $\mathbb{K}^n$ d'une norme $\|\cdot\|$, on en déduit donc une norme sur l'espace des matrices (parfois notée avec une triple barre) :

$$\forall M \in \mathcal{M}_n(\mathbb{K}), \quad |\!|\!| M |\!|\!| = \sup_{x\in\mathbb{K}^n, \|x\|\le 1} \|Mx\| = \sup_{x\in\mathbb{K}^n\setminus\{0\}} \frac{\|Mx\|}{\|x\|}, \tag{2.10}$$

et on la nomme norme subordonnée à $\|\cdot\|$. On calcul à l'exercice suivant cette norme pour trois choix classiques de normes sur $\mathbb{K}^n$.

Exercice 2.55 (Exemples de normes matricielles)**.** Si $\mathbb{K}^n$ est muni des normes $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_\infty$, alors on note respectivement $|\!|\!|\cdot|\!|\!|_1, |\!|\!|\cdot|\!|\!|_2, |\!|\!|\cdot|\!|\!|_\infty$ (définie en (2.10)). Montrer que pour tout $M \in \mathcal{M}_n(\mathbb{K})$,

1. $|\!|\!| M |\!|\!|_1 = \max_{j\in[\![1,n]\!]} \left[\sum_{i=1}^n |m_{ij}|\right]$.
2. $|\!|\!| M |\!|\!|_2 = \sqrt{\rho(M^*.M)}$ où $\rho(A)$ est le rayon spectral d'une matrice $A \in \mathcal{M}_n(\mathbb{K})$, à savoir
$$\rho(A) := \max\Big\{|\lambda|, \ \ \lambda \text{ valeur propre complexe de } A\Big\}.$$
3. $|\!|\!| M |\!|\!|_\infty = \max_{i\in[\![1,n]\!]} \left[\sum_{j=1}^n |m_{ij}|\right]$.

Exercice 2.56 (Remarques numériques sur les normes matricielles)**.** Soit $A \in \mathcal{M}_n(\mathbb{C})$.

1. Montrer que si $|||\cdot|||$ est une norme subordonnée (à une norme $\|\cdot\|$ sur $\mathbb{C}^n$) et si $|||A||| < 1$, alors $A^k \xrightarrow[k \to +\infty]{} 0$.
2. Montrer que si $|||\cdot|||$ est une norme subordonnée, alors
$$\rho(A) \leq |||A|||.$$
3. (*voir [Dem16, Page 106]) Montrer que pour tout $\varepsilon > 0$ il existe $|||\cdot|||$ une norme subordonnée [45] telle que
$$|||A||| \leq \rho(A) + \varepsilon.$$
Autrement dit on a démontré
$$\rho(A) = \inf \Big\{ |||A|||, \ \ |||\cdot||| \text{ norme subordonnée sur } \mathcal{M}_n(\mathbb{C}) \Big\}.$$
4. En déduire les équivalences suivantes :
$$\lim_{k \to +\infty} A^k = 0 \quad \Longleftrightarrow \quad \rho(A) < 1.$$

3 Espace des fonctions continues sur un compact

Dans ce chapitre, nous nous concentrons sur l'étude d'un espace vectoriel normé de dimension infinie, à savoir celui des fonctions continues. Soit $[a, b]$ ($a \leq b$ deux réels) un intervalle fermé borné, donc compact de $\mathbb{R}$. On note
$$C^0([a, b]) = \{f \colon [a, b] \to \mathbb{R} \text{ continue}\}.$$
Il s'agit clairement d'un espace vectoriel. Pour tout $f \in C^0([a, b])$, on note
$$\|f\|_\infty = \sup_{x \in [a,b]} |f(x)|.$$

Avec le théorème des bornes 3.13 du Chapitre I, cette quantité est toujours finie quel que soit $f \in C^0([a, b])$. On peut même dire que le supremum est atteint, et est donc un maximum. On montre par ailleurs aisément qu'elle définit une norme sur $C^0([a, b])$, ce qui fait de $C^0([a, b])$ un espace vectoriel normé. La convergence d'une suite au sens de cette norme n'est autre que sa convergence uniforme sur l'intervalle $[a, b]$.

Exercice 3.1. Soit $E = C^0([a, b])$ muni de la norme uniforme.

1. Soit $F = \{f \in E, \ f \geq 0 \text{ sur } [a, b]\}$. Montrer que F est fermé. Décrire le complémentaire de F.
2. Soit $O = \{f \in E, \ f > 0 \text{ sur } [a, b]\}$. Montrer que O est ouvert (on pourra faire deux preuves, une directe, une qui consiste à montrer que O^c est fermé).
3. On remplace E par $E = C^0_b(\mathbb{R}_+)$ (espace des fonctions continues bornées sur $\mathbb{R}_+$), et on considère cette fois-ci $O = \{f \in E, \ f > 0 \text{ sur } \mathbb{R}_+\}$. Cet ensemble est-il ouvert ?

3.1 Propriétés

Proposition 3.2. *L'espace $C^0([a, b])$ est un espace de Banach.*

Remarque 3.3. On a vu ce résultat à l'exercice 2.16, en montrant que l'ensemble des fonctions bornées muni de la norme uniforme est un espace de Banach, et en concluant par le fait que $C^0([a, b])$ est un sous-espace fermé de l'ensemble des fonctions bornées. On reproduit ici une preuve directe ; les arguments sont les mêmes, simplement présentés un peu différemment.

45. Attention, cette norme subordonnée dépend de ε et de A !

Démonstration. Il s'agit de montrer que $C^0([a,b])$ est complet pour la distance induite par sa norme. Pour ce faire, considérons une suite de Cauchy $(f_n)_{n\in\mathbb{N}}$ dans $C^0([a,b])$ et montrons qu'elle converge uniformément sur $[a,b]$ vers une fonction $f \in C^0([a,b])$. Par définition, pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que pour tout $p \geq N$ et tout $q \geq N$, $\|f_q - f_p\|_\infty \leq \varepsilon$. On en déduit que, pour tout $x \in [a,b]$,

$$|f_q(x) - f_p(x)| \leq \|f_q - f_p\|_\infty \leq \varepsilon.$$

Il s'ensuit que la suite numérique $(f_n(x))_{n\in\mathbb{N}}$ est de Cauchy dans $\mathbb{R}$, lequel est complet. Ceci assure l'existence d'un nombre réel que l'on note $f(x) \in \mathbb{R}$ tel que $f_n(x) \xrightarrow[n\to+\infty]{} f(x)$ dans $\mathbb{R}$. Par passage à la limite quand $q \to +\infty$ dans l'inégalité précédente, on en déduit que $|f - f_p|$ est borné sur $[a,b]$, puis par passage au sup en x, que pour tout $p \geq N$,

$$\|f - f_p\|_\infty \leq \varepsilon,$$

ce qui assure que $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers f sur $[a,b]$. Il reste à montrer que la fonction f est continue. Mais comme il s'agit d'une limite uniforme de fonctions continues, par le théorème 4.10 du Chapitre I, c'est bien le cas. □

Poursuivons avec un résultat célèbre de densité :

Théorème 3.4 (Théorème de Weierstrass). *Pour toute fonction $f \in C^0([a,b])$, il existe une suite de fonctions polynomiales $(P_n)_{n\in\mathbb{N}}$ telle que $P_n \xrightarrow[n\to+\infty]{} f$ uniformément sur $[a,b]$.*

Remarque 3.5. Ce théorème possède de nombreuses démonstrations. Nous donnons ici une version relativement classique, due à Bernstein (les polynômes qui apparaissent dans la preuve sont appelés polynômes de Bernstein), et qui est constructive (les polynômes qui approchent f sont explicites). On peut trouver cette preuve par exemple dans [Gou08b, page 231].

Cette preuve a une interprétation probabiliste ; notamment en faisant apparaitre la loi binomiale, on peut simplifier certains calculs en reconnaissant espérance et variance de cette loi. Notez quand même que cette version de la preuve, désormais très classique à l'agrégation, nécessite de connaître l'existence d'une suite de variable indépendante identiquement distribuée ayant une loi donnée, résultat qu'on est prêt à accepter au regard du programme, mais qui n'est pas trivial. On trouvera cette version de la preuve par exemple dans [QZ13, page 518]

Démonstration. On commence par supposer $[a,b] = [0,1]$. Soit $n \in \mathbb{N}$. Pour tout $0 \leq j \leq n$, on définit les polynômes de Bernstein par

$$B_j^n(x) = \binom{n}{j} x^j (1-x)^{n-j}, \text{ pour tout } x \in [0,1].$$

On remarque que $B_j^n \geq 0$ sur $[0,1]$ pour tout $0 \leq j \leq n$. Enfin on définit la fonction polynomiale

$$P_n(x) = \sum_{j=0}^{n} f\Big(\frac{j}{n}\Big) B_j^n(x).$$

À l'aide de la formule du binôme de Newton, on voit que $\sum_{j=0}^n B_j^n = 1$, et donc

$$|f(x) - P_n(x)| = \Big| f(x) - \sum_{j=0}^{n} f\Big(\frac{j}{n}\Big) B_j^n(x) \Big| \leq \sum_{j=0}^{n} \Big| f(x) - f\Big(\frac{j}{n}\Big) \Big| B_j^n(x). \tag{3.1}$$

La fonction f étant continue sur le compact $[0,1]$, elle y est uniformément continue par le théorème de Heine 3.18 du Chapitre I. Soit $\varepsilon > 0$. Il existe donc $\delta > 0$ tel que,

$$\forall (x,y) \in [0,1]^2, \;\; |y - x| \leq \delta \Rightarrow |f(y) - f(x)| \leq \frac{\varepsilon}{2}.$$

Pour $x \in [0,1]$ fixé, on définit deux ensembles d'indices complémentaires l'un de l'autre :

$$I = \Big\{ j = 0, \dots, n : \Big| x - \frac{j}{n} \Big| < \delta \Big\} \text{ et } J = \Big\{ j = 0, \dots, n : \Big| x - \frac{j}{n} \Big| \geq \delta \Big\}.$$

En utilisant le fait que $B_j^n \geq 0$ et $\sum_{j=0}^n B_j^n = 1$, on obtient d'une part que

$$\sum_{j\in I}\left|f(x) - f\left(\frac{j}{n}\right)\right|B_j^n(x) \leq \frac{\varepsilon}{2}\sum_{j\in I} B_j^n(x) \leq \frac{\varepsilon}{2}. \tag{3.2}$$

Par ailleurs, pour $j \in J$ on a $1 \leq \left(\frac{nx-j}{n\delta}\right)^2$, d'où

$$\begin{aligned}\sum_{j\in J}\left|f(x) - f\left(\frac{j}{n}\right)\right|B_j^n(x) &\leq 2\|f\|_\infty \sum_{j\in J} B_j^n(x)\\ &\leq \frac{2\|f\|_\infty}{n^2\delta^2}\sum_{j\in J}(nx-j)^2 B_j^n(x) \leq \frac{2\|f\|_\infty}{n^2\delta^2}\sum_{j=0}^n (nx-j)^2 B_j^n(x).\end{aligned}$$

Admettons l'espace d'un instant que

$$\sum_{j=0}^n (nx-j)^2 B_j^n(x) = nx(1-x), \tag{3.3}$$

alors on obtient que

$$\sum_{j\in J}\left|f(x) - f\left(\frac{j}{n}\right)\right|B_j^n(x) \leq \frac{2\|f\|_\infty}{n^2\delta^2} nx(1-x) \leq \frac{\|f\|_\infty}{2n\delta^2}, \tag{3.4}$$

car la fonction $x \mapsto x(1-x)$ atteint son maximum sur $[0,1]$ en $x = 1/2$. En regroupant (3.1), (3.2) et (3.4), il vient, pour tout $x \in [0,1]$

$$|f(x) - P_n(x)| \leq \frac{\varepsilon}{2} + \frac{\|f\|_\infty}{2n\delta^2}.$$

On prend $n \geq \frac{\|f\|_\infty}{\varepsilon\delta^2}$, on a donc $\|f - P_n\|_\infty \leq \varepsilon$ et ε étant arbitraire, on constate effectivement que la suite (P_n) converge uniformément vers f sur $[0,1]$.

Il reste à montrer l'identité (3.3). Pour ce faire, on développe le carré

$$\begin{aligned}\sum_{j=0}^n (j-nx)^2 B_j^n(x) &= \sum_{j=0}^n (j^2 - 2njx + n^2x^2)B_j^n(x)\\ &= \sum_{j=0}^n j^2 B_j^n(x) - 2nx\sum_{j=0}^n jB_j^n(x) + n^2x^2\\ &= \sum_{j=0}^n j(j-1)B_j^n(x) + (1-2nx)\sum_{j=0}^n jB_j^n(x) + n^2x^2.\end{aligned}$$

en jouant avec le fait que $\sum_{j=0}^n B_j^n = 1$ et en ajoutant et retranchant $\sum_{j=0}^n jB_j^n$. En remarquant que pour $j \geq 1$ et $n \geq 1$, $j\binom{n}{j} = n\binom{n-1}{j-1}$, on voit que

$$\begin{aligned}\sum_{j=0}^n jB_j^n(x) &= \sum_{j=1}^n j\binom{n}{j}x^j(1-x)^{n-j} = \sum_{j=1}^n nx\binom{n-1}{j-1}x^{j-1}(1-x)^{(n-1)-(j-1)}\\ &= nx(x+1-x)^{n-1} = nx.\end{aligned}$$

Enfin, en écrivant dans la même veine pour $j \geq 2$ et $n \geq 2$, $j(j-1)\binom{n}{j} = n(n-1)\binom{n-2}{j-2}$, on obtient

$$\begin{aligned}\sum_{j=0}^n j(j-1)B_j^n(x) &= \sum_{j=2}^n n(n-1)\binom{n-2}{j-2}x^j(1-x)^{n-j}\\ &= n(n-1)x^2\sum_{j=2}^n \binom{n-2}{j-2}x^{j-2}(1-x)^{(n-2)-(j-2)}\\ &= n(n-1)x^2(x+1-x)^{n-2} = n(n-1)x^2.\end{aligned}$$

Finalement, on en déduit que

$$\sum_{j=0}^{n}(j-nx)^2 B_j^n(x) = n(n-1)x^2 + (1-2nx)nx + n^2x^2 = nx(1-x),$$

ce qui conclut la preuve du théorème dans le cas $[a,b]=[0,1]$.

Si $f \in C^0([a,b])$ avec $a<b$, on considère un changement de variable affine et bijectif entre $[a,b]$ et $[0,1]$: plus précisément on pose $g : s \in [0,1] \mapsto f((b-a)s+a)$ qui est encore continue. Par le cas précédent, il existe $(P_n)_{n\in\mathbb{N}}$ une suite de polynômes qui converge vers g uniformément sur $[0,1]$. On pose alors $Q_n(x) = P_n\left(\frac{x-a}{b-a}\right)$ qui est bien une suite de polynômes. On a alors

$$\forall n \in \mathbb{N}, \|Q_n - f\|_{\infty,[a,b]} = \|P_n - g\|_{\infty,[0,1]}$$

et donc la suite de polynômes $(Q_n)_{n\in\mathbb{N}}$ converge bien vers f uniformément. □

Remarque 3.6. Une question naturelle est de se poser la question de la vitesse de convergence de cette suite de polynômes. Dans les deux références que nous avons citées ([Gou08b, QZ13]), cette question est abordée, encore une fois avec une approche probabiliste dans le cas de [QZ13].

3.2 Critère de compacité

Nous établissons pour finir un critère de compacité dans l'espace $C^0([a,b])$. C'est un espace vectoriel normé de dimension infinie, donc le caractère fermé borné ne suffit pas.

Théorème 3.7 (Théorème d'Ascoli-Arzela). *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions de $C^0([a,b])$ telle que*

i) (bornitude) il existe $M>0$ telle que $\sup_n \|f_n\|_\infty \le M$;

ii) (uniforme équicontinuité) pour tout $\varepsilon>0$, il existe un $\delta>0$ tel que pour tout x, $y \in [a,b]$,

$$|x-y| \le \delta \Longrightarrow \sup_{n\in\mathbb{N}} |f_n(x) - f_n(y)| \le \varepsilon.$$

Alors, il existe une sous-suite $(f_{\varphi(n)})_{n\in\mathbb{N}}$ et une fonction $f \in C^0([a,b])$ telles que $f_{\varphi(n)}$ converge uniformément vers f sur $[a,b]$.

Réciproquement, si $(f_n)_{n\in\mathbb{N}}$ est une suite de fonctions de $C^0([a,b])$ qui converge uniformément, alors elle est bornée et uniformément équicontinue.

Démonstration. **Étape 1 : définition de la fonction f sur $D = \mathbb{Q}\cap[a,b]$.** L'ensemble D étant dénombrable, on peut énumérer ses éléments en une suite $(a_i)_{i\in\mathbb{N}}$. D'après l'hypothèse *i)*, pour tout $i \in \mathbb{N}$, la suite numérique $(f_n(a_i))_{n\in\mathbb{N}}$ est bornée. Nous allons appliquer le procédé d'extraction diagonale de sous-suite. On raisonne par récurrence. Pour $i=0$, d'après le théorème de Bolzano-Weierstrass, il existe une sous-suite $\varphi_0 \colon \mathbb{N} \to \mathbb{N}$ et une valeur que l'on décide de noter $f(a_0) \in \mathbb{R}$, telles que $f_{\varphi_0(n)}(a_0) \to f(a_0)$ quand $n \to +\infty$.

Pour un certain $i \in \mathbb{N}$, on suppose avoir construit des extractions $\varphi_0,\dots,\varphi_i \colon \mathbb{N}\to\mathbb{N}$ strictement croissantes et des réels que l'on note $f(a_0),\dots,f(a_i) \in \mathbb{R}$ tels que

$$f_{\varphi_0\circ\cdots\circ\varphi_k(n)}(a_k) \to f(a_k) \quad \text{pour tout } 0 \le k \le i.$$

La suite $(f_{\varphi_0\circ\cdots\circ\varphi_i(n)}(a_{i+1}))_{n\in\mathbb{N}}$ étant bornée, d'après le théorème de Bolzano-Weierstrass, on peut de nouveau en extraire une sous-suite φ_{i+1} telle que $(f_{\varphi_0\circ\cdots\circ\varphi_i\circ\varphi_{i+1}(n)}(a_{i+1}))_{n\in\mathbb{N}}$ converge vers un certain $f(a_{i+1}) \in \mathbb{R}$, ce qui conclut la construction récurrente.

Pour tout $n \in \mathbb{N}$, posons alors

$$\varphi(n) = \varphi_0 \circ \cdots \circ \varphi_n(n),$$

dont on vérifie rapidement (voir le lemme 3.12 ci-dessous) qu'il s'agit bien d'une extraction, c'est-à-dire une fonction de $\mathbb{N}$ dans $\mathbb{N}$ strictement croissante. Ainsi, pour tout $i\in\mathbb{N}$, la suite $(f_{\varphi(n)}(a_i))_{n\ge i}$ est une sous-suite de la suite $(f_{\varphi_0\circ\cdots\circ\varphi_i(n)}(a_i))_{n\ge i}$. Par conséquent,

$$\lim_{n\to+\infty} f_{\varphi(n)}(a_i) = f(a_i) \text{ pour tout } i \in \mathbb{N}. \tag{3.5}$$

Étape 2 : convergence simple. Montrons que pour tout $x \in [a,b]$, la suite $(f_{\varphi(n)}(x))_{n\in\mathbb{N}}$ est de Cauchy dans $\mathbb{R}$. Soit $\varepsilon > 0$, et notons $\delta > 0$ venant de la définition de l'uniforme équicontinuité. Par densité de D dans $[a,b]$, il existe un $a_i \in D$ tel que $|x - a_i| \leq \delta$. Par conséquent, pour tout n et $m \in \mathbb{N}$,

$$\begin{aligned}|f_{\varphi(n)}(x) - f_{\varphi(m)}(x)|& \\ &\leq |f_{\varphi(n)}(x) - f_{\varphi(n)}(a_i)| + |f_{\varphi(n)}(a_i) - f_{\varphi(m)}(a_i)| + |f_{\varphi(m)}(a_i) - f_{\varphi(m)}(x)| \\ &\leq 2\varepsilon + |f_{\varphi(n)}(a_i) - f_{\varphi(m)}(a_i)|.\end{aligned}$$

Comme $a_i \in D$, d'après l'étape 1, la suite numérique $(f_{\varphi(n)}(a_i))_{n\in\mathbb{N}}$ converge et donc est de Cauchy. Il existe donc un $N \in \mathbb{N}$ tel que pour tous m, $n \geq N$, on a $|f_{\varphi(n)}(a_i) - f_{\varphi(m)}(a_i)| \leq \varepsilon$, ce qui implique que

$$|f_{\varphi(n)}(x) - f_{\varphi(m)}(x)| \leq 3\varepsilon.$$

Ceci montre effectivement que $(f_{\varphi(n)}(x))_{n\in\mathbb{N}}$ est de Cauchy dans $\mathbb{R}$ et donc qu'il existe une valeur que l'on décide de noter $f(x) \in \mathbb{R}$ telle que $f_{\varphi(n)}(x) \to f(x)$. On a ainsi construit une limite ponctuelle $f\colon [a,b] \to \mathbb{R}$ pour la suite $f_{\varphi(n)}$.

Étape 3 : uniforme continuité de f. D'après la propriété d'uniforme équicontinuité de f, pour tout $\varepsilon > 0$, il existe un $\delta > 0$ tel que pour tout x, $y \in [a,b]$ avec $|x - y| \leq \delta$, alors

$$|f_{\varphi(n)}(x) - f_{\varphi(n)}(y)| \leq \varepsilon.$$

Par passage à la limite quand $n \to +\infty$, on obtient que

$$|f(x) - f(y)| \leq \varepsilon,$$

ce qui montre bien l'uniforme continuité de f sur $[a,b]$. En particulier, la fonction f ainsi construite est continue sur $[a,b]$.

Étape 4 : convergence uniforme. Soit $\varepsilon > 0$, et notons $\delta > 0$ venant de la définition de l'uniforme équicontinuité. Par compacité de l'intervalle $[a,b]$, il existe un entier $N_\varepsilon \in \mathbb{N}$ et $a_1, \ldots, a_{N_\varepsilon} \in D$ tels que $[a,b] \subset \bigcup_{i=1}^{N_\varepsilon}]a_i - \delta/2, a_i + \delta/2[$. Donc si $x \in [a,b]$, il existe $i \in \{1, \ldots, N_\varepsilon\}$ tel que $x \in]a_i - \delta/2, a_i + \delta/2[$ et

$$\begin{aligned}|f_{\varphi(n)}(x) - f(x)| &\leq |f_{\varphi(n)}(x) - f_{\varphi(n)}(a_i)| + |f_{\varphi(n)}(a_i) - f(a_i)| + |f(a_i) - f(x)| \\ &\leq 2\varepsilon + \max_{1\leq i\leq N_\varepsilon} |f_{\varphi(n)}(a_i) - f(a_i)|,\end{aligned}$$

où l'on a utilisé l'uniforme équicontinuité de la suite $(f_{\varphi(n)})_{n\in\mathbb{N}}$ et l'uniforme continuité de f établie à l'étape 3. D'après l'étape 1, on en déduit que $|f_{\varphi(n)}(a_i) - f(a_i)| \leq \varepsilon$ dès lors que $n \geq n_i$ (qui ne dépend que de ε et de a_i). En notant $n_\varepsilon = \max\{n_1, \ldots, n_{N_\varepsilon}\}$, il vient : pour tout $\varepsilon > 0$, il existe $n_\varepsilon \in \mathbb{N}$ tel que $|f_{\varphi(n)}(x) - f(x)| \leq 3\varepsilon$ pour tout $n \geq n_\varepsilon$ et tout $x \in [a,b]$. On en déduit la convergence uniforme de $f_{\varphi(n)}$ vers f sur $[a,b]$.

Étape 5 : Réciproque. Soit $(f_n)_{n\in\mathbb{N}}$ est une suite de fonctions de $C^0([a,b])$ qui converge uniformément vers une fonction $f \in C^0([a,b])$. Alors $(f_n)_{n\in\mathbb{N}}$ est bornée dans $C^0([a,b])$. Par ailleurs, pour tout $\varepsilon > 0$ il existe $N \in \mathbb{N}$ tel que $\|f_n - f\|_\infty \leq \varepsilon/3$ pour tout $n \geq N$. Comme les fonctions f, $f_0, \ldots, f_N$ sont continues sur le compact $[a,b]$, elles sont uniformément continues. Par conséquent, il existe $\eta > 0$ et $\delta_0, \ldots, \delta_N > 0$ tels que pour tout $x, y \in [a,b]$

$$|x - y| \leq \eta \Longrightarrow |f(x) - f(y)| \leq \frac{\varepsilon}{3}$$

et

$$|x - y| \leq \delta_i \Longrightarrow |f_i(x) - f_i(y)| \leq \frac{\varepsilon}{3}, \quad \text{pour tout } 0 \leq i \leq N.$$

On définit $\delta = \min(\eta, \delta_0, \ldots, \delta_N)$ de sorte que si x et $y \in [a,b]$, alors

$$|x - y| \leq \delta \Longrightarrow |f(x) - f(y)| \leq \frac{\varepsilon}{3} \text{ et } |f_i(x) - f_i(y)| \leq \frac{\varepsilon}{3} \quad \text{pour tout } 0 \leq i \leq N.$$

Ainsi, pour tout $n \geq N$ et pour tout $x, y \in [a,b]$ tel que $|x-y| \leq \delta$, on a

$$|f_n(x) - f_n(y)| \leq |f_n(x) - f(x)| + |f(x) - f(y)| + |f(y) - f_n(y)| \\ \leq 2\|f - f_n\|_\infty + |f(x) - f(y)| \leq \varepsilon.$$

On obtient finalement que la famille $(f_n)_{n\in\mathbb{N}}$ est bien uniformément équicontinue. □

Remarque 3.8. L'énoncé du théorème peut être amélioré en constatant qu'on peut remplacer l'hypothèse d'être uniformément borné par ponctuellement borné, et l'hypothèse d'uniforme équicontinuité par la simple équicontinuité. Plus précisément, on peut traiter l'exercice suivant :

Exercice 3.9. Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions de $C^0([a,b])$ telle que

i') pour tout $x \in [a,b]$, il existe $M_x > 0$ tel que pour tout $n \in \mathbb{N}, |f_n(x)| \leq M_x$ *(on dit que la suite est ponctuellement bornée)*

ii') pour tout $x \in [a,b]$ et $\varepsilon > 0$, il existe $\delta > 0$ tel que pour tout $y \in [a,b]$,

$$|x-y| \leq \delta \Longrightarrow \sup_{n\in\mathbb{N}} |f_n(x) - f_n(y)| \leq \varepsilon$$

(qui correspond à la (simple) équicontinuité).

1. Montrer que $(f_n)_{n\in\mathbb{N}}$ est uniformément bornée, c'est-à-dire qu'il existe M telle que

$$\forall n \in \mathbb{N}, \|f_n\|_\infty \leq M.$$

2. [Généralisation du théorème de Heine] Montrer que la suite $(f_n)_{n\in\mathbb{N}}$ est uniformément équicontinue.

En conséquence, le théorème d'Ascoli reste valable en remplaçant $i), ii)$ par $i'), ii')$.

Remarque 3.10. On a donné une formulation du théorème pour des suites. On aurait pu donner une formulation pour des sous-ensembles de $C^0([a,b])$ plus généraux. Combiné à la remarque 3.8, on donne une autre version du théorème d'Ascoli, qui se déduit immédiatement du théorème 3.7 :

Théorème 3.11. *Soit $A \subset C^0([a,b])$. Alors*

A est relativement compacte[46] *dans $C^0([a,b]) \Longleftrightarrow A$ est pontuellement bornée et équicontinue*

c'est-à-dire satisfait

i) pour tout $x \in [a,b]$, il existe $M_x \in \mathbb{R}$ tel que $\forall f \in A, |f(x)| \leq M_x$,

ii) pour tout $x \in [a,b]$ et $\varepsilon > 0$, il existe $\delta > 0$ tel que pour tout $y \in [a,b]$,

$$|x-y| \leq \delta \Longrightarrow \forall f \in A, |f(y) - f(x)| \leq \varepsilon$$

Autrement dit, les compacts de $C^0([a,b])$ sont exactements les parties fermées, bornées et équicontinues.

Commentaires sur la preuve du théorème d'Ascoli : La preuve donnée ci-dessus est très naturelle et chaque étape est intéressante en soi :

1. Le première étape présente le concept d'extraction diagonale. Le point de départ ressemble à la preuve de la proposition 1.106 sur la compacité d'un produit fini de compacts. L'idée est d'avoir un nombre dénombrable de suites $\forall i \in \mathbb{N}, (x_n^{(i)})_{n\in\mathbb{N}}$ telle que pour chaque $i \in \mathbb{N}$, on peut extraire une sous-suite convergente de $(x_n^{(i)})_{n\in\mathbb{N}}$, et on veut pouvoir construire une extraction valable pour tous les i. Dans la preuve, on a posé

$$\forall n \in \mathbb{N}, \quad \varphi(n) = \varphi_0 \circ \ldots \circ \varphi_n(n) \tag{3.6}$$

qui convient par le lemme suivant :

46. Un ensemble est dit relativement compact si son adhérence est compacte.

Lemme 3.12. *Si $(\varphi_i)_{i\in\mathbb{N}}$ est une famille de fonctions strictement croissantes de $\mathbb{N}$ dans $\mathbb{N}$, alors φ définie par (3.6) est une fonction strictement croissante de $\mathbb{N}$ dans $\mathbb{N}$.*

Attention, ce lemme ne peut pas se résumer à “une composée de fonctions strictement croissantes est strictement croissante”, car φ n'est pas une composition, le nombre de fonctions intervenant dans la composition dépendant de la variable $n \in \mathbb{N}$.

Démonstration. On commence par constater que φ est bien définie et à valeurs dans $\mathbb{N}$. Ensuite, on écrit

$$\forall n \in \mathbb{N}, \quad \varphi(n+1) = \varphi_0 \circ \ldots \varphi_n(\varphi_{n+1}(n+1)) \geq \varphi_0 \circ \ldots \varphi_n(n+1) > \varphi_0 \circ \ldots \varphi_n(n)$$

où on a utilisé pour la première inégalité que $\varphi_{n+1}(n+1) \geq n+1$ (voir exercice 2.20 au Chapitre I) et ensuite on a utilisé la stricte croissance de $\varphi_0 \circ \ldots \varphi_n$. □

Ce principe est très important pour aborder les questions de compacité en dimension infinie. Dans ce qui précède, on n'a pas vraiment dit où vivaient les suites $(x_n^{(i)})_{n\in\mathbb{N}}$. Dans la preuve du théorème d'Ascoli, il s'agit de $(f_n(a_i))_{n\in\mathbb{N}}$ pour $i \in \mathbb{N}$, qui vivent toutes dans $\mathbb{R}$, mais on pourrait traiter le cas d'espaces plus généraux. En fait, le principe d'extraction diagonale permet d'aborder la question de la compacité d'un produit dénombrable d'espaces métriques, voir le paragraphe 4.2.

2. Les étapes 2 et 3 sont très proches du théorème de prolongement (1.87). D'ailleurs, on aurait pu dans la preuve appliquer ce théorème pour prolonger la limite f de $(a_j)_{j\in\mathbb{N}}$ à $[a,b]$ et ensuite montrer que $(f_n(x))_{n\in\mathbb{N}}$ converge bien vers $f(x)$ pour tout $x \in [a,b]$, mais cela n'aurait pas vraiment simplifié la preuve.
3. L'étape 4 peut se reformuler avec le lemme suivant :

   Lemme 3.13. *Soit $(f_n)_{n\in\mathbb{N}} \in C^0([a,b])$ une suite équicontinue et $f : [a,b] \to \mathbb{R}$. Alors*

   $$f_n \text{ converge simplement vers } f \iff f_n \text{ converge uniformément vers } f.$$

   La preuve est exactement celle donnée ci-dessus à l'étape 4, à ceci près qu'on le fait pour $(f_{\varphi(n)})_{n\in\mathbb{N}}$ au lieu de $(f_n)_{n\in\mathbb{N}}$. Enfin, notons que ce lemme est lié au premier théorème de Dini : en effet, combiné à un autre lemme qui affirme qu'une suite monotone de fonctions continues qui converge simplement vers une fonction continue est équicontinue [47], on en déduit une preuve du premier théorème de Dini.
4. La dernière étape qui consiste à montrer que le critère de compacité du théorème d'Ascoli est optimal, critère qui sert rarement. C'est plutôt pour montrer la force du théorème que l'on énonce cette réciproque.

Concluons avec des exemples très standards de familles pour lesquelles on peut appliquer le théorème précédent :

Exercice 3.14.
1. Soit $M \in \mathbb{R}_+$ et $k \in \mathbb{R}_+$. On pose

   $$\mathrm{Lip}(M,k) = \Big\{ f : [a,b] \to \mathbb{R}, \quad \|f\|_\infty \leq M, \quad \text{et } \ f \text{ est } k\text{-lipschitzienne sur } [a,b] \Big\}.$$

   Montrer que $\mathrm{Lip}(M,k)$ est un compact de $C^0([a,b])$.
2. On travaille dans $C^1([a,b])$ muni de la norme $\forall f \in C^1([a,b]), \|f\|_{C^1} = \|f\|_\infty + \|f'\|_\infty$ (voir l'exercice 2.17). Montrer que pour $M \in \mathbb{R}_+$, les ensembles

   $$\overline{B}(0,M) = \{f \in C^1([a,b]),\ \|f\|_{C^1} \leq M\}$$

   sont relativement compacts dans $(C^0([a,b]), \|\cdot\|_\infty)$ [48].

47. Pour ce lemme on n'a pas besoin de la compacité de l'ensemble de départ.
48. Notez le jeu de normes ici : un ensemble borné pour la norme C^1 sera relativement compact pour la norme C^0. Ces situations sont fréquentes pour pallier au théorème de Riesz de non-compacité des boules pour une norme donnée. On dit que l'injection $C^1([a,b]) \subset C^0([a,b])$ est compacte.

3.3 *Généralisation sur les espaces de départ et d'arrivée

On a volontairement présenté le paragraphe précédent dans un cadre simple où les espaces de départ et d'arrivée sont de dimension 1. Il n'est en effet pas toujours mieux de chercher à tout prix à prendre le cadre le plus général, notamment quand un cas particulier pourrait suffire pour les applications que vous connaissez. Surtout quand les idées principales se retrouvent dans ce cas particulier (par exemple pour le théorème d'Ascoli), il n'y a pas de mal à l'oral à ne présenter que ce dernier. Dans votre apprentissage, il peut aussi être raisonnable de commencer à s'intéresser à ce cas particulier pour éviter de cumuler les difficultés.

Cela n'empêche pas une fois ce cas bien compris, de se demander quels énoncés se généralisent facilement, lesquels se généralisent avec plus d'effort, et lesquels ne se généralisent pas.

— On peut considérer $C^0(X,Y)$ pour X,Y deux espaces métriques.

  — Si F est un espace vectoriel normé, alors $C^0(X,F)$ devient un espace vectoriel (dont le corps de base est le même que celui de F).

  — Si F est un espace vectoriel normé et X est un espace métrique compact, on peut munir $C^0(X,F)$ de la norme uniforme, et l'espace est alors un espace de Banach, voir l'exercice 2.16.

  — Si F est un espace vectoriel normé et X est un espace métrique non compact, il faut se retreindre à l'espace de fonctions continues et bornées, qu'on note par exemple $C^0_b(X,F)$, et qui munit de la norme uniforme est encore un espace de Banach.

— [*]Le théorème de Weierstrass est plus délicat à généraliser. Le théorème suivant donne une généralisation :

  Théorème 3.15 (Théorème de Stone-Weierstrass). *Soit* (X,d) *un espace métrique compact, et* $C^0(X) = C^0(X,\mathbb{R})$. *Soit* A *une sous-algèbre de* $C^0(X)$, *c'est-à-dire un sous-espace vectoriel de* $C^0(X)$ *qui est stable par produit. Alors*

$$A \text{ est dense dans } C^0(X) \iff \begin{cases} A \text{ sépare les points, c'est-à-dire} \\ \qquad\qquad \forall (x,y)\in X,\quad x\neq y,\quad \exists f\in A,\quad f(x)\neq f(y) \\ \forall x\in X, \exists f\in A \text{ telle que } f(x)\neq 0 \end{cases}$$

  Attention, ici le corps d'arrivée était $\mathbb{R}$. Le théorème reste valable si on remplace par $\mathbb{C}$, à condition de demander que A est une sous-algèbre stable par conjugaison. Dans le cas $X=[a,b]$, on peut appliquer ce théorème à A l'ensemble des fonctions polynomiales à coefficients réels (ou complexe) qui est bien une sous-algèbre (stable par conjugaison) qui satisfait les deux conditions requises, ce qui redonne bien le théorème de Weierstrass 3.4.

  Attention également, il existe de nombreuses versions du théorème, certaines plus générales (sur les sous-ensembles réticulés), d'autres un peu moins générales (souvent la deuxième condition est remplacée par le fait que A contient les constantes, mais alors il n'y a plus une équivalence dans l'énoncé). On peut consulter par exemple [HL99, page 27].

  Si vous évoquez un tel énoncé à l'oral, je vous invite à connaître des applications plus élaborées que le théorème de Weierstrass 3.4 : on peut citer la densité des fonctions lipschitziennes de X dans $C^0(X)$ (si X est compact), la densité des polynômes à n variables si X est un compact de $\mathbb{R}^n$, ou encore la densité de l'espace vectoriel engendré par les fonctions à variables séparées $(\{(x,y)\mapsto f(x)g(y), f\in C^0(Y), g\in C^0(Z)\})$ si $X=Y\times Z$ est un produit de compacts.

— [*] Le théorème d'Ascoli peut également être facilement généralisé, mais contrairement au théorème de Stone-Weierstrass, cela recquiert peu d'efforts, on peut en effet garder essentiellement la même preuve que celle donnée au paragraphe 3.2. On renvoit à [Mat] pour plus de détails, et aussi [BB18, II.1 et II.2] [49]

49. Même si dans cette référence, ils ne vont pas jusqu'à "réunir" dans un même énoncé les théorèmes d'Ascoli et de Banach-Alaoglu séparable ; il n'empêche que les méthodes de démonstrations proposées sont précisément les mêmes que celles proposées dans notre preuve du théorème 3.7.

1. En premier lieu, on peut adapter le théorème d'Ascoli et sa preuve à $C^0(K)$ où K est un compact de $\mathbb{R}^n$. L'énoncé et la preuve sont exactement les mêmes, il faut simplement adapter la première étape de la démonstration en voyant que $K \cap \mathbb{Q}^n$ est dénombrable et dense dans K.
2. Ensuite, on peut se demander si on peut traiter $C^0(X)$ où (X, d) est un espace métrique compact. La réponse est oui, et la preuve reste la même, en invoquant simplement à la première étape qu'un espace compact est séparable (voir le paragraphe 4.3).
3. Pour modifier l'espace d'arrivée et traiter $C^0(K, F)$ où F est un espace vectoriel normé, on peut à nouveau adapter le théorème, mais il faut changer l'hypothèse d'être ponctuellement bornée avec :

   i') pour tout $x \in X$, l'ensemble $\{f_n(x), n \in \mathbb{N}\}$ est relativement compact dans F.

   Notez qu'on utilise en apparence la complétude de l'espace F dans l'étapes 2 ; mais on peut en fait se passer de cette hypothèse car l'hypothèse *i')* ci-dessus implique que $\overline{\{f_n(x), n \in \mathbb{N}\}}$ est complet pour la norme de F.
4. On peut enfin se demander si on peut choisir un espace de départ (X, d) non compact. Dans ce cas, comme on le verra au paragraphe 4.4.2, la bonne convergence n'est pas la convergence uniforme sur X, mais la convergence uniforme sur tout compact de X. Pour faire fonctionner l'étape 1 néanmoins, on a besoin de supposer X séparable.

   On obtient donc au final, avec une preuve très similaire, le résultat suivant :

   Théorème 3.16 (Théorème d'Ascoli précisé)**.** *Soit (X, d) un espace métrique séparable, $(F, \|\cdot\|)$ un espace vectoriel normé, et $(f_n)_{n\in\mathbb{N}}$ une suite de $C^0(X, F)$. Si*

   i) pour tout $x \in X$, l'ensemble $\{f_n(x), n \in \mathbb{N}\}$ est relativement compact dans F,

   ii) la suite $(f_n)_{n\in\mathbb{N}}$ est équicontinue,

   alors il existe φ une extraction et $f_\infty \in C^0(X, F)$ telle que $(f_{\varphi(n)})_{n\in\mathbb{N}}$ converge vers f_∞ uniformément sur tout compact de X.

   **Remarquons que cette version du théorème d'Ascoli est très forte, et a pour corollaire le théorème de Banach-Alaoglu dans le cas d'un espace séparable :

   Théorème 3.17 (*Théorème de Banach-Alaoglu dans le cas séparable)**.** *Soit $(E, \|\cdot\|)$ un espace de Banach séparable. Si $(f_n)_{n\in\mathbb{N}}$ est une suite bornée de E', i.e.,*

   $$\sup_{n\in\mathbb{N}} \|f_n\|_{E'} < +\infty,$$

   alors on peut en extraire une sous-suite $(f_{\sigma(n)})_{n\in\mathbb{N}}$ et $f \in E'$ telle que

   $$\forall x \in E, \quad f_{\sigma(n)}(x) \xrightarrow[n\to+\infty]{} f(x),$$

   c'est-à-dire que $(f_{\sigma(n)})_{n\in\mathbb{N}}$ converge faiblement-$*$ vers $f \in E'$.

   Voir par exemple [HL99, page 17] pour une approche plus directe de ce dernier résultat.

4 Sujets de travail et de développement

4.1 *Précompacité

On conseille par exemple [HL99, page 13].

Définition 4.1. Soit (X, d) un espace métrique. On dit qu'une partie A de X est précompacte si pour tout $\rho > 0$, A peut être recouvert par un nombre fini de boules de rayon ρ, c'est-à-dire s'il existe un nombre fini $m \in \mathbb{N}$ de points $x_1, \ldots, x_m \in A$ tels que

$$A \subset \bigcup_{i=1}^{m} B(x_i, \rho).$$

Remarque 4.2. Attention, contrairement à la compacité, cette notion dépend de la métrique, ce n'est pas une notion topologique : par exemple dans $\mathbb{R}$ muni de $d(x,y) = |x-y|$ et de $\delta(x,y) = |\arctan(x) - \arctan(y)|$ pour tout $(x,y) \in \mathbb{R}^2$, on vérifie facilement que

$$(\mathbb{R}, d) \text{ n'est pas précompact, mais } (\mathbb{R}, \delta) \text{ est précompact.}$$

Exemple 4.3. Il est facile de voir avec les définitions que cette notion est plus faible que la compacité : si $A \subset X$, alors

$$A \text{ relativement compact} \implies A \text{ précompacte.}$$

Remarque 4.4. On peut reprendre quelques éléments des sections précédents à la lumière de cette notion :

— L'étape 3 de la preuve du théorème 1.100 consiste à montrer que si un espace est séquentiellement compact, alors il est précompact.

— La preuve du théorème 2.29 consistait à voir que si la boule unité d'un espace vectoriel normé est précompacte, alors l'espace en question est de dimension finie.

On peut en fait montrer le théorème :

Théorème 4.5. *Soit (X,d) un espace métrique. Alors*

$$X \textit{ est compact} \iff X \textit{ est précompact et complet.}$$

Corollaire 4.6. *Soit (X,d) un espace métrique complet. Alors*

$$A \textit{ relativement compact} \iff A \textit{ précompacte.}$$

On peut obtenir une preuve différente du théorème d'Ascoli 3.7 avec la notion de précompacité, voir par exemple [Ska04].

4.2 Produit dénombrable d'espaces métriques

Jusqu'à présent, on a surtout évoqué le cas de produit fini d'espaces métriques ; en fait, à l'image de l'exercice 1.10, étant donnés $(X_n, d_n)_{n\in\mathbb{N}}$ une infinité dénombrable d'espaces métriques, on peut définir [50] sur $X = \prod_{i\in\mathbb{N}} X_n$:

$$\forall (x,y) \in X^2, \quad \delta(x,y) := \sum_{n\in\mathbb{N}} \frac{1}{2^n} \min\big\{1, d_n(x_n,y_n)\big\}.$$

Exercice 4.7. 1. Montrer que δ est une métrique sur X.

2. Montrer qu'étant donnée une suite $(x^{(p)})_{p\in\mathbb{N}}$ d'éléments de X et $x^{(\infty)} \in X$, on a

$$x^{(p)} \xrightarrow[n\to+\infty]{\delta} x^{(\infty)} \iff \Big[\forall n \in \mathbb{N},\ x_n^{(p)} \xrightarrow[n\to+\infty]{d_n} x_n^{(\infty)}\Big].$$

3. Montrer que si les (X_n, d_n) sont complets, alors (X, δ) est complet.

Avec la notion d'extraction diagonale pour la preuve du théorème d'Ascoli, nous pouvons démontrer :

Théorème 4.8 (Théorème de Tychonov dénombrable)**.** *Avec les notations précédentes, on suppose que pour tout $n \in \mathbb{N}$, (X_n, d_n) est dénombrable. Alors*

$$(X,d) \textit{ est compact.}$$

50. Notons que ce n'est évidemment pas la seule définition possible ; un autre choix fréquent dans la littérature est

$$\sum_{n\in\mathbb{N}} \frac{1}{2^n} \frac{d_n(x_n,y_n)}{1+d_n(x_n,y_n)}.$$

4.3 *Séparabilité

Voir par exemple [HL99, page 7].

Définition 4.9. Soit (X, d) un espace métrique. On dit que X est séparable s'il existe $(x_n)_{n\in\mathbb{N}}$ un ensemble dénombrable d'éléments de X qui est dense dans X.

Exercice 4.10.
1. Montrer que si (X, d) est un espace métrique compact, alors X est séparable.
2. Montrer que si (X, d) est séparable et $Y \subset X$, alors Y muni de la métrique induite est séparable.

En ce qui concerne les espaces vectoriels normés de dimension infinie, la séparabilité est un moyen de mesurer que la dimension correspond au plus petit infini possible :

Exercice 4.11. Soit E un $\mathbb{K}$-espace vectoriel normé et D une partie de E. On dit que D est totale si elle engendre une partie dense, c'est-à-dire que pour tout $x \in E$ et tout $\varepsilon > 0$, il existe une famille finie $\{x_1, \dots, x_m\} \in D^m$ et des scalaires $\{\lambda_1, \lambda_2, \dots, \lambda_m\} \in \mathbb{K}^m$ tes que

$$\left\| x - \sum_{i=1}^{m} \lambda_i x_i \right\| \leq \varepsilon.$$

Montrer que E est séparable si et seulement s'il possède une famille totale dénombrable.

Voici une conséquence classique du théorème de Weierstrass :

Proposition 4.12. *L'espace $(C^0([a, b]), \|\cdot\|_\infty)$ est séparable.*

Démonstration. Soit $\mathbf{P}$ l'espace des fonctions polynomiales sur $[a, b]$. Montrons d'abord que $\mathbf{P}$ est séparable pour la topologie de $C^0([a, b])$. On note que $\mathbf{P} = \bigcup_{n\in\mathbb{N}} \mathbf{P}_n$, où $\mathbf{P}_n$, l'ensemble des fonctions polynomiales de degré inférieur ou égal à n, est un espace vectoriel de dimension $n+1$. Par conséquent $\mathbf{P}_n$ est isomorphe à $\mathbb{R}^{n+1}$ et comme $\mathbb{Q}^{n+1}$ est dense dans $\mathbb{R}^{n+1}$, il s'ensuit que l'ensemble des fonctions polynomiales à coefficients rationnels de degré inférieur ou égal à n, noté $\mathbf{Q}_n$, est dense dans $\mathbf{P}_n$ pour la topologie de $C^0([a, b])$ (on utilise ici le fait que toutes les normes sur $\mathbb{R}^{n+1}$ sont équivalentes). Comme $\mathbb{Q}^{n+1}$ est dénombrable et isomorphe à $\mathbf{Q}_n$, on en déduit que $\mathbf{Q}_n$ est dénombrable et donc que $\mathbf{Q} = \bigcup_{n\in\mathbb{N}} \mathbf{Q}_n$ l'est aussi. On vient de voir que $\mathbf{Q}$ est dense dans $\mathbf{P}$, et le théorème de Weierstrass assure que $\mathbf{P}$ est dense dans $C^0([a, b])$. On conclut que $\mathbf{Q}$ est dense dans $C^0([a, b])$ par un argument de double approximation. □

Exercice 4.13. Montrer que $(\ell^\infty(\mathbb{N}), \|\cdot\|_\infty)$ n'est pas séparable.

4.4 *Autres espaces fonctionnels

On pourra consulter [QZ13, Chapitre VIII].

4.4.1 Espaces de fonctions $C^{k,\alpha}$ sur un segment

Dans ce paragraphe, on travaille dans I un intervalle de $\mathbb{R}$ pour simplifier. Notons que pour généraliser en dimension supérieure, il faut faire attention à la définition des fonctions régulières sur un espace fermé, que nous n'abordons pas ici, voir par exemple [QZ13, page 272].

Exercice 4.14. On suppose $I = [a, b]$ un segment de $\mathbb{R}$.

1. Montrer que
$$\|f\|_{C^k([a,b])} := \sum_{i=0}^{k} \|f^{(i)}\|_\infty$$
est une norme sur l'espace $C^k([a, b])$ des fonctions de classe C^k sur $[a, b]$. Lorsqu'il n'y a pas d'ambiguité, on notera seulement $\|\cdot\|_{C^k}$.
2. Étant donné une suite $(f_n)_{n\in\mathbb{N}}$ de fonctions de classe $C^k([a, b])$ et $f_\infty \in C^k([a, b])$, montrer
$$f_n \xrightarrow[n\to+\infty]{\|\cdot\|_{C^k}} f_\infty \iff \forall i \in [\![0, k]\!], f_n^{(i)} \xrightarrow[n\to+\infty]{\|\cdot\|_\infty} f_\infty.$$

3. Montrer que $(C^k([a,b]), \|\cdot\|_{C^k})$ est un espace de Banach.

Définition 4.15. Soit I un intervalle de $\mathbb{R}$.

— Soit $\alpha \in]0,1]$. On dit que $f : I \to \mathbb{R}$ est hölderienne d'exposant α (ou α-hölderienne) s'il existe M tel que
$$\forall (x,y) \in I^2, \quad |f(x) - f(y)| \leq M|x-y|^\alpha, \tag{4.1}$$
et on note $f \in C^{0,\alpha}(I)$.

— Étant donné $k \in \mathbb{N}$, on dit que f est de classe $C^{k,\alpha}$ si f est de classe C^k et si $f^{(k)}$ est α-hölderienne.

Remarque 4.16. Ceci est une généralisation du caractère lipschitzien qui correspond au choix $\alpha = 1$.

Exercice 4.17. 1. Dans cette question nous expliquons la restriction à $\alpha \leq 1$: montrer que si $f : I \to \mathbb{R}$ satisfait (4.1) pour $\alpha > 1$, alors f est constante.

2. Montrer que si $\alpha \in]0,1]$ et si $f \in C^{0,\alpha}(I)$, alors f est uniformément continue sur I.
3. On suppose désormais que $I = [a,b]$ est un segment.
   (a) Soit $\alpha \in]0,1]$. Montrer
   $$C^1([a,b]) \subset C^{0,\alpha}([a,b]) \subset C^0([a,b])$$
   et exhiber des fonctions permettant de voir que ces inclusions sont strictes.

   (b) Pour $\alpha \in]0,1]$, on pose
   $$\forall f \in C^{0,\alpha}([a,b]), \quad \|f\|_{C^{0,\alpha}} = \|f\|_\infty + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x-y|^\alpha}.$$
   Montrer qu'il s'agit d'une norme sur $C^{0,\alpha}([a,b])$ et que ce dernier est complet pour cette norme.

   (c) Montrer que les ensembles bornés de $C^{0,\alpha}([a,b])$ sont relativement compacts dans $C^0([a,b])$.

   (d) Généraliser à $C^{k,\alpha}$ pour $k \in \mathbb{N}$ et $\alpha \in]0,1]$.

Concluons avec un dernier exemple pour lequel on construit une métrique, mais pas une norme :

Exercice 4.18. On pose
$$\forall (f,g) \in C^\infty([a,b]), \quad d(f,g) = \sum_{n \in \mathbb{N}} \min\left\{1, \|(f-g)^{(i)}\|_\infty\right\}.$$

1. Montrer que d est une métrique sur $C^\infty([a,b])$.
2. Soit (f_n) une suite d'éléments de $C^\infty([a,b])$. Montrer que
$$f_n \xrightarrow[n \to +\infty]{d} f_\infty \quad \Longleftrightarrow \quad \left[\forall k \in \mathbb{N}, \; f_n^{(k)} \xrightarrow[n \to +\infty]{\|\cdot\|_\infty} f_\infty^{(k)}\right].$$
3. Montrer que $(C^\infty([a,b]), d)$ est complet.

4.4.2 Espace des fonctions continues sur un ouvert

Soit I un intervalle ouvert de $\mathbb{R}$. Il n'est pas si incongru de travailler avec les fonctions continues sur I. Mais attention, on ne peut pas munir cet espace de la norme uniforme, puisque les fonctions continues sur un intervalle ouvert (borné ou non) ne sont pas bornées.

Exercice 4.19. Soit I un intervalle ouvert de $\mathbb{R}$.

1. Montrer que I peut s'écrire comme l'union dénombrable de segments $(K_n)_{n \in \mathbb{N}}$ tels que pour tout $n \in \mathbb{N}$, $K_n \subset \mathring{K}_{n+1}$.

2. Montrer que
$$\forall n \in \mathbb{N}, \quad p_n(f) := \|f\|_{\infty, K_n}$$
est une famille séparante [51] de semi-normes [52].
3. Montrer que
$$\forall (f,g) \in C^0(I), \quad d(f,g) := \sum_{n \in \mathbb{N}} \frac{1}{2^n} \min\left\{1, p_n(f-g)\right\}$$
est une métrique sur $C^0(I)$.
4. Montrer que la convergence pour la métrique d équivaut à la convergence uniforme sur tout compact de I.
5. Montrer que $(C^0(I), d)$ est un espace complet.
6. Généraliser au cas où on remplace I par un ouvert Ω de $\mathbb{R}^d$. En particulier, on montrera que tout ouvert de $\mathbb{R}^d$ est une union dénombrable de compacts $(K_n)_{n\in\mathbb{N}}$ tels que pour tout $n \in \mathbb{N}$, $K_n \subset \mathring{K}_n$. [53]
7. [*] Dans cette question, on se demande si la convergence uniforme sur tout compact pourrait être induite par une norme.
   (a) Soit $\varepsilon > 0$. Montrer qu'il existe $f \in C^0(I) \setminus \{0\}$ tel que pour tout $\lambda \in \mathbb{R}, d(\lambda f, 0) < \varepsilon$.
   (b) En déduire qu'il n'existe pas de norme sur $C^0(I)$ telle que la convergence pour cette norme soit équivalente à la convergence uniforme sur tout compact de I.
8. Généraliser en exhibant une métrique sur $C^k(I)$ telle que $C^k(I)$ soit un espace complet.

4.4.3 Espace des fonctions holomorphes

De la même manière que pour les fonctions continues sur un ouvert de $\mathbb{R}$, la convergence naturelle pour les fonctions holomorphes est la convergence uniforme sur tout compact :

Exercice 4.20. Soit U un ouvert de $\mathbb{C}$.

1. Montrer que U est une union dénombrable de compacts, qu'on note $(K_n)_{n\in\mathbb{N}}$.
2. On note $H(U)$ l'ensemble des fonctions holomorphes sur U. On pose
$$\forall (f,g) \in H(U), \quad d(f,g) := \sum_{n \in \mathbb{N}} \frac{1}{2^n} \min\left\{1, \|f-g\|_{\infty, K_n}\right\}.$$
Montrer que d est une métrique sur $H(U)$.
3. Montrer que la convergence pour la métrique d équivaut à la convergence uniforme sur tout compact de U.
4. Montrer que $(H(U), d)$ est un espace complet.

Pour conclure, on cite le résultat suivant qui permet de caractériser la compacité dans le cadre des fonctions holomorphes :

51. On dit qu'une famille $(p_i)_{i\in I}$ de semi-normes est séparante si
$$\left[\ \forall i \in\ I, p_i(x) = 0\ \right] \implies x = 0.$$

52. Une application $p : E \to \mathbb{R}_+$ définie sur un espace vectoriel E est une semi-norme si elle satisfait :
(a) $\forall x \in E, \forall \lambda \in \mathbb{K}, \ \ p(\lambda x) = |\lambda| p(x)$,
(b) $\forall (x,y) \in E^2, \ \ p(x+y) \leq p(x) + p(y)$.
c'est-à-dire les mêmes axiomes qu'une norme, sans le caractère "défini".
53. Une telle suite s'appelle une suite exhaustive de compact, car si K est un compact inclus dans Ω, alors il existe $n \in \mathbb{N}$ tel que $K \subset K_n$.

Théorème 4.21 (Théorème de Montel)**.** *Soit U un ouvert de $\mathbb{C}$, et $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions holomorphes sur U. On suppose que pour tout K compact inclus dans U, il existe M_K tel que*

$$\forall n \in \mathbb{N}, \quad \sup_K |f_n| \leq M_K.$$

Alors on peut extraire de $(f_n)_{n\in\mathbb{N}}$ une suite qui converge uniforément sur tout compact de U.

Sketch de preuve : On se donne $(K_p)_{p\in\mathbb{N}}$ une suite exhaustive de compacts, c'est-à-dire telle que pour tout $p \in \mathbb{N}, K_p$ est un compact, $K_p \subset \mathring{K}_{p+1}$, et $\bigcup_{p\in\mathbb{N}} K_p = U$.

Alors étant donné $p \in \mathbb{N}$, la suite $(f_n)_{n\in\mathbb{N}}$ est uniformément bornée sur K_{p+1}. Par les inégalités de Cauchy (voir l'exercice 3.64 au Chapitre II) on en déduit que $(f'_n)_{n\in\mathbb{N}}$ est uniformément bornée sur K_p, et donc $(f_n)_{n\in\mathbb{N}}$ est une suite uniformément lipschitzienne sur K_p. Par le théorème d'Ascoli, on peut extraire une sous-suite qui converge uniformément sur K_p.

Pour avoir une extraction qui convient pour tous les compacts K_p avec $p \in \mathbb{N}$, on fait une extraction diagonale. □

Remarque 4.22. La limite dans le théorème de Montel sera une fonction holomorphe sur U, d'après le théorème 3.47.

Exercice 4.23 (Partie bornée d'un espace vectoriel topologique)**.** Soit (E, d) un espace vectoriel muni d'une métrique, et $A \subset E$. L'ensemble A est dit borné si

$$\text{pour tout voisinage } V \text{ de } 0 \text{ il existe } t > 0 \text{ tel que } A \subset tV := \{tv, v \in V\}.$$

1. Montrer que si d est associée à une norme, on retrouve la définition habituelle d'ensemble borné dans un espace vectoriel normé.
2. Montrer que si $(p_n)_{n\in\mathbb{N}}$ est une famille séparante de semi-normes sur E, et d est la métrique définie par
$$\forall (x, y) \in E^2, \quad d(x, y) := \sum_{n\in\mathbb{N}} \frac{1}{2^n} \min \big\{1, p_n(y - x)\big\},$$
alors A est bornée si et seulement si
$$\forall n \in \mathbb{N}, \ \exists M_n \in \mathbb{R}, \ \ \forall x \in A, \ \ p_n(x) \leq M_n.$$
3. On travaille dans cette question dans $H(U)$ l'espace des fonctions holomorphes sur un ouvert U de $\mathbb{C}$, muni de la métrique définie à l'exercice 4.20.
   (a) Soit A une partie de $H(U)$. Montrer avec le théorème de Montel 4.21 que
   $$A \text{ est compacte} \quad \Longleftrightarrow \quad A \text{ est fermée et bornée.}$$
   (b) En déduire qu'il n'existe pas de norme sur $H(U)$ telle que la convergence pour cette norme soit équivalente à la convergence uniforme sur tout compact de U.

Chapitre IV

Intégration de Lebesgue

Dans ce chapitre, nous étudions la théorie de l'intégration dûe à Lebesgue (1904). En première approche, il semble surtout que cette théorie va permettre d'intégrer plus de fonctions, c'est à dire des fonctions moins régulières que ce qui a été fait pour l'intégrale de Riemann [1]. De telles fonctions peu régulières peuvent apparaître naturellement comme des limites de fonctions plus régulières, par exemple comme somme de séries de Fourier. Précisément, le premier succès de cette théorie sera de fournir des théorèmes de passage à la limite dans les intégrales, qui sont plus naturels et plus performants que ce qu'on a vu au Chapitre I dans le cadre de l'intégrale de Riemann. D'ailleurs, comme on l'avait déjà évoqué, dans le cadre de l'agrégation un réflexe raisonnable est de systématiquement se ramener à ces théorèmes de la théorie de Lebesgue, même quand ceux de la théorie de Riemann pourraient s'appliquer ; on a d'ailleurs de ce fait peu développé ces derniers (voir quand même la proposition 4.11 et l'exercice 5.28 du Chapitre I).

Mais en seconde lecture, on se rend compte que le cadre proposé par Lebesgue a un impact majeur sur deux champs mathématiques qui s'inscrivent tout-à-fait dans le cadre du programme de l'agrégation :

- la théorie de probabilités : on voit une probabilité comme la mesure d'un évènement, ce qui amène à la notion de loi d'une variable aléatoire, et on voit une moyenne comme une intégrale par rapport à une mesure ;
- l'analyse fonctionnelle : en effet, la théorie de Lesbegue fournit des espaces complets pour les normes de type intégrale, ce qui est d'importance capitale. Au niveau de l'agrégation, les principales applications de ces faits se situeront dans l'analyse de Fourier.

Dans la liste des leçons d'oral, la leçon **234** est évidemment étroitement liée à ce chapitre. Les leçons **201-205-208-213-235-239-246-250** vont naturellement amener à des questions liées à l'intégration de Lebesgue. Précisons que pour certaines de ces leçons, il n'y a pas d'obligation de se placer dans le cadre de la théorie de Lebesgue, mais ce serait beaucoup plus naturel et commode de le faire [2]. Les leçons sur les probabilités ont également une grande importance, mais nous n'aborderons ici que brièvement le vocabulaire dans ce cadre ; un chapitre [3] sera entièrement consacré à cela.

Comme pour le chapitre sur les fonctions holomorphes et tous les sujets qui touchent clairement au programme de L3, nous conseillons vivement aux étudiants de ne pas faire une impasse totale sur ce chapitre ; il est indiscutable, à l'écrit et à l'oral, que les étudiants doivent savoir énoncer et utiliser sans faille les théorèmes de convergence de ce chapitre, ainsi que leurs corollaires (comme les théorèmes sur les intégrales à paramètres) pour pouvoir prétendre réussir le concours. Au-delà de ce minimum, disons que chaque étudiant fera en fonction de son temps et de ses moyens ; mais assurément, s'il arrive à passer la couche d'abstraction qui permet de s'autoriser à parler sans crainte de "fonctions

1. Rappelons qu'on a essentiellement 3 espaces naturelles pour considérer l'intégrale de Riemann : les fonctions continues par morceaux, les fontions réglées, et les fonctions Riemann-intégrables

2. Citons par exemple la leçon **246** sur les séries de Fourier ; on peut tout à fait développer et défendre cette leçon dans le cadre de la théorie de Riemann. Cela ne pose pas de réel problème pour définir les coefficients de Fourier et l'étude des séries de Fourier de fonctions explicites. Néanmoins les théorèmes de convergence seront peu naturels et peu performants (les hypothèses seront fortes sur la fonction, et la convergence obtenue assez faible en comparaison), et de ce fait les applications qu'on pourra développer nécessiteront de préciser des hypothèses assez restrictives. Encore une fois, l'étudiant souhaitant défendre cette approche doit surtout en avoir conscience et l'assumer.

3. Fait par M. Berger à la suite du présent chapitre.

mesurables", de "mesure de Lebesgue", de "presque partout", de "fonction L^1", de "convergence L^1", alors son investissement sera très payant, à la fois pour l'écrit et pour l'oral. Afin de motiver l'étudiant à faire cet investissement, précisons que le programme n'a pas vocation à titiller les étudiants sur les subtilités délicates qu'amène l'approche de Lebesgue [4]. Il n'est d'ailleurs pas anodin que le programme officiel précise par deux fois que certaines constructions de ce chapitre seront admises [5]. Comme d'habitude, on abordera via des remarques "étoilées" les questions difficiles soulevées par la théorie, afin que les étudiants puissent en avoir un rapide aperçu ; mais l'étudiant ne doit pas se décourager s'il lui semble ne pas maîtriser tous les contours de la théorie abordée ici, qui sont délicats et dépassent les attentes du concours.

Concluons par des références : j'utilise souvent le livre [GH13] qui me paraît bien adfapté à l'agrégation. J'utilise aussi [Wag12a] qui est plus difficile. Les livres [Rud98, Rev97, BP12] sont des classiques souvent cités. Pour tout ce qui est lié au formalisme des probabilités, j'utilise beaucoup [Ouv07, Ouv09].

1 Notions de théorie de la mesure

Dans ce paragraphe, nous allons aborder les questions naturelles (et étroitement liées) suivantes : quels ensembles pouvons-nous mesurer ? Quelles fonctions allons-nous pouvoir intégrer ?

Peut-être, l'exemple le plus naturel de "mesure" est la notion de longueur pour un sous-ensemble de $\mathbb{R}$. À titre d'exemple, on peut chercher à étendre pour des ensembles plus généraux de $\mathbb{R}$ l'idée que la longueur de l'intervalle $[a, b]$ pour $a \leq b$ est $b - a$. Une grosse difficulté, qui va être une première motivation pour les notions abordées ici, est qu'il n'existe pas de moyen de mesurer tous les sous-ensembles de $\mathbb{R}$ de façon raisonnable, en accord avec cette première intuition de longueur. La seconde motivation sera certainement plus claire dans le cadre de la théorie des probabilités, ou on réutilisera la notion de mesure pour représenter la notion de probabilité d'un évènement, ce qui fournira de nombreux exemples pertinents.

1.1 Préambule : sur la dénombrabilité

Faisons un petit point rapide sur la notion de dénombrabilité : l'objectif n'est pas de rentrer dans la théorie des cardinaux, mais bien d'avoir une bonne intuition des opérations qu'on peut effectuer pour rester dans la classe des ensembles dénombrables qui ont une grande importance dans le formalisme de la théorie de Lebesgue.

Définition 1.1. Un ensemble E est dit dénombrable [6] s'il existe une injection de E dans $\mathbb{N}$.
Il sera dit non dénombrable sinon.

Remarque 1.2. L'intuition qu'on peut avoir d'un ensemble dénombrable, est un ensemble dont on pourra étiqueter les éléments avec des éléments de $\mathbb{N}$. En effet la donnée d'une injection $\varphi : E \to \mathbb{N}$ permet d'associer à chaque élément $x \in E$ une étiquette $\varphi(x) \in \mathbb{N}$, et x sera le seul à porter cette étiquette.

Exemple 1.3. Tous les sous-ensembles de $\mathbb{N}$ sont dénombrables : si $E \subset \mathbb{N}$, il suffit de considérer $n \in E \mapsto n \in \mathbb{N}$ qui est injective.

Un premier exemple d'ensemble qui paraît plus gros que $\mathbb{N}$, mais qui sera en fait dénombrable, est

4. La situation, poussée un peu plus à l'extrême, est comparable à la la théorie des distributions, pour laquelle le programme annonce officiellement qu'il est attendu une capacité de manipulation des objets, mais pas une réelle maîtrise de leurs constructions ou de leur étude abstraite.

5. Le mot "admis" n'apparaissant que trois fois ; deux fois donc pour la théorie de Lebesgue, et une fois dans le chapitre sur les distributions, où on précise que les étudiants peuvent admettre la formule d'intégration par partie sur des ouverts réguliers de $\mathbb{R}^n$ avec $n \geq 2$.

6. Il existe en fait deux terminologies raisonnables : certains diront plutôt "au plus dénombrable" pour la définition qu'on a donné ici. En fait ils préfèrent réserver le mot "dénombrable" pour les ensembles qui sont en bijection avec $\mathbb{N}$. On peut en fait montrer qu'un ensemble "au plus dénombrable" est soit fini, soit en bijection avec $\mathbb{N}$, voir l'exercice 1.7. Dans la terminologie que nous avons choisi, nous dirons "infini-dénombrable" dans ce deuxième cas.

$\mathbb{Z}$: en effet l'application

$$f : \mathbb{Z} \longrightarrow \mathbb{N}$$
$$n \longmapsto \begin{cases} -2n & \text{si} \quad n \leq 0 \\ 2n-1 & \text{si} \quad n > 0 \end{cases}$$

est une bijection (donc une injection).

Remarque 1.4. [*[7]] On peut utiliser la notation $\mathrm{Card}(E) \leq \mathrm{Card}(F)$ pour signifier qu'il existe une injection de E dans F. Il vaut mieux le prendre juste pour une notation, car nous ne définissons pas la notion de cardinal à proprement parler. Ainsi un ensemble dénombrable est tel que $\mathrm{Card}(E) \leq \mathrm{Card}(\mathbb{N})$. On peut aussi lire la notation $\aleph_0 = \mathrm{Card}(\mathbb{N})$.

Exercice 1.5. Soit E et F deux ensembles. Montrer qu'il existe une injection de E dans F si et seulement s'il existe une surjection de F dans E.

Remarque 1.6. L'exercice précédent appliqué aux ensembles dénombrables nous dit qu'un ensemble E sera dénombrable s'il existe une fonction surjective de $\mathbb{N}$ dans E.

[*]Si on veut noter $\mathrm{Card}(F) \geq \mathrm{Card}(E)$ pour signifier le fait qu'il existe une surjection de F dans E, l'exercice précédent affirme que $\mathrm{Card}(E) \leq \mathrm{Card}(F)$ si et seulement si $\mathrm{Card}(F) \geq \mathrm{Card}(E)$.

Exercice 1.7. Soit E un ensemble dénombrable. Montrer que

— soit E est fini,

— soit E est infini et est en bijection avec $\mathbb{N}$.

Exercice 1.8 (Transitivité)**.** Soit E un ensemble, et F un ensemble dénombrable. Montrer que s'il existe une injection de E dans F, alors E est dénombrable.[8]

Exercice 1.9 (**Théorème de Cantor-Bernstein)**.** Le résultat évoqué ici est assez difficile ; on ne l'utilisera en fait pas par la suite, on pourra toujours s'en passer.

On dit que deux ensembles E et F sont équipotent s'il existe une bijection entre eux.

Montrer que E et F sont équipotents si et seulement s'il existe à la fois une injection de E dans F et une injection de F dans E.

Autrement dit, si on veut noter $\mathrm{Card}(E) = \mathrm{Card}(F)$ quand deux ensembles E et F sont en bijection, l'exercice affirme

$$\mathrm{Card}(E) = \mathrm{Card}(F) \quad \Leftrightarrow \quad \mathrm{Card}(E) \leq \mathrm{Card}(F) \text{ et } \mathrm{Card}(E) \geq \mathrm{Card}(F).$$

Exemple 1.10 (Exemple fondamental)**.** Nous pouvons démontrer que l'ensemble $\mathbb{N}^2$ est dénombrable ; c'est un exemple fondamental car on l'utilise souvent pour démontrer que d'autres ensembles sont dénombrables.

On peut donner deux preuves de la dénombrabilité de $\mathbb{N}^2$:

— On peut montrer que l'application[9]

$$f : \quad \mathbb{N}^2 \longrightarrow \mathbb{N}$$
$$(p,q) \longmapsto \frac{(p+q)(p+q+1)}{2} + q = \sum_{k=1}^{p+q} k + q$$

7. Ces notations sont plutôt commodes pour avoir une représentation concise de ce qui est dit, mais je conseille quand même de les éviter en situation de concours (pendant les épreuves), on a vite fait de trouver certaines choses évidentes alors qu'elles ne le sont pas.

8. Plus généralement, s'il existe une injection de E dans F et une injection de F dans G, alors il existe une injection de E dans G, ce qui s'écrit :

$$\mathrm{Card}(E) \leq \mathrm{Card}(F) \text{ et } \mathrm{Card}(F) \leq \mathrm{Card}(G) \quad \Longrightarrow \quad \mathrm{Card}(E) \leq \mathrm{Card}(G).$$

9. Qui se comprend bien sur un dessin !

est une injection[10]. En effet, si $(p,q) \in \mathbb{N}^2$ et $(p',q') \in \mathbb{N}^2$ sont tels que $f(p,q) = f(p',q')$, alors on suppose par l'absurde que $p+q < p'+q'$ et alors

$$\sum_{k=1}^{p+q} k + q = \sum_{k=1}^{p'+q'} k + q' \quad \Rightarrow \quad \sum_{k=p+q+1}^{p'+q'} k = q - q' \quad \Rightarrow \quad p+q+1 \leq q - q'$$

ce qui est une contradiction puisque $q' \leq 0$ et $p+1 \geq 1$. Ainsi $p+q \geq p'+q'$, et en échangeant les rôles on obtient en fait $p+q = p'+q'$, ce qui donne facilement $q = q'$ puis $p = p'$.

— Une autre démonstration astucieuse est de considérer la fonction

$$\begin{aligned} g : \quad \mathbb{N}^2 &\longrightarrow \mathbb{N} \\ (p,q) &\longmapsto 2^p 3^q \end{aligned}$$

dont on voit qu'elle est injective par unicité de la décomposition en nombres premiers.

À partir de cet exemple fondamental, nous allons pouvoir déduire plusieurs exemples intéressants :

Proposition 1.11. *L'ensemble $\mathbb{Q}$ est dénombrable.*

Démonstration. Rappelons que pour tout $x \in \mathbb{Q}^*$, il existe un unique couple $(p,q) \in \mathbb{Z}^* \times \mathbb{N}^*$ tel que $x = p/q$ et $p \wedge q = 1$. Ainsi la fonction qui à 0 associe $(0,1)$ et qui est définie sur $\mathbb{Q}^*$ par

$$\begin{aligned} f : \ \mathbb{Q}^* &\longrightarrow \mathbb{Z} \times \mathbb{N}^* \\ p/q &\longmapsto (p,q) \end{aligned}$$

est une injection de $\mathbb{Q}$ dans $\mathbb{Z} \times \mathbb{N}^*$. Or par l'exemple 1.3, il existe une injection $g : \mathbb{Z} \to \mathbb{N}$, donc l'application qui à (x,y) associe $(g(x),y)$ est une injection de $\mathbb{Z} \times \mathbb{N}^*$ dans $\mathbb{N}^2$. Avec l'exercice 1.8 et l'exemple 1.10, cela permet de conclure. □

Proposition 1.12. *Toute réunion dénombrable d'ensembles dénombrables est dénombrable.*

Démonstration. Soit $E = \bigcup_{n \in \mathbb{N}} E_n$, où pour tout $n \in \mathbb{N}$, E_n est dénombrable. Alors par définition, pour tout $n \in \mathbb{N}$ il existe une injection $\varphi_n : E_n \to \mathbb{N}$. Pour tout $x \in E$ on définit alors

$$N(x) := \min\{n \in \mathbb{N} : x \in E_n\}$$

qui est bien défini car l'ensemble considéré est un ensemble non vide d'entiers naturels, qui admet donc un plus petit élément. Alors la fonction

$$\begin{aligned} \phi : E &\longrightarrow \mathbb{N}^2 \\ x &\longmapsto (N(x), \varphi_{N(x)}(x)) \end{aligned}$$

est une injection : en effet, soit $x,y \in E$ tels que $\phi(x) = \phi(y)$. Alors $N(x) = N(y)$ et en notant n la valeur commune, on obtient $\varphi_n(x) = \varphi_n(y)$, donc $x = y$ par injectivité de φ_n. Avec l'exercice 1.8 et l'exemple 1.10, cela permet de conclure. □

Proposition 1.13. *Tout produit cartésien fini d'ensembles dénombrables est dénombrable.*

Démonstration. Commençons[11] par montrer par récurrence que $\mathbb{N}^n$ est dénombrable pour $n \in \mathbb{N}^*$. Notons $H(n)$: "$\mathbb{N}^n$ est dénombrable" pour $n \in \mathbb{N}^*$.

$H(1)$ est bien sûr vraie.

10. Le lecteur consciencieux pourra montrer que c'est en fait une bijection.

11. Notons qu'on peut donner une autre preuve expéditive, comme on l'a fait dans l'exemple 1.10 : en effet pour $n \in \mathbb{N}^*$ donné, on sait qu'il existe $p_1, p_2, \dots, p_n$ des nombres premiers distincts, et alors

$$\begin{aligned} f : \quad \mathbb{N}^n &\longrightarrow \mathbb{N} \\ (x_1, \dots, x_n) &\longmapsto p_1^{x_1} p_2^{x_2} \dots p_n^{x_n} \end{aligned}$$

est injective.

Soit donc $n \in \mathbb{N}^*$ tel que $H(n)$. Pour voir que $\mathbb{N}^{n+1}$ est dénombrable, on utilise l'hypothèse de récurrence qui donne l'existence de $f : \mathbb{N}^n \to \mathbb{N}$ injective, on considère aussi $g : \mathbb{N}^2 \to \mathbb{N}$ injective donnée par l'exemple 1.10, et on pose

$$\begin{aligned} h : \mathbb{N}^n \times \mathbb{N} &\longrightarrow \mathbb{N} \\ (x, m) &\longmapsto g(f(x), m) \end{aligned}$$

qui est facilement injective et montre $H(n+1)$.

Pour le cas général on se donne $n \in \mathbb{N}^*$ et $(E_i)_{i \in [\![1,n]\!]}$ des ensembles dénombrables, pour lesquels il existe des injections $\varphi_i : E_i \to \mathbb{N}$ pour $i \in [\![1, n]\!]$. Alors la fonction

$$\begin{aligned} \phi : \quad \Pi_{i=1}^n E_i &\longrightarrow \mathbb{N}^n \\ (x_1, \ldots, x_n) &\longmapsto (\varphi_1(x_1), \ldots, \varphi_n(x_n)) \end{aligned}$$

est injective donc avec l'exercice 1.8 et la dénombrabilité de $\mathbb{N}^n$ on conclut que $\Pi_{i=1}^n E_i$ est injective. $\square$

Exercice 1.14.
1. Montrer que l'ensemble des sous-ensembles finis de $\mathbb{N}$ est dénombrable.
2. Montrer que l'ensemble des polynômes à coefficients entiers $\mathbb{Z}[X]$ est dénombrable.
3. Montrer que les ensembles $\mathbb{Q} \cap [0,1]$ et $\mathbb{Q} \cap]0,1[$ sont en bijection [12].

Remarque 1.15. Attention, un produit cartésien infini d'ensembles non vides et non réduits à des singletons est toujours non dénombrable. Un bon exemple pour se convaincre de ce fait est de voir que $\{0,1\}^{\mathbb{N}}$ est non dénombrable, comme on peut le voir à l'exercice suivant.

Exercice 1.16. [*Non dénombrabilité de $\mathbb{R}$]

1. Dans cette question, on montre le résultat classique que E ne peut pas être en bijection avec $\mathcal{P}(E)$ l'ensemble de ses parties.
   (a) Montrer que E s'injecte dans $\mathcal{P}(E)$.
   (b) Montrer que si E est fini, alors $\mathcal{P}(E)$ possède plus d'éléments que E.
   (c) [13]Considérons f une fonction de E dans $\mathcal{P}(E)$. Montrer que
   $$A = \{x \in E, x \notin f(x)\}$$
   ne peut pas avoir d'antécédent par f.
   (d) En déduire qu'il n'existe pas d'injection de $\mathcal{P}(E)$ dans E.
2. On va maintenant montrer la non-dénombrabilité de $\mathbb{R}$.
   (a) Montrer que les ensembles $\{0,1\}^{\mathbb{N}}$ et $\mathcal{P}(\mathbb{N})$ sont équipotents [14].
   (b) Montrer que $\mathbb{R}$ et $[0,1[$ sont équipotents.
   (c) Rappelez la notion de développement décimal d'un nombre, et décrire les réels admettant deux développements décimaux. Faire de même pour le développement binaire, c'est-à-dire en base 2.
   (d) Montrer que les ensembles $\mathbb{R}$, $\{0,1\}^{\mathbb{N}}$ et $\mathcal{P}(\mathbb{N})$ sont équipotents.
   (e) En déduire que l'ensemble des irrationnels est non dénombrable [15].

1.2 Tribus

La notion de tribu est la structure qu'on pourra mettre sur l'ensemble des parties qui pourront être mesurées.

---

12. On pourra donner 2 arguments : un premier utilisant l'exercice 1.7, un second exhibant explicitement une telle bijection)
13. **Cette question est liée au paradoxe de Russell, qui implique qu'il ne peut pas exister d'ensemble contenant tous les ensembles (ici on se place dans l'"axiomatique ZF" (pour Zermelo-Fraenkel) qui est le formalisme habituellement utilisé pour manipuler les ensembles). Du coup on est amené à parler de la *catégorie* des ensembles.
14. C'est-à-dire qu'il existe une bijection entre eux.
15. Ceci prouve que l'ensemble des rationnels est non vide sans en exhiber un explicitement ! On parle de preuve non-constructive d'existence

1.2.1 Définition et exemples

Définition 1.17. Une classe $\mathscr{A}$ de parties d'un ensemble E est appelée tribu[16] (sur E) si

(i) elle contient l'ensemble vide : $\emptyset \in \mathscr{A}$;

(ii) elle est stable par passage au complémentaire : $\forall A \in \mathscr{A},\ A^c \in \mathscr{A}$;

(iii) elle est stable par réunion dénombrable : si $(A_n)_{n\in\mathbb{N}}$ est une famille dénombrable d'éléments de $\mathscr{A}$, alors $\displaystyle\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{A}$.

On dit alors que $(E, \mathscr{A})$ est un *espace mesurable.*

Remarque 1.18. Cette définition a les premières conséquences suivantes :

— $E \in \mathscr{A}$ car $E = \emptyset^c$;

— stabilité par intersection dénombrable car $\displaystyle\bigcap_{n\in\mathbb{N}} A_n = (\bigcup_{n\in\mathbb{N}} A_n^c)^c$;

— stabilité par différence car $A \setminus B = A \cap B^c$;

— stabilité par différence symétrique car $A\Delta B = (A \setminus B) \cup (B \setminus A)$;

Exemple 1.19. Quelques exemples de tribus : on considère E un ensemble quelconque.

— $\{\emptyset, E\}$ est une tribu sur E ;

— $\mathcal{P}(E)$ est une tribu sur E ;

— si $A \subset E$, la plus petite tribu contenant A est $\{\emptyset, A, A^c, E\}$.

Remarque 1.20. [*Comparaison avec la structure topologique] Remarquons qu'on a ici une structure bien différente de ce qu'on a constaté pour la structure de l'ensemble des ouverts d'un espace métrique, qui n'est pas stable par passage au complémentaire, est stable par union quelconque, mais pas par intersection infinie.

Exercice 1.21. On considère E un ensemble.

1. Soit $(A_n)_{n\in\mathbb{N}}$ une partition dénombrable[17] de E. Montrer que
$$\mathscr{A} := \left\{ \bigcup_{i\in I} A_i,\ \ I \text{ sous-ensemble de } \mathbb{N} \right\}$$
est une tribu sur E.

2. On pose
$$\mathscr{A} := \{A \subset E : A \text{ ou } A^c \text{ est dénombrable}\}.$$
Montrer que $\mathscr{A}$ est une tribu.

Définition-Proposition 1.22 (Tribu engendrée)**.** — L'intersection d'une collection non vide quelconque[18] de tribus de parties de E est une tribu.

— Pour toute classe $\mathscr{C}$ de parties de E, l'intersection de toutes les tribus contenant[19] $\mathscr{C}$ est donc[20] une tribu : elle est appelée tribu engendrée[21] par $\mathscr{C}$, et notée $\sigma(\mathscr{C})$:
$$\sigma(\mathscr{C}) := \bigcap_{\mathscr{A} \text{ tribu}, \mathscr{C}\subset\mathscr{A}} \mathscr{A}.$$

16. On rencontre aussi la dénomination σ-algèbre ; le caractère σ fait référence au fait qu'on autorise une stabilité par union dénombrable. Si on autorisait seulement les unions finies, on parlerait d'algèbre sur E.

17. On pourrait penser qu'on a supposé la partition infinie dénombrable car on a indexé par $\mathbb{N}$. Le lecteur peut préférer indexer par I avec I dénombrable, mais en fait nous pouvons voir que notre terminologie n'exclut pas le cas d'une partition finie $(A_n)_{n\in[\![1,]\!]p}$, car on peut compléter cette dernière par des ensembles tous vides $A_n = \emptyset$ pour tout $n > p$.

18. Rappelons qu'on dit quelconque quand on ne fait aucune hypothèse sur le nombre (le cardinal) d'éléments pris en compte ; en particulier ici on ne prétend pas qu'il y a un nombre fini ou dénombrable de tribus qui satisfont aux hypothèses.

19. Au sens de l'inclusion

20. Cette collection est effectivement non vide car un de ses éléments est $\mathscr{P}(E)$.

21. On peut aussi dire la plus petite tribu contenant $\mathscr{C}$.

Remarque 1.23. Faisons quelques commentaires élémentaires à partir de cette définition :

— Si $\mathscr{C}$ est une classe de parties de E et $\mathscr{A}$ est une tribu de parties de E telle que $\mathscr{C} \subset \mathscr{A}$, alors $\mathscr{A}$ est élément de la collection des tribus contenant $\mathscr{C}$, donc contient son intersection $\sigma(\mathscr{C})$, autrement dit $\sigma(\mathscr{C}) \subset \mathscr{A}$.

— Par exemple si $\mathscr{A}$ est une tribu de parties de E, alors $\sigma(\mathscr{A}) = \mathscr{A}$.

— On en déduit aussi que si $\mathscr{C} \subset \mathscr{B}$ alors $\sigma(\mathscr{C}) \subset \sigma(\mathscr{B})$.

Remarque 1.24 (Méthodologie)**.** — Pour montrer une égalité du type $\mathscr{A} = \sigma(\mathscr{C})$, il suffit de montrer que $\mathscr{A} \subset \sigma(\mathscr{C})$, que $\mathscr{C} \subset \mathscr{A}$ et que $\mathscr{A}$ est une tribu.

— Pour montrer que $\sigma(\mathscr{C}_1) = \sigma(\mathscr{C}_2)$, il suffit de montrer que $\mathscr{C}_1 \subset \sigma(\mathscr{C}_2)$ et que $\mathscr{C}_2 \subset \sigma(\mathscr{C}_1)$.

Exercice 1.25 (Tribu trace, tribu image réciproque)**.** 1. Soit $(E, \mathscr{A})$ un ensemble mesurable et X une partie de E. Montrer que la classe

$$\mathscr{C} = \{A \cap X : A \in \mathscr{A}\}$$

est une tribu sur X. On l'appelle tribu trace de $\mathscr{A}$ sur X.

2. Soit $f : E_1 \longrightarrow E_2$.

(a) Si $\mathscr{A}_2$ est une tribu sur E_2, alors

$$f^{-1}(\mathscr{A}_2) := \{f^{-1}(A), A \in \mathscr{A}_2\}$$

est une tribu sur E_1, appelée tribu image réciproque de $\mathscr{A}_2$ par f.

(b) Retrouver le résultat de la question 1 en voyant la tribu trace comme une tribu image réciproque.

(c) Vérifier que si $\mathscr{A}_1$ est une tribu sur E_1, alors $f(\mathscr{A}_1) := \{f(A), A \in \mathscr{A}_1)$ n'est pas une tribu sur E_2 en général.

Lemme 1.26 (Lemme de transport)**.** *Soit $f : E_1 \longrightarrow E_2$ et $\mathscr{C}$ une classe de parties de E_2. Alors*

$$\sigma(f^{-1}(\mathscr{C})) = f^{-1}(\sigma(\mathscr{C})).$$

Démonstration. Montrons d'abord l'inclusion $\sigma(f^{-1}(\mathscr{C})) \subset f^{-1}(\sigma(\mathscr{C}))$. Tout d'abord $\mathscr{C} \subset \sigma(\mathscr{C})$, donc $f^{-1}(\mathscr{C}) \subset f^{-1}(\sigma(\mathscr{C}))$. Mais par l'exercice 1.25, $f^{-1}(\sigma(\mathscr{C}))$ est une tribu, et comme elle contient $f^{-1}(\mathscr{C})$, on obtient $\sigma(f^{-1}(\mathscr{C})) \subset f^{-1}(\sigma(\mathscr{C}))$ (voir la remarque 1.24).

Inversement, posons

$$\mathscr{A} := \sigma(f^{-1}(\mathscr{C})), \qquad \text{et} \qquad \mathscr{B} := \{Y \subset E_2 : f^{-1}(Y) \in \mathscr{A}\}.$$

Vérifions que $\mathscr{B}$ est une tribu car $\mathscr{A}$ l'est :

— $\emptyset \in \mathscr{B}$ car $f^{-1}(\emptyset) = \emptyset \in \mathscr{A}$ car cette dernière est une tribu.

— Si $Y \in \mathscr{B}$, alors $f^{-1}(Y) \in \mathscr{A}$, mais alors $f^{-1}(Y^c) = (f^{-1}(Y))^c \in \mathscr{A}$, donc par stabilité par passage au complémentaire de $\mathscr{A}$, $Y^c \in \mathscr{B}$.

— Si $(Y_n)_{n\in\mathbb{N}}$ est une famille dénombrable d'éléments de $\mathscr{B}$, alors

$$f^{-1}\left(\bigcup_{n\in\mathbb{N}} Y_n\right) = \bigcup_{n\in\mathbb{N}} f^{-1}(Y_n)$$

est bien dans $\mathscr{A}$ par stabilité de $\mathscr{A}$ par union dénombrable.

Alors $\mathscr{C} \subset \mathscr{B}$, et $\mathscr{B}$ est une tribu, donc $\sigma(\mathscr{C}) \subset \mathscr{B}$, puis $f^{-1}(\sigma(\mathscr{C})) \subset f^{-1}(\mathscr{B})$. Mais par définition de $\mathscr{B}$, $f^{-1}(\mathscr{B}) \subset \mathscr{A} = \sigma(f^{-1}(\mathscr{C}))$ donc $f^{-1}(\sigma(\mathscr{C})) \subset \sigma(f^{-1}(\mathscr{C}))$. $\square$

1.2.2 Tribu borélienne

Nous avons évoqué à la remarque 1.20 que l'ensemble des ouverts d'un espace métrique n'est pas une tribu. Cela amène à la construction de ce paragraphe. Pour simplifier, commençons par détailler un premier exemple en construisant une tribu intéressante de sous-ensembles de $\mathbb{R}$:

Définition 1.27. On note $\mathscr{B}(\mathbb{R})$ [22], et on appelle tribu de Borel sur $\mathbb{R}$ la tribu engendrée par les intervalles ouverts de $\mathbb{R}$, c'est-à-dire de la forme $]a,b[$ avec $a \in \mathbb{R} \cup \{-\infty\}$ et $b \in \mathbb{R} \cup \{+\infty\}$.

Proposition 1.28. *Soit S une partie dense de $\mathbb{R}$* [23]. *Alors $\mathscr{B}(\mathbb{R})$ est la tribu engendrée par l'une des familles suivantes :*

$$a)\ \big\{[a,+\infty[, a \in S\big\} \qquad b)\ \big\{]a,+\infty[, a \in S\big\} \qquad c)\ \big\{]-\infty,a[, a \in S\big\} \qquad d)\ \big\{]-\infty,a], a \in S\big\}.$$

Démonstration. On va prouver l'énoncé pour la famille *a)* ; les autres cas sont laissés en exercice. Soit $\mathscr{I}_S$ l'ensemble des intervalles de la forme $[a,+\infty[$ pour $a \in S$. Tout d'abord, $\mathscr{B}(\mathbb{R})$ contient tous les intervalles fermés de $\mathbb{R}$ car est stable par passage au complémentaire ; on a donc l'inclusion $\sigma(\mathscr{I}_S) \subset \mathscr{B}(\mathbb{R})$.

Soit maintenant $a \in \mathbb{R} \cup \{-\infty\}$. Comme S est dense dans $\mathbb{R}$, il existe une suite décroissante $(a_n)_{n\in\mathbb{N}}$ d'éléments de S qui converge vers a et qui sont tels que $a_n > a$ pour tout $n \in \mathbb{N}$. Comme $[a_n,+\infty[\ \in\ \mathscr{I}_S$, on a $[a_n,+\infty[\in \sigma(\mathscr{I}_S)$, donc par stabilité par réunion dénombrable de la tribu $\sigma(\mathscr{I}_S)$,

$$]a,+\infty[= \bigcup_{n\in\mathbb{N}} [a_n,+\infty[\ \in\ \sigma(\mathscr{I}_S).$$

On démontre de la même manière mais avec une suite croissante et avec la stabilité par intersection que $[b,+\infty[\in \sigma(\mathscr{I}_S)$ pour tout $b \in \mathbb{R}$. De plus, pour tous $a \in \mathbb{R} \cup \{-\infty\}$ et $b \in \mathbb{R}$, l'intervalle $]a,b[$ s'écrit $]a,+\infty[\setminus[b,+\infty[\in \sigma(\mathscr{I}_S)$. Par conséquent tous les intervalles ouverts de $\mathbb{R}$ sont dans $\sigma(\mathscr{I}_S)$, et donc $\mathscr{B}(\mathbb{R}) \subset \sigma(\mathscr{I}_S)$, ce qui conclut la preuve. □

Remarque 1.29. On sera parfois amené à travailler dans $\mathbb{R} \cup \{\pm\infty\}$ plutôt que $\mathbb{R}$. La tribu de Borel sur $\mathbb{R} \cup \{\pm\infty\}$ est l'ensemble des parties prenant l'une des formes A, $A \cup \{+\infty\}$, $A \cup \{-\infty\}$ ou $A \cup \{-\infty,+\infty\}$, où $A \in \mathscr{B}(\mathbb{R})$. On laisse le lecteur vérifier qu'il s'agit bien d'une tribu, et aussi qu'il s'agit de la tribu borélienne associée à l'espace métrique $(\mathbb{R} \cup \{\pm\infty\}, d)$ défini à l'exemple 1.6 du Chapitre III.

Remarquons qu'on peut avoir une construction des boréliens dans tout espace métrique :

Définition 1.30. Soit (E,d) un espace métrique. On note $\mathscr{B}(E)$ et on appelle tribu borélienne sur E la tribu engendrée par les ouverts de E. Les éléments de $\mathscr{B}(E)$ sont appelés boréliens de E.

Proposition 1.31. *La tribu $\mathscr{B}(\mathbb{R}^d)$ est la tribu engendrée par la classe des pavés ouverts* [24].

Remarque 1.32. En particulier pour $d = 1$, on constate que les définitions 1.27 et 1.30 sont consistantes, au sens où la tribu engendrée par les intervalles ouverts de $\mathbb{R}$ est bien la même que la tribu engendrée par les ouverts de $\mathbb{R}$.

Démonstration. Notons $\mathscr{A}$ la tribu engendrée par les pavés ouverts de $\mathbb{R}^d$. Comme ce sont des ouverts, on a $\mathscr{A} \subset \mathscr{B}(\mathbb{R}^d)$.

Réciproquement, soit O est un ouvert de $\mathbb{R}^d$. On munit $\mathbb{R}^d$ de la norme $\|\cdot\| := \|\cdot\|_\infty$ car les boules pour cette norme sont des pavés ; bien sûr on peut faire ce choix car le caractère ouvert de O ne dépend pas de la norme du fait du théorème 2.25 du Chapitre III sur l'équivalence des normes dans les espaces vectoriels de dimension finie. Comme O est ouvert, pour tout $x \in O$ il existe $r_x > 0$ tel que $B(x,r_x) = B_{\|\cdot\|_\infty}(x,r_x) \subset O$. Alors on peut écrire

$$O = \bigcup_{x\in O} B(x,r_x).$$

22. En cas de confusion on peut aussi noter $\mathscr{B}or(\mathbb{R})$
23. On peut prendre $S = \mathbb{R}$ qui donne déjà une propriété intéressante, mais on peut aussi considérer $S = \mathbb{Q}$.
24. Où on appelle pavé un produit d'intervalles.

Le problème est que cette union n'est pas dénombrable : nous allons donc utiliser la densité de $\mathbb{Q}$ dans $\mathbb{R}$. Pour $x \in O$ fixé, il existe $r'_x \in \mathbb{Q}$ tel que $\frac{r_x}{4} < r'_x < \frac{3r_x}{4}$. Par densité de $\mathbb{Q}^d$ dans $\mathbb{R}^d$, il existe $a_x \in \mathbb{Q}^2$ tel que $\|a_x - x\| < \frac{r_x}{4}$. Par conséquent, si $y \in B(a_x, r'_x)$, alors l'inégalité triangulaire implique que $\|x - y\| \leq \|x - a_x\| + \|a_x - y\| < \frac{r_x}{4} + r'_x < r_x$, ce qui montre que $B(a_x, r'_x) \subset B(x, r_x) \subset O$. Par ailleurs, comme $\|x - a_x\| < \frac{r_x}{4} < r'_x$, on a également que $x \in B(a_x, r'_x)$. Tout ceci montre que

$$O = \bigcup_{x \in O} B(a_x, r'_x).$$

Comme l'ensemble $\{B(a_x, r'_x), x \in U\}$ est contenu dans l'ensemble $\{B(a, r'), a \in \mathbb{Q}^2,\ r' \in \mathbb{Q}^+\}$ et que ce dernier ensemble est dénombrable, on en déduit que $\{B(a_x, r'_x), x \in O\}$ est également dénombrable [25]. On a donc établi que O peut s'écrire comme une union dénombrable de pavés ouverts, et donc $O \in \mathscr{A}$, ce qui conclut la preuve. □

Exercice 1.33. [*Ouverts de $\mathbb{R}$] Dans la preuve précédente, on a montré que tout ouvert de $\mathbb{R}^d$ est union dénombrable de pavés ouverts, ce qui donne quand $d = 1$ que tout ouvert est union dénombrable d'intervalles ouverts. On peut affiner ce résultat quand $d = 1$: montrer que tout ouvert de $\mathbb{R}$ est union dénombrable d'intervalles ouverts 2 à 2 disjoints.

1.3 Mesure

Maintenant qu'on a pu définir la structure des ensembles qui pourront être mesurés, nous pouvons définir la notion de mesure elle-même.

1.3.1 Définitions et propriétés

Définition 1.34. Soit E un ensemble et $\mathscr{A}$ une tribu sur E. Une mesure [26] sur l'espace mesurable $(E, \mathscr{A})$ est une application $\mu : \mathscr{A} \to [0, +\infty]$ telle que :

(i) $\mu(\emptyset) = 0$;

(ii) μ est σ-additive : pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de $\mathscr{A}$ deux à deux disjoints,

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n).$$

On dit que $(E, \mathscr{A}, \mu)$ est un *espace mesuré*, et pour tout $A \in \mathscr{A}$, on appelle $\mu(A)$ la mesure de A.

Remarque 1.35. Notons que la σ-additivité implique l'additivité finie grâce à (i) : si l'on définit $A_i = \emptyset$ pour tout $i \geq n + 1$, alors

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i \geq 1} \mu(A_i) = \sum_{i=1}^{n} \mu(A_i).$$

Remarque 1.36. [*] On pourrait se demander pourquoi on a souhaité avoir une σ-additivité et non seulement une additivité pour une union finie. En fait on aura fréquemment besoin de la σ-additivité pour pouvoir calculer la mesure de parties compliquées construites comme limites d'ensembles plus simples que l'on sait mesurer. Dans le cadre de la construction de l'intégrale, on aura bien besoin de ce type d'ensemble.

Remarque 1.37. [*] Dans l'égalité $\mu(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$, l'ordre de sommation (membre de droite) n'intervient pas car la série est à termes positifs (voir l'exercice 2.54 du Chapitre I), ce qui est cohérent avec le membre de gauche qui bien sûr ne dépend pas de l'ordre d'énumération des $(A_n)_{n \in \mathbb{N}}$.

Exemple 1.38. Soit E un ensemble. Voici quelques exemples classiques de mesures :

25. Attention, ce n'est pas parce que l'ensemble des indices est non dénombrable que l'ensemble est lui-même non dénombrable. Ce qui se passe c'est que les pavés obtenus sont très souvent les mêmes.

26. Dans ce cours nous ne considérerons que des mesures positives, mais les notions de mesure réelle, complexes ou vectorielles existent

— pour tout $a \in E$, la masse de Dirac au point a est définie pour tout $A \in \mathscr{P}(E)$ par

$$\mu(A) := \begin{cases} 1 & \text{si } a \in A \\ 0 & \text{sinon.} \end{cases}$$

Cette mesure est souvent notée δ_a.

— la *mesure de comptage* sur $\mathscr{P}(E)$:

$$\mu(A) := \begin{cases} \mathrm{Card}(A) & \text{si } A \text{ est fini} \\ +\infty & \text{sinon,} \end{cases}$$

où $\mathrm{Card}(A)$ désigne ici le nombre d'éléments de l'ensemble A.

— soit un espace mesuré $(E, \mathscr{A}, \mu)$ et X une partie de E. Si $X \in \mathscr{A}$, alors on peut définir la *mesure trace* μ_X de μ sur X par $\mu_X(A) := \mu(A \cap X)$ pour tout $A \in \mathscr{A}$.

Proposition 1.39. *Soit $(E, \mathscr{A}, \mu)$ un espace mesuré. Alors pour tout $(A, B) \in \mathscr{A}^2$:*

(i) $\mu(A) = \mu(A \setminus B) + \mu(A \cap B)$,

(ii) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$,

(iii) $\mu(A \cup B) \leq \mu(A) + \mu(B)$ *; on dit que μ est finiment sous-additive,*

(iv) si $A \subset B$, $\mu(A) \leq \mu(B)$; On dit que μ est croissante pour l'inclusion.

Remarque 1.40. Attention, μ est à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$, donc le signe - est à proscrire en général. Par exemple dans (ii), on ne peut pas écrire $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$, qui pourrait être une forme indéterminée, si $\mu(A \cap B) = +\infty$.

Remarque 1.41. La propriété (iii) se généralise bien sûr à un nombre fini de parties par récurrence. Au corollaire 1.43 nous verrons que cela se généralise en fait à une union infinie dénombrable.

Démonstration. (i) $A \setminus B$ et $A \cap B$ sont disjoints et leur réunion est A.

(ii) $A \setminus B$, $A \cap B$ et $B \setminus A$ sont disjoints et leur réunion est $A \cup B$, donc

$$\mu(A \setminus B) + \mu(A \cap B) + \mu(B \setminus A) = \mu(A \cup B),$$

donc en ajoutant $\mu(A \cap B)$ à chaque membre on obtient

$$\mu(A \setminus B) + \mu(A \cap B) + \mu(B \setminus A) + \mu(A \cap B) = \mu(A \cup B) + \mu(A \cap B).$$

Or, dans le premier membre, grâce à (i), la somme des deux premiers termes vaut $\mu(A)$ et la somme des deux derniers termes vaut $\mu(B)$.

(iii) est une conséquence de (ii).

(iv) D'après (ii), si $A \subset B$, alors

$$\mu(B) = \mu(B \setminus A) + \mu(B \cap A) = \mu(B \setminus A) + \mu(A) \geq \mu(A),$$

qui est l'inégalité souhaitée. □

Proposition 1.42. *Une application $\mu : \mathscr{A} \to [0, +\infty]$ est une mesure ssi :*

(i) $\mu(\emptyset) = 0$ *;*

(ii) μ est finiment additive : pour toute famille finie d'éléments $(A_i)_{i \in I}$ deux à deux disjoints de $\mathscr{A}$ avec I est fini, alors $\mu\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mu(A_i)$.

(iii) μ est continue à gauche[27] *: pour toute suite croissante $(A_n)_{n \in \mathbb{N}}$ d'éléments de $\mathscr{A}$,*

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lim_{n \to +\infty} \mu(A_n).$$

27. Il s'agit d'une expression figurée qui signifie 'continue pour les suites croissantes' et est utilisée par analogie avec les fonctions définies sur $\mathbb{R}$ pour lesquels ces deux expressions sont synonymes.

Démonstration. Montrons d'abord le sens $\Rightarrow$ et supposons donc que μ est une mesure. On a déjà vu que (i) et (ii) sont vraies. Montrons la continuité à gauche. Soit $(A_n)_{n\in\mathbb{N}}$ une suite croissante de parties de $\mathscr{A}$: posons $B_0 := A_0$, et $B_n := A_n \setminus A_{n-1}$ pour tout $n \in \mathbb{N}^*$. Alors par une récurrence on montre que pour tout $n \in \mathbb{N}, \bigcup_{k=0}^n B_k = A_n$. De plus les $(B_n)_{n\in\mathbb{N}}$ sont des éléments de $\mathscr{A}$ deux à deux disjoints, donc

$$\mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{n\in\mathbb{N}} \mu(B_n).$$

Mais d'une part, $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A_n$ et d'autre part,

$$\sum_{n\in\mathbb{N}} \mu(B_n) = \lim_{n\to+\infty} \sum_{k=0}^n \mu(B_k) = \lim_{n\to+\infty} \mu\left(\bigcup_{k=0}^n B_k\right) = \lim_{n\to+\infty} \mu(A_n),$$

d'où le résultat.

Montrons maintenant $\Leftarrow$. Soit donc μ vérifiant les trois propriétés de la proposition. Il nous suffit de montrer que μ est bien σ-additive. Soient $(A_n)_{n\in\mathbb{N}}$ des éléments de $\mathscr{A}$ deux à deux disjoints. On pose $B_n := \bigcup_{k=0}^{n} A_k$: $(B_n)_{n\in\mathbb{N}}$ est une suite croissante d'éléments de $\mathscr{A}$ et donc $\mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \lim_{n\to+\infty} \mu(B_n)$. Mais d'une part

$$\mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} \bigcup_{k=0}^n A_k\right) = \mu\left(\bigcup_{n\in\mathbb{N}} A_n\right),$$

et d'autre part, comme μ est finiment additive, $\mu(B_n) = \mu\left(\bigcup_{k=0}^n A_k\right) = \sum_{k=0}^n \mu(A_k)$. Ainsi

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \lim_{n\to+\infty} \mu(B_n) = \lim_{n\to+\infty} \sum_{k=0}^n \mu(A_k) = \sum_{n\in\mathbb{N}} \mu(A_n),$$

ce qui montre la σ-additivité de μ. $\square$

Corollaire 1.43. *Toute mesure μ est* sous σ-additive, *au sens où pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathscr{A}$, on a*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n\in\mathbb{N}} \mu(A_n).$$

Démonstration. On pose $B_n := \bigcup_{k=0}^n A_k$. Par sous-additivité (voir la proposition 1.39 et la remarque 1.41), $\mu(B_n) \leq \sum_{k=0}^n \mu(A_k)$. Mais comme la suite $(B_n)_{n\in\mathbb{N}}$ croît et que son union est égale à $\bigcup_{n\in\mathbb{N}} A_n$, on obtient

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \lim_{n\to+\infty} \mu(B_n) \leq \lim_{n\to+\infty} \sum_{k=0}^n \mu(A_k) = \sum_{n\in\mathbb{N}} \mu(A_n),$$

la deuxième égalité étant due à la continuité à gauche des mesures prouvée à la proposition 1.42. $\square$

Définition 1.44. Une mesure μ sur un espace mesurable $(E, \mathscr{A})$

- est dite finie, ou bornée, si $\mu(E) < \infty$[28]. Le nombre réel $\mu(E)$ est alors appelé masse totale de μ ;
- est appelée (mesure de) probabilité si elle est finie et sa masse totale vaut 1 ;
- est dite σ-finie s'il existe une suite $(E_n)_{n\in\mathbb{N}}$ de parties mesurables de E telles que pour tout $n \in \mathbb{N}$, $\mu(E_n) < \infty$, et $\bigcup_{n\in\mathbb{N}} E_n = E$;

Proposition 1.45 (Continuité à droite, c'est-à-dire pour les suites décroissantes de mesure finie)**.** *Soit $(E, \mathscr{A}, \mu)$ un espace mesuré. Soit $(A_n)_{n\in\mathbb{N}}$ une suite décroissante de $\mathscr{A}$ telle qu'il existe n_0 tel que $\mu(A_{n_0}) < \infty$. Alors*

$$\lim_{n\to+\infty} \mu(A_n) = \mu\left(\bigcap_{n\in\mathbb{N}} A_n\right). \tag{1.1}$$

28. Ceci équivaut à : $\mu(A) < \infty$ pour tout $A \in \mathscr{A}$.

Remarque 1.46. En particulier, si $(A_n)_{n\in\mathbb{N}}$ est une suite décroissante de $\mathscr{A}$ et μ est finie, (1.1) est toujours vraie. Mais pour la mesure de Lebesgue λ, on peut trouver un contre-exemple à la formule (1.1) : on considère $A_n := [n, +\infty[$ pour $n \in \mathbb{N}$. Alors $(A_n)_{n\in\mathbb{N}}$ est une suite décroissante, et pourtant

$$\forall n \in \mathbb{N},\ \ \lambda(A_n) = +\infty, \qquad \text{mais} \qquad \lambda\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \mu(\emptyset) = 0.$$

Démonstration. On pose $B_n := A_{n_0} \setminus A_n$ pour tout $n \geq n_0$. La suite $(B_n)_{n\geq n_0}$ est croissante et d'union égale à $A_{n_0} \setminus \bigcap_{n\geq n_0} A_n$, donc en utilisant la proposition 1.42, on obtient

$$\begin{aligned}
\mu(A_{n_0}) - \mu\left(\bigcap_{n\geq n_0} A_n\right) &= \mu\left(\bigcup_{n\geq n_0} B_n\right) = \lim_{n\to+\infty} \mu(B_n) = \lim_{n\to+\infty} \left[\mu(A_{n_0}) - \mu(A_n)\right] \\
&= \mu(A_{n_0}) - \lim_{n\to+\infty} \mu(A_n),
\end{aligned}$$

ce qui donne bien $\mu\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcap_{n\geq n_0} A_n\right) = \lim_{n\to+\infty} \mu(A_n)$. □

Exercice 1.47. Si $(A_n)_{n\in\mathbb{N}}$ est une suite de parties de E, on définit

$$\limsup_{n\to+\infty} A_n := \lim_{n\to+\infty} \bigcup_{k\geq n} A_k = \bigcap_n \bigcup_{k\geq n} A_k,$$

et

$$\liminf_{n\to+\infty} A_n := \lim_{n\to+\infty} \bigcap_{k\geq n} A_k = \bigcup_n \bigcap_{k\geq n} A_k,$$

1. Montrer que pour tout $x \in E$,

$$\begin{aligned}
x \in \limsup_{n\to+\infty} A_n \quad &\Leftrightarrow \quad \forall n\ \exists k \geq n,\ x \in A_k \quad \Leftrightarrow \quad \{n : x \in A_n\} \text{ est infini.} \\
x \in \liminf_{n\to+\infty} A_n \quad &\Leftrightarrow \quad \exists n\ \forall k \geq n,\ x \in A_k \quad \Leftrightarrow \quad \{n : x \notin A_n\} \text{ est fini.}
\end{aligned}$$

2. Montrer que $\liminf_n A_n \subset \limsup_n A_n$.
3. On suppose que $(A, \mathscr{A}, \mu)$ est un espace mesuré, et pour tout $n \in \mathbb{N}$, $A_n \in \mathscr{A}$.
   (a) Montrer que $\limsup_{n\to+\infty} A_n$ et $\liminf_{n\to+\infty} A_n$ sont des éléments de $\mathscr{A}$.
   (b) Montrer que
   $$\mu\left(\liminf_n A_n\right) \leq \liminf_n \mu(A_n)$$
   (c) Montrer que si μ est finie ou plus généralement s'il existe B de mesure finie et $n_0 \in \mathbb{N}$ tel que $A_n \subset B$ pour tout $n \geq n_0$, alors
   $$\limsup_n \mu(A_n) \leq \mu\left(\limsup_n A_n\right).$$

Proposition 1.48. *Si $(\mu_n)_{n\in\mathbb{N}}$ est une suite croissante de mesures, au sens où pour tout $A \in \mathscr{A}$ et tout $n \in \mathbb{N}$, $\mu_n(A) \leq \mu_{n+1}(A)$, alors l'égalité $\mu(A) := \lim_{n\to+\infty} \mu_n(A) \in [0, +\infty]$ définit une mesure μ sur $\mathscr{A}$.*

Démonstration. Nous allons utiliser la proposition 1.42.

(i) comme $\mu_n(\emptyset) = 0$ pour tout $n \in \mathbb{N}$, $\mu(\emptyset) = \lim_{n\to+\infty} \mu_n(\emptyset) = 0$.

(ii) pour tout ensemble d'indices fini I, pour toutes parties mesurables $(A_i)_{i\in I}$ deux à deux disjointes, l'additivité finie de chaque μ_n s'écrit

$$\mu_n\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mu_n(A_i).$$

L'additivité finie de μ s'obtient en faisant tendre $n \to +\infty$ dans chaque membre (car le membre de droite est une somme finie).

(iii) soit maintenant une suite croissante $(A_k)_{k\in\mathbb{N}}$ d'éléments de la tribu $\mathscr{A}$. La suite doublement indicée $(\mu_n(A_k))_{(n,k)\in\mathbb{N}^2}$ est croissante en k et en n, ce qui garantit que l'on peut intervertir les limites en n et en k [29], d'où :

$$\mu\left(\bigcup_{k\in\mathbb{N}} A_k\right) = \lim_{n\to+\infty} \mu_n\left(\bigcup_{k\in\mathbb{N}} A_k\right) = \lim_{n\to+\infty}\lim_{k\to+\infty} \mu_n(A_k) = \lim_{k\to+\infty}\lim_{n\to+\infty} \mu_n(A_k) = \lim_{k\to+\infty} \mu(A_k),$$

où la deuxième égalité est due à la continuité à gauche de chaque mesure μ_n pour $n\in\mathbb{N}$.

□

Exemple 1.49. En conséquence, toute combinaison linéaire dénombrable, à coefficients positifs, de mesures, est une mesure.

Par exemple, si E est un ensemble, alors pour toute suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de E, pour toute suite $(\alpha_n)_{n\in\mathbb{N}}$ de nombre réels positifs, $\sum\limits_{n\in\mathbb{N}} \alpha_n\delta_{x_n}$ est une mesure sur $(E,\mathcal{P}(E))$.

Exercice 1.50 (Fonction de répartition d'une mesure)**.** Soient μ une mesure finie sur $\mathscr{B}(\mathbb{R})$ et $F : \mathbb{R}\to\mathbb{R}_+$ la fonction définie, pour tout $x\in\mathbb{R}$, par $F(x)=\mu(]-\infty,x])$.

1. Montrer que F est croissante et continue à droite sur $\mathbb{R}$ et calculer ses limites en $\pm\infty$.
2. Pour tout $x\in\mathbb{R}$, montrer que F est continue en x si et seulement si $\mu(\{x\})=0$.
   En déduire que $\{x\in\mathbb{R} : \mu(\{x\})\neq 0\}$ (l'ensemble des atomes de μ) est dénombrable.

1.3.2 Mesure de Lebesgue

La mesure de Lebesgue est une mesure définie sur la tribu de Borel de $\mathbb{R}^d$. Elle donne un sens mathématique à la notion physique de longueur si $d=1$, de surface si $d=2$, de volume si $d=3$.

Théorème 1.51. *Il existe une unique mesure μ définie sur les boréliens de $\mathbb{R}^d$ telle que pour tout pavé ouvert $\prod_{i=1}^d]a_i,b_i[\subset\mathbb{R}^d$ on a*

$$\mu\left(\prod_{i=1}^d]a_i,b_i[\right) = \prod_{i=1}^d (b_i-a_i).$$

Cette mesure est appelée mesure de Lebesgue et est notée λ_d, voire λ s'il n'y a pas d'ambiguïté sur la dimension.

De plus, λ_d est l'unique mesure μ qui est telle que

(i) μ est invariance par translation : pour tout borélien A et tout $x\in\mathbb{R}^d$, $\mu(A+x)=\mu(A)$;

(ii) la mesure du pavé unité est 1 : $\mu\left([0,1]^d\right)=1$.

La démonstration de ce théorème est explicitement exclue du programme. Il existe plusieurs approches, et nous renvoyons le lecteur curieux au paragraphe 6.

Exemple 1.52. — Pour $d=1$ et (a,b) deux réels avec $a\leq b$, on en déduit sans difficulté :

$$\lambda([a,b]) = \lambda([a,b[) = \lambda(]a,b]) = \lambda(]a,b[) = b-a.$$

— En toute dimension $d\geq 1$, la mesure de Lebesgue d'un singleton (qui est bien un borélien) est nul. Par propriété des mesures, on en déduit que tout ensemble dénombrable est de mesure nulle. Ainsi,

$$\lambda_d(\mathbb{Q}^d)=0.$$

29. Car si $(u_{k,n})_{k,n\in\mathbb{N}^2}$ est croissante en k et en n, alors

$$\lim_{n\to+\infty}\lim_{k\to+\infty} u_{k,n} = \sup_{n\in\mathbb{N}}\left\{\sup_{k\in\mathbb{N}} u_{k,n}\right\} = \sup_{k\in\mathbb{N}}\left\{\sup_{n\in\mathbb{N}} u_{k,n}\right\} = \lim_{k\to+\infty}\lim_{n\to+\infty} u_{k,n}.$$

Exercice 1.53 (*Ensemble de Cantor)**.** L'ensemble triadique de Cantor est un sous-ensemble de l'intervalle $[0,1]$. C'est un exemple de partie de $\mathbb{R}$ qui ne contient ni point isolé ni intervalle ouvert, et c'est aussi un exemple d'ensemble de mesure nulle pour la mesure de Lebesgue, mais qui est non-dénombrable [30]. Il est défini comme suit : soit A_0 l'intervalle $[0,1]$, A_1 la réunion de l'intervalle $[0,1/3]$ et de l'intervalle $[2/3,1]$, et plus généralement A_{n+1} la partie de A_n obtenue en divisant chaque composante connexe de A_n en trois sous-intervalles de tailles égales et en lui en ôtant le sous-intervalle central. On pose alors

$$K := \bigcap_{n\in\mathbb{N}} A_n \text{ qui est appelé } \textit{ensemble triadique de Cantor}.$$

1. Montrer que
$$K = \left\{ \sum_{n\geq 1} \frac{x_n}{3^n}, x_n \in \{0,2\} \right\}.$$
2. Montrer que K est compact, d'intérieur vide, équipotent à $\mathbb{R}$, de mesure de Lebesgue nulle.

Exercice 1.54.
1. Montrer que si O est un ouvert de $\mathbb{R}^d$, alors soit O est vide et de mesure nulle, soit O est non vide et de mesure strictement positive.
2. Pour tout $\varepsilon > 0$, montrer qu'il existe O un ouvert dense dans $\mathbb{R}$, de mesure inférieure à ε.
3. [*] Améliorer l'argument pour voir qu'il existe O un ouvert dense dans $\mathbb{R}$ de mesure exactement égale à ε.

Remarque 1.55 (**Ensemble non borélien)**.** Les tribus sont des familles de parties qui sont destinées à être mesurées. Dans le cadre de $\mathbb{R}$, rappelons qu'on souhaite généraliser la notion de longueur, naturellement définie sur les intervalles. Pour pouvoir mesurer des parties suffisamment compliquées comme celles qui ne peuvent être définies que par des passages à la limite (comme l'ensemble triadique de Cantor), nous avons autorisé des opérations relativement générales comme le passage au complémentaire, les réunions et intersections dénombrables. Néanmoins, nous allons voir que nous ne pouvons pas mettre trop d'ensembles dans cette tribu, par exemple on ne pourrait pas choisir la tribu $\mathcal{P}(\mathbb{R})$: en effet, on peut construire des ensembles qui ne peuvent pas admettre de mesure de Lebesgue, et donc qu'on doit exclure de la tribu. L'ensemble construit ici sera en particulier non borélien (car la mesure de Lebesgue sera bien définie sur l'ensemble des boréliens).

On définit la relation d'équivalence $\sim$ sur $\mathbb{R}$:

$$x \sim y \Leftrightarrow x - y \in \mathbb{Q}.$$

En se servant de l'axiome du choix, on peut supposer l'existence d'une partie A de $]0,1[$ qui contient exactement un représentant et un seul de chaque classe d'équivalence de la relation $\sim$. On suppose par l'absurde que A peut admettre une mesure de Lebesgue : soit donc $\lambda(A) \in [0,+\infty]$ la mesure de A. Soit

$$L := \bigcup_{r\in\mathbb{Q}\cap]-1,1[} (r + A),$$

où $r + A = \{r + x, x \in A\}$. Comme A admet une mesure, alors chaque partie $r + A$ en admet une aussi, qui vaut $\lambda(A)$ par invariance par translation de la mesure de Lebesgue. Comme L est réunion dénombrable de parties admettant une mesure, ce doit être également son cas.

Montrons que $]0,1[\subset L$. Pour tout $x \in]0,1[$, désignons par $a = a(x)$ le représentant de sa classe d'équivalence contenu dans A. Alors en particulier, $x - a \in \mathbb{Q}$, et $x - a \in]-1,1[$, donc $r := x - a \in \mathbb{Q}\cap]-1,1[$, et comme $x \in r + A$, $x \in L$. On a aussi $L \subset]-1,2[$, donc on en déduit

$$1 \leq \lambda(L) \leq 3.$$

Montrons que les parties $(r+A)_{r\in\mathbb{Q}}$ sont deux à deux disjointes. Soient $r, s \in \mathbb{Q}$. Si $(r+A)\cap(s+A) \neq \emptyset$, alors il existe $a, b \in A$ tels que $z = r + a = s + b$, donc $b - a = r - s \in \mathbb{Q}$. Par conséquent $a \sim b$, mais comme $a, b \in A$ qui ne contient qu'un représentant de chaque classe d'équivalence, $a = b$, donc $r = s$.

30. La mesure de Lebesgue est introduite au paragraphe 1.3.

Par σ-additivité, nous en déduisons

$$\lambda(L) = \lambda\left(\bigcup_{r\in\mathbb{Q}}(r+A)\right) = \sum_{r\in\mathbb{Q}}\lambda(r+A) = \sum_{r\in\mathbb{Q}}\lambda(A).$$

Cette somme ne peut être qu'infinie (si $\lambda(A)\neq 0$) ou nulle (si $\lambda(A)=0$), ce qui contredit l'inégalité $1\leq\lambda(L)\leq 3$.

1.4 Applications mesurables

1.4.1 Définition

Définition 1.56. Soit $(E_1,\mathscr{A}_1)$ et $(E_2,\mathscr{A}_2)$ deux espaces mesurables. Une fonction $f : E_1 \longrightarrow E_2$ est dite mesurable[31] si $f^{-1}(\mathscr{A}_2)\subset\mathscr{A}_1$, c'est-a-dire si

$$\forall B\in\mathscr{A}_2,\quad f^{-1}(B)\in\mathscr{A}_1. \tag{1.2}$$

Remarque 1.57. On invite le lecteur à comparer avec la définition équivalente de la continuité donnée dans la proposition 1.59 du Chapitre III.

Exercice 1.58. Suivant que E_1 et E_2 sont munis des tribus $\{\emptyset, E_i\}$ ou $\mathcal{P}(E_i)$ pour $i\in\{1,2\}$, décrire les fonctions mesurables dans les quatres situations possibles.

Exemple 1.59. Soit $(E,\mathscr{A})$ un espace mesurable, et $A\subset E$. On peut considérer la fonction indicatrice $\mathbb{1}_A : E \longrightarrow \{0,1\}$ où $\{0,1\}$ est muni de la tribu $\mathcal{P}(\{0,1\})$[32]. On constate sans difficulté que cette fonction est mesurable ssi $A\in\mathscr{A}$.

Remarque 1.60. Dans le cas où E_1 et E_2 sont $\mathbb{R}$ ou $\mathbb{R}\cup\{\pm\infty\}$ (ou plus généralement des espaces métriques), on les munit par défaut des tribus boréliennes. Si $f : (E_1,\mathscr{B}(E_1)) \to (E_2,\mathscr{B}(E_2))$ est mesurable dans ce cas, alors on dira que f est une fonction borélienne.

[*]Précisons dans ce contexte qu'imaginer une fonction de $\mathbb{R}$ dans $\mathbb{R}$ qui n'est pas borélienne est aussi difficile que d'imaginer un ensemble non borélien (voir la remarque 1.55). Intuitivement, il n'est pas trop grave de penser que toute fonction est borélienne, même si cela est faux. Voir aussi la remarque 1.71 à ce sujet.

Remarque 1.61. Soit $f : E_1 \longrightarrow E_2$ et $B\subset E_2$. On utilisera fréquemment la notation $\{f\in B\}$ à la place de $f^{-1}(B)$, ce qui peut se voir comme une écriture condensée de $\{x : f(x)\in B\}$. Par exemple, dans le cas où $E_2=\mathbb{R}$ et $B=[a,+\infty[$, on pourra écrire $f^{-1}(B)$ sous la forme $\{f\geq a\}$.

Cette terminologie est très répandue dans le cadre des probabilités.

1.4.2 Exemples et opérations stables pour la mesurabilité

En premier lieu, on peut voir que si l'espace d'arrivée est engendré par une partie $\mathscr{C}$, alors il suffit de vérifier la propriété (1.2) pour les éléments de cette partie.

Proposition 1.62. *Soit $\mathscr{C}$ une classe de parties de F et $\mathscr{B} := \sigma(\mathscr{C})$. Alors $f : (E,\mathscr{A}) \to (F,\mathscr{B})$ est mesurable ssi $f^{-1}(\mathscr{C})\subset\mathscr{A}$.*

Démonstration. L'application f est mesurable ssi $f^{-1}(\mathscr{B})\subset\mathscr{A}$, mais d'une part $f^{-1}(\mathscr{B}) = f^{-1}(\sigma(\mathscr{C})) = \sigma(f^{-1}(\mathscr{C}))$ par le lemme de transport 1.26, et d'autre part $\sigma(f^{-1}(\mathscr{C}))\subset\mathscr{A}$ ssi $f^{-1}(\mathscr{C})\subset\mathscr{A}$. □

Exemple 1.63. L'exemple suivant peut s'avérer bien utile : soit S une partie dense de $\mathbb{R}$[33]. Alors d'après les propositions 1.28 et 1.62, la fonction $f : (E,\mathscr{A}) \to (\mathbb{R},\mathscr{B}(\mathbb{R}))$ est mesurable ssi $\{f\geq a\}\in\mathscr{A}$ pour tout $a\in S$. On peut remplacer $\{f\geq a\}$ par $\{f>a\}$, $\{f\leq a\}$ ou $\{f<a\}$.

31. Il est sous-entendu que c'est par rapport aux deux tribus $\mathscr{A}_1$ et $\mathscr{A}_2$, comme on sous-entend la continuité d'une fonction par rapport aux métriques qui munissent les espaces de départ et d'arrivée. Pour lever l'ambiguïté, on écrira parfois $f : (E_1,\mathscr{A}_1) \to (E_2,\mathscr{A}_2)$ mesurable.

32. Qui se trouve être aussi la tribu trace de $\mathbb{R}$ sur $\{0,1\}$.

33. Le cas $S=\mathbb{R}$ est déjà intéressant.

Exercice 1.64. Montrer que si $f:\mathbb{R}\to\mathbb{R}$ est monotone, alors f borélienne.

Établissons le lien entre fonction continue et borélienne :

Proposition 1.65. *Soit $(E_1,\mathscr{A}_1)$ un espace mesuré, et (E_2,d_2) un espace métrique, que l'on munit de la tribu des boréliens. Alors $f:(E_1,\mathscr{A}_1)\to(E_2,\mathscr{B}(E_2))$ est mesurable si et seulement si pour tout ouvert O de E_2, $f^{-1}(O)\in\mathscr{A}_1$.*

Démonstration. C'est une conséquence de la définition des boréliens et de la proposition 1.62. □

Corollaire 1.66. *Soit (E_1,d_1) et (E_2,d_2) deux espaces métriques. Si $f:E_1\to E_2$ est continue, alors f est borélienne.*

Démonstration. Cela découle de la proposition 1.65 et le fait que si O est un ouvert de E_2, alors par continuité de $f:E_1\to E_2$, $f^{-1}(O)$ est un ouvert de E_1, donc un élément de $\mathscr{B}(E_1)$. □

Comme pour les fonctions continues, on peut voir que de nombreuses opérations conservent la mesurabilité :

Proposition 1.67.
1. *Soient $f_1:(E_1,\mathscr{A}_1)\to(E_2,\mathscr{A}_2)$ et $f_2:(E_2,\mathscr{A}_2)\to(E_3,\mathscr{A}_3)$. Si f_1 et f_2 sont mesurables, alors $f_2\circ f_1:(E_1,\mathscr{A}_1)\to(E_3,\mathscr{A}_3)$ est aussi mesurable.*
2. *Soit $f=(f_1,f_2):(E,\mathscr{A})\to(\mathbb{R}^2,\mathscr{B}(\mathbb{R}^2))$. Alors f est mesurable si et seulement si f_1 et f_2 sont mesurables.*
3. *Soient $f:(E,\mathscr{A})\to(\mathbb{R},\mathscr{B}(\mathbb{R}))$ et $g:(E,\mathscr{A})\to(\mathbb{R},\mathscr{B}(\mathbb{R}))$ deux fonctions mesurables, et $\alpha\in\mathbb{R}$. Alors $\alpha f+g$ et $f.g$ sont mesurables.*

Remarque 1.68. Le second énoncé peut être remplacé par le fait qu'une fonction $f:(E,\mathscr{A})\to(\mathbb{C},\mathscr{B}(\mathbb{C}))$ est mesurable si et seulement si $\mathrm{Re}(f)$ et $\mathrm{Im}(f)$ sont mesurables.
Dans le troisème énoncé on peut remplacer $\mathbb{R}$ par $\mathbb{C}$.
L'ensemble des applications mesurables de $(E,\mathscr{A})$ dans $(\mathbb{K},\mathscr{B}(\mathbb{K}))$ (avec $\mathbb{K}=\mathbb{R}$ ou $\mathbb{C}$) muni des opérations $+,\cdot,\times$ est une algèbre sur $\mathbb{K}$.

Démonstration.
1. Pour tout $A_3\in\mathscr{A}_3$, on vérifie que $(f_2\circ f_1)^{-1}(A_3)=f_1^{-1}(f_2^{-1}(A_3))$. Or $f_2^{-1}(A_3)\in\mathscr{A}_2$ car f_2 est mesurable, puis $f_1^{-1}(f_2^{-1}(A_3))\in\mathscr{A}_1$ car f_1 est mesurable.
2. Supposons f mesurable. Comme les applications $\pi_i:(x_1,x_2)\in\mathbb{R}^2\mapsto x_i\in\mathbb{R}$ pour $i\in\{1,2\}$ sont continues, on en déduit avec le premier point que $f_i=\pi_i\circ f$ sont mesurables pour $i\in\{1,2\}$.
   Réciproquement, supposons que f_1 et f_2 sont mesurables. On va utiliser la proposition 1.31 : soit $R=]a,b[\times]c,d[$ un rectangle ouvert de $\mathbb{R}^2$. On a alors
   $$f^{-1}(R)=f_1^{-1}(]a,b[)\cap f_2^{-1}(]c,d[)$$
   donc par mesurabilité de f_1 et f_2, on en déduit que $f^{-1}(R)\in\mathscr{A}$. Mais la proposition 1.31 montre que l'ensemble des rectangles ouverts engendre la tribu $\mathscr{B}(\mathbb{R}^2)$, donc avec la proposition 1.62 on peut conclure à la mesurabilité de f.
3. Les applications $S_\alpha:(x,y)\in\mathbb{R}^2\mapsto\alpha x+y\in\mathbb{R}$ et $P:(x,y)\in\mathbb{R}^2\mapsto x.y$ sont continues donc boréliennes, par la proposition 1.65. En notant $\Phi:x\in E\mapsto(f(x),g(x))$ qui est mesurable par le point précédent, on a donc $\alpha f+g=S_\alpha\circ\Phi$ et $f.g=P\circ\Phi$ qui sont mesurables par composition.

□

Maintenant, on va voir que la notion de mesurabilité est beaucoup plus stable par passage à la limite que la notion de continuité : en effet, on a vu (théorème 4.10 du Chapitre I) que l'on préservait la continuité si on avait la convergence uniforme, mais que la convergence simple ne suffisait pas en général. Ici la mesurabilité sera conservée par la convergence simple :

Proposition 1.69. *Soit $(E,\mathscr{A})$ un espace mesuré, et $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions mesurables de $(E,\mathscr{A})$ dans $(\mathbb{R}\cup\{\pm\infty\},\mathscr{B}(\mathbb{R}\cup\{\pm\infty\}))$. Alors*

(i) $\sup\limits_{n\in\mathbb{N}} f_n$ *et* $\inf\limits_{n\in\mathbb{N}} f_n$ *sont mesurables ;*

(ii) $\limsup\limits_{n\to+\infty} f_n$ *et* $\liminf\limits_{n\to+\infty} f_n$ *sont mesurables ;*

(iii) si (f_n) *converge simplement vers une fonction* f [34] *(définie à valeurs dans* $\mathbb{R}\cup\{\pm\infty\}$*), alors* f *est mesurable.*

Démonstration. (i) Pour tout $a\in\mathbb{R}$, on a

$$\left\{\sup_{n\in\mathbb{N}} f_n \le a\right\} = \bigcap_{n\in\mathbb{N}}\{f_n\le a\}\in\mathscr{A}, \quad \text{et} \quad \left\{\inf_{n\in\mathbb{N}} f_n \ge a\right\} = \bigcap_{n\in\mathbb{N}}\{f_n\ge a\}\in\mathscr{A}.$$

(ii) D'après (i), pour tout $n\in\mathbb{N}$, la fonction $\sup_{k\ge n} f_k$ est mesurable, donc la fonction $\limsup\limits_{n\to+\infty} f_n = \inf\limits_{n\in\mathbb{N}}\sup\limits_{k\ge n} f_k$ est mesurable. De même pour $\liminf_{n\to+\infty} f_n$.
(iii) Si $(f_n)_{n\in\mathbb{N}}$ converge simplement vers f, alors $f=\limsup_{n\to+\infty} f_n$, qui est mesurable d'après (ii). □

Exercice 1.70. 1. Montrer que les fonctions continues par morceaux de $[a,b]$ dans $\mathbb{R}$ sont boréliennes. [35]

2. [*] Montrer qu'une fonction réglée de $[a,b]$ dans $\mathbb{R}$ est borélienne.
3. [**] En admettant qu'il existe un ensemble non-borélien de mesure nulle [36], montrer qu'il existe une fonction de $[0,1]$ dans $\mathbb{R}$ qui est Riemann-intégrable mais non borélienne [37].
4. Soit I un intervalle de $\mathbb{R}$. Montrer qu'une fonction continues par morceaux de I dans $\mathbb{R}$ (voir la définition donnée à la note 49 du Chapitre I) est borélienne.

Remarque 1.71. On a dit à la remarque 1.60 qu'il était difficile d'imaginer une fonction de $\mathbb{R}$ dans $\mathbb{R}$ qui ne soit pas borélienne. Il n'empêche qu'il faut savoir rapidement justifier qu'une fonction est borélienne pour pouvoir justifier le cadre d'application de la théorie de Lebesgue. Dans le cadre des épreuves de l'agrégation, on peut résumer les choses ainsi : dans plus de 90% des cas, les fonctions seront explicites et facilement continues par morceaux, donc boréliennes. Dans la plupart des cas restants, les fonctions seront limites simples de fonctions boréliennes, et donc seront également boréliennes.

Exercice 1.72. Soit $(E,\mathscr{A})$ un espace mesurable et $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions mesurables de $(E,\mathscr{A})$ dans $\mathbb{R}$. Montrer que l'ensemble

$$A=\{x\in E : \text{la suite } (f_n(x))_{n\in\mathbb{N}} \text{ est convergente}\}$$

est un élément de $\mathscr{A}$.

Exercice 1.73 (Mesurabilité par rapport à une tribu image réciproque)**.** Soit $f:E\to\mathbb{R}$ et $\mathscr{A}_f=f^{-1}(\mathscr{B}(\mathbb{R}))$ la tribu image réciproque de $\mathscr{B}(\mathbb{R})$ par f.

1. Soit $h:\mathbb{R}\to\mathbb{R}$ une fonction borélienne. Montrer que $g=h\circ f$ est une fonction mesurable de $(E,\mathscr{A}_f)$ dans $\mathbb{R}$.
2. Soit $s:(E,\mathscr{A}_f)\to\mathbb{R}$ une fonction étagée mesurable. Montrer qu'il existe une fonction borélienne t telle que $s=t\circ f$.

---

34. Autrement dit : $\forall x\in E, f_n(x)\xrightarrow[n\to+\infty]{} f(x)$.

35. On peut proposer deux démonstrations : l'une qui revient à la définition, l'autre qui utilise le lemme 5.4 du Chapitre I.

36. **On peut montrer l'existence d'un tel ensemble par deux arguments. Un argument de cardinal, car si on note C l'ensemble de Cantor, on peut justifier les inégalités :

$$\mathrm{Card}(\mathscr{B}(C))=\mathrm{Card}(\mathscr{B}(\mathbb{R}))<\mathrm{Card}(\mathcal{P}(\mathbb{R}))=\mathrm{Card}(\mathcal{P}(C))$$

les deux égalités découlant du fait que C est équipotent à $\mathbb{R}$, et l'inégalité stricte découlant de l'égalité $\mathrm{Card}(\mathscr{B}(\mathbb{R}))=\mathrm{Card}(\mathbb{R})$ qui se montre par récurrence transfinie. L'autre argument consiste à utiliser la fonction de Cantor et l'ensemble de la remarque 1.55, voir [BP12, page 291].

37. **Attention, cet exemple pourrait suggérer que la théorie de Lebesgue n'est pas “meilleure” que celle de Riemann en ce qui concerne l'étendue des fonctions qu'on peut intégrer. Il n'en est rien, car pour pallier à ce problème, on peut compléter la tribu de l'espace de départ. On dira alors que les fonctions sont Lebesgue-mesurables. Et on peut montrer que toute fonction Riemann-intégrable est Lebesgue-mesurable.

3. Montrer que si $g : (E, \mathscr{A}_f) \to \mathbb{R}$ est mesurable, alors il existe h borélienne telle que $g = h \circ f$.

Exercice 1.74 (Théorème d'Egoroff)**.** Soit $(E, \mathscr{A}, \mu)$ un espace mesuré tel que $\mu(E) < +\infty$ et soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions mesurables de $(E, \mathscr{A})$ dans $\mathbb{R}$.

1. On suppose que la suite $(f_n)_{n\in\mathbb{N}}$ converge μ-p.p. vers une fonction mesurable f, au sens où $\forall x \in C, f_n(x) \xrightarrow[n\to+\infty]{} f(x)$ et $\mu(E \setminus C) = 0$. Pour tout $k \in \mathbb{N}^*$ et tout $n \in \mathbb{N}$, on pose
$$E_n^k = \bigcap_{i\geq n} \left\{|f_i - f| \leq \frac{1}{k}\right\}.$$
Montrer que $C \subset \bigcup_{n\geq 1} E_n^k$. En déduire que, pour tout réel $\varepsilon > 0$, pour tout $k \in \mathbb{N}^*$, il existe $n_{k,\varepsilon} \in \mathbb{N}^*$ tel que $\mu\left(E \setminus E^k_{n_{k,\varepsilon}}\right) < \frac{\varepsilon}{2^k}$.
2. (Théorème d'Egoroff) En déduire que, pour tout $\varepsilon > 0$, il existe $E_\varepsilon \in \mathscr{A}$ tel que $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers f sur E_ε et tel que $\mu(E \setminus E_\varepsilon) < \varepsilon$.
3. Donner un contre-exemple lorsque $\mu(E) = +\infty$.

1.4.3 Fonctions étagées et approximation

Définition 1.75. Une fonction $f : (A, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ est dite étagée si elle est mesurable et ne prend qu'un nombre fini de valeurs. Dans ce cas il existe une partition finie $\{A_i\}_{i\in I}$ de E telle que $\forall i \in I, A_i \in \mathscr{A}$ et des nombres réels $(\alpha_i)_{i\in I}$ tels que $f = \sum_{i\in I} \alpha_i \mathbb{1}_{A_i}$.

Exemple 1.76. Une fonction indicatrice d'un ensemble mesurable est étagée, et on a $\mathbb{1}_A = 1 \cdot \mathbb{1}_A + 0 \cdot \mathbb{1}_{A^c}$.

Remarque 1.77. La partition qui apparaît dans la définition 1.75 n'est pas unique, mais si on écrit f sous la forme $\sum_{i\in I} \alpha_i \mathbb{1}_{A_i}$ où on demande en plus que les α_i sont deux à deux distincts, alors on récupère l'unicité, et de plus $A_i = \{f = \alpha_i\}$ pour tout $i \in I$.

Proposition 1.78. *Soit f, g deux fonctions étagée et $\alpha \in \mathbb{R}$. Les fonctions $\lambda f + g$, fg, $f \wedge g$ et $f \vee g$*[38] *sont étagées.*

Démonstration. On écrit f et g sous la forme $f = \sum_{i\in I} \alpha_i \mathbb{1}_{A_i}$ et $g = \sum_{j\in J} \beta_j \mathbb{1}_{B_j}$. Alors $(A_i \cap B_j; (i,j) \in I \times J)$ est une partition finie de E et on peut écrire
$$\lambda f + g = \sum_{(i,j)\in I\times J} (\lambda\alpha_i + \beta_j)\mathbb{1}_{A_i\cap B_j}\ , \qquad fg = \sum_{(i,j)\in I\times J} \alpha_i\beta_j \mathbb{1}_{A_i\cap B_j},$$
$$f \wedge g = \sum_{(i,j)\in I\times J} (\alpha_i \wedge \beta_j)\mathbb{1}_{A_i\cap B_j}\ , \qquad f \vee g = \sum_{(i,j)\in I\times J} (\alpha_i \vee \beta_j)\mathbb{1}_{A_i\cap B_j}.$$
□

Exercice 1.79.
1. Montrer qu'une fonction de $[a, b]$ dans $\mathbb{R}$ qui est en escalier est étagée[39].
2. Montrer que la fonction $\mathbb{1}_{\mathbb{Q}}$ est étagée, mais n'est en escalier sur aucun segment $[a, b] \subset \mathbb{R}$ avec $a < b$.

Lemme 1.80 (lemme fondamental d'approximation)**.** *Soit $f : (E, \mathscr{A}) \to (\mathbb{R} \cup \{\pm\infty\}, \mathscr{B}(\mathbb{R} \cup \{\pm\infty\}))$ mesurable. Alors il existe une suite $(f_n)_{n\in\mathbb{N}}$ de fonctions étagées convergeant simplement vers f. De plus,*

a) si f est positive, on peut choisir la suite $(f_n)_{n\in\mathbb{N}}$ positive et croissante[40]*,*

b) si f est bornée, on peut choisir $(f_n)_{n\in\mathbb{N}}$ de sorte que la convergence soit uniforme.

38. $a \wedge b$ est une notation alternative pour $\min(a, b)$, et $a \vee b$ pour $\max(a, b)$.
39. Comme on l'a dit à la remarque 1.60, les tribus choisies par défaut sont les tribus boréliennes.
40. Autrement dit : $\forall x \in E, \forall n \in \mathbb{N}, 0 \leq f_n(x) \leq f_{n+1}(x)$. Attention, cela n'a rien à voir avec le fait de demander que les fonctions $(f_n)_{n\in\mathbb{N}}$ soient croissantes, ce qui n'aurait d'ailleurs pas de sens ici.

Démonstration. Commençons par le cas où f est positive. On définit alors

$$f_n := \sum_{k=0}^{n2^n-1} \frac{k}{2^n}\mathbb{1}_{\left\{\frac{k}{2^n}\leq f<\frac{k+1}{2^n}\right\}} + n\mathbb{1}_{\{f\geq n\}}.$$

Alors pour tout $x \in E$, la suite $(f_n(x))_{n\in\mathbb{N}}$ est bien positive et croissante et converge vers $f(x)$; en effet si $f(x) = +\infty$, alors $f_n(x) = n \xrightarrow[n\to+\infty]{} +\infty$ et sinon il existe $n_0 \in \mathbb{N}$ tel que $f(x) < n_0$, ce qui implique que pour tout $n \geq n_0$, il existe un unique $k \in [\![0, n2^n - 1]\!]$ tel que $x \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[$, et alors $|f_n(x) - f(x)| \leq \dfrac{1}{2^n} \xrightarrow[n\to+\infty]{} 0$. De plus, par mesurabilité de f, les fonctions $(f_n)_{n\in\mathbb{N}}$ sont mesurables, et comme elles prennent un nombre fini de valeurs, elles sont bien étagées.

Si f est bornée et positive, alors il existe $n_0 \in \mathbb{N}$ tel que pour tout $x \in E$, $f(x) < n_0$. Comme précédemment, on a alors pour tout $x \in E$ et pour tout $n \geq n_0$, $|f_n(x) - f(x)| \leq 2^{-n}$. Ainsi $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers f.

Si f est de signe quelconque, on écrit f sous la forme $f = f^+ - f^-$, où

$$f^+ := f\mathbb{1}_{\{f>0\}} \qquad \text{et} \qquad f^- := -f\mathbb{1}_{\{f<0\}}.$$

La somme $f^+ - f^-$ n'est jamais indéterminée, car pour tout $x \in E$, au moins un des deux termes $f^+(x)$ ou $f^-(x)$ est nul. On notera également que f^+ (et f^-, par un même raisonnement) est mesurable car pour tout $a \geq 0$, $\{f^+ \geq a\} = \{f \geq a\}$ et pour tout $a < 0$, $\{f^+ \geq a\} = E$. À présent, comme f^+ et f^- sont positives, il existe deux suites croissantes $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ de fonctions étagées positives convergeant resp. vers f^+ et f^-. De plus, si l'on utilise la construction de ces suites proposée plus haut, on a $u_n v_n = 0$, de sorte que l'on peut toujours définir $f_n := u_n - v_n$, qui définit une suite de fonctions étagées convergeant vers $f^+ - f^- = f$.

Si f est de signe quelconque mais bornée, alors f^+ et f^- sont bornées, donc on peut choisir les suites $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ pour que les convergences vers f^+ et f^- soient toutes deux uniformes. Alors la suite $(u_n - v_n)_{n\in\mathbb{N}}$ converge uniformément vers f. □

Remarque 1.81. Le lemme précédent est à mettre en comparaison du lemme 5.4 du Chapitre I. Pour la construction de l'intégrale de Riemann, on a utilisé l'approximation des fonctions continues par morceaux par des fonctions en escaliers, et l'intégrale est vue comme limite des intégrales de fonctions en escalier ; pour cela on a découpé l'espace de départ (un segment) en petits morceaux (les subdivisions). Pour l'intégrale de Lebesgue, c'est l'espace d'arrivée ($\mathbb{R}$ ou $\mathbb{R}\cup\{\pm\infty\}$) qui sera découpé, et on verra l'intégrale de la fonction comme limite des intégrales de fonctions étagées. Cette différence est fondamentale, et nous oblige notamment à mesurer la "taille" de l'ensemble $\{a \leq f < b\} = f^{-1}([a, b[)$; ceci explique la nécessité de "mesurer" des ensembles possiblement compliqués.

2 Construction de l'intégrale de Lebesgue

2.1 Intégrale des fonctions étagées positives

Définition 2.1. Soit $(E, \mathscr{A}, \mu)$ un espace mesuré, et $f : E \to \mathbb{R}_+$ une fonction étagée[41] positive. On appelle intégrale (sur E) de f par rapport à la mesure μ, et l'on note $\displaystyle\int_E f d\mu$ l'élément de $[0, +\infty]$

$$\int_E f\, d\mu := \sum_{\alpha\in f(E)} \alpha\mu(\{f = \alpha\})\,^{42}, \tag{2.1}$$

41. Rappelons qu'on a défini une fonction étagée comme étant mesurable, donc ici il est sous-entendu que E est muni de la tribu $\mathscr{A}$ et $\mathbb{R}_+$ de la tribu $\mathscr{B}(\mathbb{R}_+)$.
42. C'est une somme finie !

où l'on utilise la convention $0 \times \infty = 0$. On notera indifféremment l'intégrale de f par rapport à μ sous une des formes suivantes

$$\int_E f\,d\mu, \qquad \int_E f(x)\,d\mu(x), \qquad \int_E f(x)\,\mu(dx), \qquad \int f d\mu$$ [43].

Remarque 2.2. La définition précédente ne dépend pas de la représentation de f sous la forme $f = \sum_{i\in I} \alpha_i \mathbb{1}_{A_i}$: on peut vérifier qu'on a toujours l'égalité

$$\int_E f\,d\mu = \sum_{i\in I} \alpha_i \mu(A_i).$$

Exemple 2.3.

1. Si $\mu = \delta_a$ et $f : E \to \mathbb{R}_+$ est étagée, alors on trouve
$$\int_E f\,d\mu = \sum_{\alpha\in f(E)} \alpha\mu(\{f = \alpha\}) = f(a).$$

2. Notons m la mesure de comptage sur $(E, \mathcal{P}(E))$. Si $f : E \to \mathbb{R}_+$ est étagée [44], alors
$$\int_E f dm = \sum_{\alpha\in f(E)} \alpha \mathrm{Card}(\{f = \alpha\}) = \sum_{x\in E} f(x)$$
où la somme possède seulement un nombre fini de termes non nuls. Par exemple, si $E = \mathbb{N}$, on aura
$$\int_E f dm = \sum_{n\in\mathbb{N}} f(n).$$

3. Si μ est la mesure de Lebesgue sur $\mathbb{R}$, alors $\mathbb{1}_{\mathbb{Q}}$ est étagée (car $\mathbb{Q}$ est borélien) et
$$\int_{\mathbb{R}} \mathbb{1}_{\mathbb{Q}}\,d\lambda = \lambda(\mathbb{Q}) = 0.$$

Exercice 2.4. Soit $f : E \to \mathbb{R}_+$ étagée. Montrer

$$\int_E f\,d\mu < \infty \Leftrightarrow \mu(\{f \neq 0\}) < \infty.$$

Proposition 2.5. *L'application* $f \mapsto \int_E f\,d\mu$ *du cône des fonctions étagées positives et à valeurs dans* $\mathbb{R}_+ \cup \{+\infty\}$ *satisfait :*

(i) additivité : $\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$ *;*

(ii) positive homogénéité : pour tout $\alpha \in \mathbb{R}_+$, $\int (\alpha f)\,d\mu = \alpha \int f\,d\mu$ *;*

(iii) croissance : si $f \leq g$ *alors* $\int f\,d\mu \leq \int g\,d\mu$.

Démonstration. Soient $f = \sum_{i\in I} \alpha_i \mathbb{1}_{A_i}$ et $g = \sum_{j\in J} \beta_j \mathbb{1}_{B_j}$, où les α_i, β_j sont des réels positifs ou nuls, et $(A_i)_{i\in I}$, $(B_j)_{j\in J}$ sont des partitions finies de E.

(i) On a vu dans la preuve de la proposition 1.78 que $(A_i \cap B_j)_{(i,j)\in I\times J}$ est une partition finie de E et que

$$f + g = \sum_{(i,j)\in I\times J} (\alpha_i + \beta_j) \mathbb{1}_{A_i\cap B_j}.$$

43. S'il n'y a pas ambiguïté sur l'espace de départ.
44. Au vu de la tribu mise sur E, toutes les fonctions sont mesurables, donc pour que f soit étagée il faut et il suffit que f ait un nombre fini de valeurs.

Par conséquent,

$$\begin{aligned}\int_E (f+g)\,d\mu &= \sum_{(i,j)\in I\times J} (\alpha_i+\beta_j)\mu(A_i\cap B_j)\\ &= \sum_{i\in I}\alpha_i\sum_{j\in J}\mu(A_i\cap B_j)+\sum_{j\in J}\beta_j\sum_{i\in I}\mu(A_i\cap B_j)\\ &= \sum_{i\in I}\alpha_i\mu(A_i)+\sum_{j\in J}\beta_j\mu(B_j)=\int_E f\,d\mu+\int_E g\,d\mu.\end{aligned}$$

(ii) Pour tout $\alpha\geq 0$, $\alpha f=\sum_{i\in I}\alpha\alpha_i\mathbb{1}_{A_i}$ est étagée, d'où

$$\int_E (\alpha f)\,d\mu=\sum_{i\in I}\alpha\alpha_i\mu(A_i)=\alpha\sum_{i\in I}\alpha_i\mu(A_i)=\alpha\int_E f\,d\mu.$$

(iii) En écrivant $g=f+(g-f)$, où $g-f$ est étagée positive, d'après (i), $\displaystyle\int g\,d\mu=\int f\,d\mu+\int(g-f)\,d\mu$ avec $\displaystyle\int(g-f)\,d\mu\in[0,+\infty]$, donc $\displaystyle\int g\,d\mu\geq\int f\,d\mu$. □

2.2 Intégrale des fonctions mesurables positives

Définition 2.6. Soit $f:(E,\mathscr{A})\to(\mathbb{R}\cup\{\pm\infty\},\mathscr{B}(\mathbb{R}\cup\{\pm\infty\}))$ mesurable et positive. On appelle intégrale de f par rapport à μ, et l'on note[45] $\displaystyle\int_E f\,d\mu$ l'élément de $[0,+\infty]$ défini par :

$$\int_E f\,d\mu:=\sup\left\{\int_E g\,d\mu : g:E\to\mathbb{R}\text{ étagée positive et telle que } g\leq f\right\}. \tag{2.2}$$

Si $\displaystyle\int_E f\,d\mu<\infty$, on dit que f est intégrable, ou μ-intégrable.

Remarque 2.7. On laisse le lecteur prouver rigoureusement qu'il s'agit bien d'un prolongement, c'est-à-dire que si f est étagée, les formules (2.1) et (2.2) coïncident.

On va voir que les propriétés de l'intégrale des fonctions étagées se conservent dans ce cadre.

Proposition 2.8 (croissance de l'intégrale)**.** *Soient f,g de $(E,\mathscr{A})$ dans $(\overline{\mathbb{R}},\mathscr{B}(\overline{\mathbb{R}}))$ mesurables positives. Si $f\leq g$, alors $\displaystyle\int_E f\,d\mu\leq\int_E g\,d\mu$.*

Démonstration. Si $\varphi:E\to\mathbb{R}_+$ est étagée et telle que $\varphi\leq f$ alors $\varphi\leq g$ donc

$$\begin{aligned}&\sup\left\{\int_E\varphi\,d\mu:\varphi:E\to\mathbb{R}\text{ étagée positive et telle que }\varphi\leq f\right\}\\ &\qquad\leq\sup\left\{\int_E\varphi\,d\mu:\varphi:E\to\mathbb{R}\text{ étagée positive et telle que }\varphi\leq g\right\}\end{aligned}$$

ce qui est l'inégalité recherchée. □

Avant de poursuivre, nous avons besoin du résultat suivant qui est le premier résultat fondamental de convergence dans le cadre de l'intégrale de Lebesgue :

Théorème 2.9 (Théorème de Beppo Levi, ou de convergence monotone)**.** *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions mesurables de $(E,\mathscr{A})$ dans $(\overline{\mathbb{R}},\mathscr{B}(\overline{\mathbb{R}}))$. On suppose*

1. *pour tout $n\in\mathbb{N}$, la fonction f_n est positive*[46]*,*
2. *la suite de fonctions $(f_n)_{n\in\mathbb{N}}$ est croissante, c'est-à-dire $\forall n\in\mathbb{N}, f_n\leq f_{n+1}$.*

45. On utilise la même notation car il s'agit d'un prolongement de l'intégrale initialement définie pour les fonctions étagées positives, aux fonctions mesurables positives, voir la remarque 2.7.

46. Du fait de l'hypothèse de monotonie, il suffit que f_0 soit positive ; en pratique, ça ne coûte pas beaucoup de vérifier que toutes les f_n sont positives. Notons aussi bien sûr qu'il suffit que f_n soit positive à partir d'un certain rang.

Alors $f := \lim_{n\to+\infty} f_n$ *est positive et mesurable et*

$$\int_E f\,d\mu = \lim_{n\to+\infty} \int_E f_n\,d\mu.$$

Remarque 2.10. Ainsi l'intégrale $\int_E f\,d\mu$ définie par (2.2) est la limite des intégrales $\int_E f_n\,d\mu$, où $(f_n)_{n\in\mathbb{N}}$ est une suite quelconque de fonctions étagées positives croissante et qui converge vers f. Or une telle suite existe toujours d'après la proposition 1.80 : ceci fournit une définition équivalente de l'intégrale des fonctions mesurables positives. Le théorème de Beppo Levi est donc essentiel pour bien comprendre la construction de l'intégrale de Lebesgue.

Démonstration. La proposition 1.69 montre que f est mesurable. On procède maintenant par double inégalité : montrons d'abord l'inégalité $\geq$. Pour tout $n \in \mathbb{N}$, on a $f_n \leq f_{n+1} \leq f$, par croissance de l'intégrale (proposition 2.8) on a également

$$\int_E f_n\,d\mu \leq \int_E f_{n+1}\,d\mu \leq \int_E f\,d\mu,$$

ce qui prouve en passant à la limite que $\lim_{n\to+\infty} \int_E f_n\,d\mu \leq \int_E f\,d\mu$.

Montrons maintenant l'autre inégalité : par définition de l'intégrale de f, il suffit de montrer que pour toute fonction étagée positive φ telle que $\varphi \leq f$, on a $\int \varphi\,d\mu \leq \lim_{n\to+\infty} \int_E f_n\,d\mu$. Soit un tel φ et $\varepsilon \in]0,1]$. On pose $E_n := \{(1-\varepsilon)\varphi \leq f_n\}$. Montrons d'abord l'égalité $E = \bigcup_{n\in\mathbb{N}} E_n$:

- sur $\{f = 0\}$, $f_n = \varphi = 0$ pour tout entier n, donc $\{f = 0\} \subset E_n$ et par conséquent $\{f = 0\} \subset \bigcup_{n\in\mathbb{N}} E_n$;
- sur $\{f > 0\}$, $(1-\varepsilon)\varphi < f$ car φ ne prend que des valeurs finies. Donc pour tout $x \in \{f > 0\}$, il existe un rang $N(x)$ tel que pour tout $n \geq N(x)$, $(1-\varepsilon)\varphi(x) \leq f_n(x)$, autrement dit $x \in E_n$, et par conséquent $x \in \bigcup_{n\in\mathbb{N}} E_n$.

En conclusion, $E = \{f = 0\} \cup \{f > 0\} \subset \bigcup_{n\in\mathbb{N}} E_n$, et l'autre inclusion est triviale.

Notons $\varphi = \sum_{i\in I} \alpha_i \mathbb{1}_{A_i}$; alors

$$\int_E (1-\varepsilon)\varphi\mathbb{1}_{E_n}\,d\mu = (1-\varepsilon)\int_E \sum_{i\in I} \alpha_i \mathbb{1}_{A_i\cap E_n}\,d\mu = \sum_{i\in I}(1-\varepsilon)\alpha_i\mu(A_i\cap E_n).$$

D'une part, par la proposition 1.42 (continuité à gauche des mesures), $\lim_{n\to+\infty} \mu(A_i \cap E_n) = \mu(A_i)$ pour tout $i \in I$, ce qui s'écrit, I étant fini,

$$\lim_{n\to+\infty} \int_E (1-\varepsilon)\varphi\mathbb{1}_{E_n}\,d\mu = \sum_{i\in I}(1-\varepsilon)\alpha_i\mu(A_i) = (1-\varepsilon)\int_E \varphi\,d\mu.$$

D'autre part pour $n_0 \in \mathbb{N}$, on a $E_{n_0} = \{(1-\varepsilon)\varphi \leq f_{n_0}\}$, donc $(1-\varepsilon)\varphi\mathbb{1}_{E_{n_0}} \leq f_{n_0}$, d'où

$$\int_E (1-\varepsilon)\varphi\mathbb{1}_{E_{n_0}}\,d\mu \leq \int_E f_{n_0}\,d\mu \leq \lim_{n\to+\infty} \int_E f_n\,d\mu.$$

En faisant tendre n_0 vers $+\infty$, on trouve

$$(1-\varepsilon)\int_E \varphi\,d\mu \leq \lim_{n\to+\infty} \int_E f_n\,d\mu.$$

On conclut en faisant tendre ε vers 0. □

Exemple 2.11 (mesure de comptage)**.** Pour poursuivre l'item 2 de l'exemple 2.3, l'intégration par rapport à la mesure de comptage sur $\mathbb{N}$ est tout simplement la sommation de série. En effet la donnée d'une fonction $u : (\mathbb{N}, \mathcal{P}(\mathbb{N})) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mesurable positive est tout simplement la donnée d'une suite $(u_n)_{n\in\mathbb{N}}$ de réels positifs, et si on définit pour tout $N \in \mathbb{N}$, $\varphi_N := (u_n\mathbb{1}_{n\leq N})_{n\in\mathbb{N}}$, on obtient une suite de "fonctions étagées positives" qui converge en croissant quand N tend vers $+\infty$ vers $(u_n)_{n\in\mathbb{N}}$. Donc si m désigne la mesure de comptage sur $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$, alors

$$\int_{\mathbb{N}} u\,dm = \lim_{N\to\infty} \int_{\mathbb{N}} \varphi_N\,dm = \lim_{N\to\infty} \sum_{n=0}^{N} u_n\,m(\{n\}) = \lim_{N\to\infty} \sum_{n=0}^{N} u_n = \sum_{n\in\mathbb{N}} u_n.$$

Grâce au théorème de Beppo Levi, on est maintenant en mesure de voir que les autres propriétés de l'intégrale des fonctions étagées sont conservées :

Proposition 2.12. *Soit $\alpha \geq 0$ et f, g de $(E, \mathscr{A})$ dans $(\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ mesurables positives. Alors*

$$\int_E (\alpha f)\, d\mu = \alpha \int_E f\, d\mu \quad \text{et} \quad \int_E (f+g)\, d\mu = \int_E f\, d\mu + \int_E g\, d\mu.$$

Démonstration. Comme f et g sont mesurables et positives, d'après le lemme d'approximation 1.80 il existe des suites croissantes $(f_n)_{n\in\mathbb{N}}$ et $(g_n)_{n\in\mathbb{N}}$ de fonctions étagées positives croissant vers f et g respectivement. La proposition 2.5 donne les propriétés de positive homogénéité et d'additivité pour ces fonctions étagées, et en appliquant à chaque suite d'intégrales le théorème de Beppo Levi on obtient le résultat. □

On va s'intéresser maintenant à savoir quand l'intégrale d'une fonction mesurable positive peut être nulle. Pour cela, on va utiliser l'inégalité importante suivante :

Proposition 2.13 (Inégalité de Markov)**.** *Soit $f : (E, \mathscr{A}) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ mesurable positive et $a > 0$. On a l'inégalité*

$$\mu\left(\{f \geq a\}\right) \leq \frac{1}{a} \int_E f\, d\mu.$$

Démonstration. Comme $f \geq a\mathbb{1}_{\{f\geq a\}}$, par croissance de l'intégrale (proposition 2.12) on obtient $\displaystyle\int_E f\, d\mu \geq a\mu(\{f \geq a\})$. □

Exercice 2.14. L'inégalité de Markov fait partie de ces résultats qu'on redémontre souvent. On a parfois besoin d'utiliser des variantes comme on le voit dans cet exercice. Soit $\varphi : I \to \mathbb{R}$ une fonction croissante et positive sur un intervalle $I \subset \mathbb{R}$. Soit $f : (E, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mesurable positive et à valeurs dans I, et $a > 0$. Montrer

$$\mu\left(\{f \geq a\}\right) \leq \frac{1}{\varphi(a)} \int_E \varphi \circ f\, d\mu.$$

Proposition 2.15. *Soit $f : (E, \mathscr{A}) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ mesurable positive. Alors*

$$\int_E f\, d\mu = 0 \Longleftrightarrow \mu(\{f \neq 0\}) = 0.$$

Remarque 2.16. Cela peut paraître évident, mais n'oubliez surtout pas l'hypothèse de positivité pour f ! Si f est sans signe ou même à valeurs complexes, on applique très souvent le résultat précédent pour $|f|$ qui est bien mesurable et positive. Ceci donne, puisque f s'annule si et seulement si $|f|$ s'annule :

$$\int_E |f|\, d\mu = 0 \Longleftrightarrow \mu(\{f \neq 0\}) = 0.$$

Démonstration. $\Rightarrow$: pour $n \in \mathbb{N}^*$ on pose $A_n := \{f \geq 1/n\}$. Par l'inégalité de Markov, pour $n \in \mathbb{N}$ on a $\mu(A_n) \leq n \int_E f\, d\mu = 0$ et donc $\mu(A_n) = 0$. Or $\{f \neq 0\} = \bigcup_{n\in\mathbb{N}} A_n$ et cette union est croissante, donc par la proposition 1.42 (continuité à gauche de μ), on obtient $\mu(\{f \neq 0\}) = \lim_{n\to+\infty} \mu(A_n) = 0$.

$\Leftarrow$:

Preuve 1 : commençons par supposer f étagée. Alors $f = \sum_{i\in I} \alpha_i \mathbb{1}_{A_i}$ où I est fini et les α_i sont positifs. Quitte à retirer les α_i nuls, on peut même supposer $\alpha_i > 0$. Comme $\int f d\mu = \sum_{i\in I} \alpha_i \mu(A_i) = 0$, on a pour tout $i \in I$, $\mu(A_i) = 0$, et donc $\mu(\{f \neq 0\}) \leq \sum_{i\in I} \mu(A_i) = 0$.

Si f est mesurable positive et telle que $\mu(\{f \neq 0\}) = 0$, alors toute fonction φ étagée et telle que $0 \leq \varphi \leq f$ sera aussi telle que $\mu(\{\varphi \neq 0\}) = 0$, et donc par le cas précédent, $\int_E \varphi d\mu = 0$. Mais alors $\int_E f d\mu = \sup\left\{\int_E \varphi d\mu, \varphi \text{ étagée telle que } 0 \leq \varphi \leq f\right\} = 0$.

Preuve 2 : On pose $A = \{f \neq 0\}$. Alors

$$\int_E f d\mu = \int_E f\mathbb{1}_A d\mu + \int_E f\mathbb{1}_{A^c} d\mu.$$

Comme f est nulle sur A^c, le deuxième morceau est égal à l'intégrale de la fonction nulle, donc est nul. Pour estimer le premier morceau, on peut écrire avec la croissance de l'intégrale : $\int_E f\mathbb{1}_A d\mu \leq (+\infty)\mu(A) = 0$. □

Remarque 2.17. Au lieu d'écrire $\mu(\{f \neq 0\}) = 0$, on notera souvent

$$f = 0 \quad \mu\text{-presque partout} \qquad \text{ou} \qquad f = 0 \quad \mu\text{-p.p.} \qquad \text{ou} \qquad f(x) = 0 \text{ pour } \mu\text{-presque tout } x \in E.$$

De manière générale, étant donnée une certaine propriété $P(x)$ qui dépend de $x \in E$, si $\{x \in E : P(x) \text{ est fausse}\}$ est un ensemble mesurable de mesure nulle, on dira que $P(x)$ est vraie « pour μ-presque tout x », « pour μ-p.t. x »ou « $\mu(dx)$-presque partout », ou encore que P est vraie μ-p.p.

L'ensemble N sera dit négligeable, ou μ-négligeable. Les ensembles dénombrables, l'ensemble triadique de Cantor, sont des ensembles λ-négligeables où λ est la mesure de Lebesgue.

Corollaire 2.18. *Soient f et g de $(E, \mathscr{A})$ dans $(\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ mesurables positives. Alors*

$$f = g \quad \mu\text{-p.p.} \Rightarrow \int_E f\, d\mu = \int_E g\, d\mu.$$

Démonstration. Attention, rappelons qu'on n'a pas le droit d'écrire le signe "$-$" sans faire attention, ni (dans cette preuve) d'écrire l'intégrale d'une fonction qui n'est pas positive. Pour cette raison, on pose

$$h := \begin{cases} \max(f,g) - \min(f,g) & \text{sur } \{\min(f,g) < \infty\} \\ 0 & \text{sur } \{f = g = \infty\}. \end{cases}$$

Comme $\{f = g\} = \{h = 0\}$, par passage au complémentaire $\{h \neq 0\} = \{f \neq g\}$ donc $\mu(\{h \neq 0\}) = 0$, et par la Proposition 2.15 $\int_E h\, d\mu = 0$. Mais comme $\max(f,g) = \min(f,g) + h$, par additivité on a

$$\int_E \max(f,g)\, d\mu = \int_E \min(f,g)\, d\mu + \int_E h\, d\mu = \int_E \min(f,g)\, d\mu.$$

Et comme $f \wedge g \leq f \leq f \vee g$ et $f \wedge g \leq g \leq f \vee g$, par croissance on a

$$\int_E \max(f,g)\, d\mu = \int_E \min(f,g)\, d\mu = \int_E f\, d\mu = \int_E g\, d\mu,$$

ce qui achève la démonstration. □

On conclut par un lemme qui servira fréquemment, et donne une condition nécessaire d'intégrabilité :

Proposition 2.19. *Soit $f : (E, \mathscr{A}) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ mesurable positive. Alors*

$$\int_E f\, d\mu < +\infty \Longrightarrow \mu(\{f = +\infty\}) = 0.$$

Démonstration. Posons $A := \{f = +\infty\}$.

Preuve 1 : On raisonne par contraposée : si $\mu(A) \neq 0$, alors $\int f\, d\mu \geq \int f \mathbb{1}_A\, d\mu = (+\infty)\mu(A) = +\infty$.

Preuve 2 : On suppose $\int_E f\, d\mu < +\infty$. Posons $A_n := \{f \geq n\}$ pour $n \in \mathbb{N}$. Alors $A = \bigcap_{n \in \mathbb{N}} A_n$ et cette intersection est décroissante. Par l'inégalité de Markov (proposition 2.13), $\mu(A_1) \leq \int_E f\, d\mu < +\infty$. Par la proposition 1.45 (continuité à droite de μ) on a $\mu(A) = \lim_{n\to+\infty} \mu(A_n)$. Mais par l'inégalité de Markov à nouveau, $\mu(A_n) \leq \dfrac{1}{n} \displaystyle\int_E f\, d\mu \xrightarrow[n\to+\infty]{} 0$. □

2.3 Intégrale des fonctions mesurables réelles de signe quelconque

Définition-Proposition 2.20. Soit $f : (E, \mathscr{A}) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ mesurable, Si $\int_E |f|\, d\mu < \infty$, alors on dit que f est μ-intégrable[47] et l'on définit alors l'intégrale de f par rapport à μ par

$$\int_E f\, d\mu = \int_E f^+\, d\mu - \int_E f^-\, d\mu.$$

qui est bien défini car les deux intégrales $\int_E f^+\, d\mu$ et $\int_E f^-\, d\mu$ sont finies[48].

De plus on a l'inégalité :

$$\left| \int_E f\, d\mu \right| \leq \int_E |f|\, d\mu.$$

47. On dira juste "intégrable" s'il n'y a pas d'ambiguïté sur la mesure à laquelle on réfère.
48. Réciproquement, on laisse le lecteur montrer que si ces deux intégrales sont finies, alors $\int_E |f| d\mu < +\infty$.

Démonstration. Les deux intégrales $\int_E f^+\,d\mu$ et $\int_E f^-\,d\mu$ sont positives et majorées par $\int_E |f| d\mu$, par la proposition 2.8 sur la croissance de l'intégrale, elles sont donc en particulier finies, donc on peut les soustraire.

De plus,

$$\left|\int_E f\,d\mu\right| = \left|\int_E f^+\,d\mu - \int_E f^-\,d\mu\right| \leq \int_E f^+\,d\mu + \int_E f^-\,d\mu = \int_E |f|\,d\mu,$$

où on a utilisé l'additivité. □

Définition-Proposition 2.21. On note $\mathscr{L}^1(E, \mathscr{A}, \mu)$ (ou $\mathscr{L}^1(E)$ ou $\mathscr{L}^1(\mu)$ s'il n'y a pas d'ambiguïté) l'espace des fonctions μ-intégrables. C'est un espace vectoriel et l'application $f \mapsto \displaystyle\int_E f\,d\mu$ est une forme linéaire positive [49].

Démonstration. Pour tous $f, g \in \mathscr{L}^1(E)$, pour tout $\alpha \in \mathbb{R}$, $|\alpha f + g| \leq |\alpha|\,|f| + |g|$ donc $\alpha f + g$ est intégrable par additivité. De plus, en se servant des égalités du type $f = f^+ - f^-$ appliquées à f, g et $f+g$, on obtient $f + g = (f+g)^+ - (f+g)^- = f^+ - f^- + g^+ - g^-$ et par conséquent

$$(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+,$$

d'où, par linéarité de l'intégrale de fonctions positives (proposition 2.12),

$$\int_E (f+g)^+\,d\mu + \int_E f^-\,d\mu + \int_E g^-\,d\mu = \int_E (f+g)^-\,d\mu + \int_E f^+\,d\mu + \int_E g^+\,d\mu.$$

Comme toutes ces quantités sont finies, on peut les retrancher, ce qui donne

$$\int_E (f+g)^+\,d\mu - \int_E (f+g)^-\,d\mu = \int_E f\,d\mu + \int_E g\,d\mu,$$

autrement dit $\displaystyle\int_E (f+g)\,d\mu = \int_E f\,d\mu + \int_E g\,d\mu$. De même, en utilisant les égalités $(\alpha f)^+ = \alpha f^+$, $(\alpha f)^- = \alpha f^-$ lorsque $\alpha > 0$ et $(\alpha f)^+ = -\alpha f^-$, $(\alpha f)^- = -\alpha f^+$ lorsque $\alpha < 0$, et en utilisant la positive homogénéité de l'intégrale des fonctions mesurables positives, on obtient :

— dans le cas $\alpha > 0$,

$$\int_E (\alpha f)\,d\mu = \int_E (\alpha f^+)\,d\mu - \int_E (\alpha f^-)\,d\mu = \alpha \int_E f^+\,d\mu - \alpha \int_E f^-\,d\mu = \alpha \int_E f\,d\mu,$$

— dans le cas $\alpha < 0$,

$$\int_E (\alpha f)\,d\mu = \int_E (-\alpha f^-)\,d\mu - \int_E (-\alpha f^+)\,d\mu = (-\alpha) \int_E f^-\,d\mu - (-\alpha) \int_E f^+\,d\mu = \alpha \int_E f\,d\mu.$$

L'intégrale est donc bien une forme linéaire.

Elle est positive car si $f \geq 0$, alors $f = f^+$ et par définition $\int_E f\,d\mu = \int_E f^+\,d\mu \geq 0$. □

Remarque 2.22. La positivité de la forme linéaire $f \mapsto \int_E f d\mu$ implique sa croissance au sens suivant : soit $g, h \in \mathscr{L}^1(E)$ avec $g \geq h$, alors $\int_E g d\mu \geq \int_E h d\mu$. En effet en écrivant $g = (g-h) + h$ avec $g - h \in \mathscr{L}^1(E)$, on a $\int g d\mu = \int (g-h) d\mu + \int h d\mu \geq \int h d\mu$ par linéarité de l'intégrale dans $\mathscr{L}^1(E)$ puis positivité de la fonction $g - h$ et donc de son intégrale.

Remarque 2.23. Si m est la mesure de comptage sur $\mathbb{N}$ alors $\mathscr{L}^1(\mathbb{N}, \mathcal{P}(\mathbb{N}), m)$ est noté $\ell^1(\mathbb{N})$ et est l'ensemble des suites dont la série est absolument convergente, c'est-à-dire

$$\ell^1(\mathbb{N}) = \Big\{(u_n)_{n\in\mathbb{N}},\ \sum_{n\in\mathbb{N}} |u_n| < +\infty\Big\}.$$

49. Si X est un espace de fonctions, on dit d'une forme linéaire L sur X qu'elle est positive si elle satisfait

$$\forall f \in X,\ \ f \geq 0 \ \Rightarrow\ L(f) \geq 0.$$

2.4 Intégrale des fonctions à valeurs complexes

Définition 2.24. Une fonction $f : (E, \mathscr{A}) \to (\mathbb{C}, \mathscr{B}(\mathbb{C}))$ est dite μ-intégrable si f est mesurable et $|f|$ est intégrable. Ceci entraîne que $\mathrm{Re}(f)$ et $\mathrm{Im}(f)$ sont intégrables et l'intégrale de f par rapport à μ est définie comme le nombre complexe

$$\int_E f\, d\mu = \int_E \mathrm{Re}(f)\, d\mu + i \int_E \mathrm{Im}(f)\, d\mu.$$

Remarque 2.25. Du corollaire 2.18 on déduit que si f et g sont deux fonctions à valeurs dans $\overline{\mathbb{R}}$ ou dans $\mathbb{C}$ telles que $f = g$ μ-presque partout, alors f est intégrable si et seulement si g est intégrable, et dans ce cas

$$\int_E f d\mu = \int_E g d\mu.$$

En particulier, il arrivera qu'on calcule des intégrales pour une fonction f qui n'est définie que presque partout sur E. On peut en effet l'étendre à E de la manière qu'on le souhaite, cela ne changera pas le résultat des intégrales ; on peut par exemple l'étendre par 0.

Définition-Proposition 2.26. On note $\mathscr{L}^1_{\mathbb{C}}(E, \mathscr{A}, \mu)$ (ou $\mathscr{L}^1_{\mathbb{C}}(E)$ ou $\mathscr{L}^1(E)$ s'il n'y a pas d'ambiguïté) l'espace des fonctions de $(E, \mathscr{A})$ à valeurs dans $(\mathbb{C}, \mathscr{B}(\mathbb{C}))$ mesurables et μ-intégrables. C'est un espace vectoriel sur $\mathbb{C}$. L'application $f \mapsto \int_E f\, d\mu$ est une $\mathbb{C}$-forme linéaire et pour tout $f \in \mathscr{L}^1_{\mathbb{C}}(E)$,

$$\left| \int_E f\, d\mu \right| \le \int_E |f|\, d\mu. \tag{2.3}$$

Démonstration. Il suffit de montrer la dernière inégalité, le reste découlant de la linéarité de l'intégrale réelle et de la définition. Il existe $z \in \mathbb{C}$ tel que $z \int f\, d\mu$ est réel et $|z| = 1$, donc

$$\left| \int f\, d\mu \right| = |z| \left| \int f\, d\mu \right| = \left| z \int f\, d\mu \right|.$$

Or par linéarité de l'intégrale complexe,

$$z \int f\, d\mu = \int (zf)\, d\mu = \int_E \mathrm{Re}(zf)\, d\mu + i \int_E \mathrm{Im}(zf)\, d\mu.$$

Or on a choisi z pour que $z \int f\, d\mu$ soit réel donc $\int_E \mathrm{Im}(zf)\, d\mu = 0$. De plus, $|\,\mathrm{Re}(zf)\,| \le |\,zf\,| = |\,f\,|$, donc comme

$$\left| \int_E \mathrm{Re}(zf)\, d\mu \right| \le \int_E |\,\mathrm{Re}(zf)\,|\, d\mu \le \int_E |\,f\,| d\mu,$$

on a

$$\left| \int_E f\, d\mu \right| = \left| z \int_E f\, d\mu \right| = \left| \int_E \mathrm{Re}(zf)\, d\mu \right| \le \int_E |\,f\,| d\mu,$$

ce qui constitue l'inégalité souhaitée. □

Exercice 2.27. Comment souvent quand on énonce une inégalité, il peut être utile de connaître une condition nécessaire et suffisante pour qu'il y ait égalité.

1. Dans le cas où f est à valeurs réelles, montrer qu'il y a égalité dans (2.3) si et seulement si f est de signe constant μ presque partout.
2. Dans le cas où f est à valeurs complexes, montrer qu'il y a égalité dans (2.3) si et seulement s'il existe $z_0 \in \mathbb{C}$ tel que $|z_0| = 1$ et $f = z_0|f|$ presque partout sur E (autrement dit, cela signifie que f a un argument constant p.p.).

2.5 Comparaison avec l'intégrale de Riemann

L'énoncé suivant permet de voir que l'intégrale de Lebesgue généralise l'intégrale de Riemann sur un segment.

Proposition 2.28. *[Comparaison Riemann-Lebesgue sur un segment] Soit $f : [a,b] \to \mathbb{R}$ une fonction continue par morceaux. Alors $f \in \mathscr{L}^1([a,b], \mathscr{B}([a,b]), \lambda)$ et*

$$\int_a^b f(t)dt = \int_{[a,b]} f d\lambda.$$

Remarque 2.29. Pour cette raison, nous pouvons utiliser les notations habituelles de l'intégrale de Riemann plutôt que d'utiliser la notation $d\lambda$; les notations de l'intégrale de Lebesgue seront en général réservées aux situations où on utilise d'autres mesures que la mesure de Lebesgue.

Remarque 2.30. L'énoncé précédent n'est pas valable si on remplace $[a,b]$ par un intervalle non compact. Néanmoins si par exemple $I = [a,b[$ avec $a \leq b \leq +\infty$, alors toute fonction $f : [a,b[\to \mathbb{R}$ continue par morceaux est borélienne et

$$f \in \mathscr{L}^1([a,b[, \mathscr{B}([a,b[), \lambda) \quad \Leftrightarrow \quad \text{l'intégrale } \int_a^b f(t)dt \text{ est absolument convergente,}$$
$$\text{i.e. } \lim_{A \overset{<}{\to} b} \int_a^A |f(t)|dt \in \mathbb{R}.$$

Il est fréquent de citer la fonction

$$x \in [0, +\infty[\longmapsto \frac{\sin(x)}{x}$$

(prolongée par 1 en 0) comme un exemple intéressant, mais il faut être vigilant avec la terminologie ; cette fonction admet une intégrale impropre [50] au sens de Riemann [51], mais l'intégrale de sa valeur absolue est divergente ; il est raisonnable d'écrire

$$\int_0^{+\infty} \left| \frac{\sin(x)}{x} \right| dx = +\infty.$$

On dit qu'elle n'est pas intégrable sur $\mathbb{R}_+$ (que ce soit au sens de Lebesgue ou au sens de Riemann [52]).

Remarque 2.31.
- La proposition 2.28 et sa preuve sont valables pour les fonctions réglées sur $[a,b]$.
- [** voir par exemple [BP12, page 124]] Le cas des fonctions Riemann intégrables sur $[a,b]$ est plus subtile ; on peut montrer que si $f : [a,b] \to \mathbb{R}$ est Riemann-intégrable, alors il existe $g \in \mathscr{L}^1([a,b], \mathscr{B}([a,b]), \lambda)$ telle que

$$f = g \quad \lambda\text{-p.p.}, \qquad \text{et} \quad \int_a^b f(t)dt = \int_{[a,b]} g d\lambda.$$

En effet comme on l'a évoqué à l'exercice 1.70, il se peut que f ne soit pas borélienne. Par contre, si on remplace $\mathscr{B}([a,b])$ par la tribu complétée notée $\overline{\mathscr{B}([a,b])}^{\lambda}$, alors on peut effectivement dire que si $f : [a,b] \to \mathbb{R}$ est Riemann-intégrable, alors $f \in \mathscr{L}^1([a,b], \overline{\mathscr{B}([a,b])}^{\lambda}, \lambda)$ et

$$\int_a^b f(t)dt = \int_{[a,b]} f d\lambda.$$

Démonstration. Les fonctions continues par morceaux sur un segment sont mesurables, et bornées (il suffit d'appliquer le théorème des bornes sur chaque subdivision de $[a,b]$), donc elles sont dans $\mathscr{L}^1([a,b], \mathscr{B}([a,b]), \lambda)$. Il reste à prouver l'égalité des intégrales.

50. On peut aussi dire "intégrale généralisée" ou "intégrale semi-convergente".
51. On peut le montrer par une intégration par partie.
52. En fait, avec l'intégrale de Riemann comme celle de Lebesgue, on conseille de réserver la mot "intégrable" au cas où l'intégrale de la valeur absolue est convergente. Si une fonction est continue par morceaux, alors elle sera (avec cette terminologie) intégrable au sens de Riemann si et seulement elle l'est au sens de Lebesgue.

Commençons par supposer $f : [a, b] \to \mathbb{R}$ en escalier. Il existe $p \in \mathbb{N}^*$ et $x_0 = a < x_1 < \ldots < x_p = b$ et $(\alpha_i)_{i \in [\![0,p-1]\!]} \in \mathbb{R}^p$ tels que

$$\forall i \in [\![0, p-1]\!],\ \forall x \in]x_i, x_{i+1}[, \quad f(x) = \alpha_i.$$

Cette fonction est borélienne (voir l'exercice 1.70) et comme les singletons sont de mesure de Lebesgue nulle, on a $|f| = \sum_{i=0}^{p-1} |\alpha_i| \mathbb{1}_{]x_i,x_{i+1}[}$ λ-presque partout et donc

$$\int_{[a,b]} |f| d\lambda = \sum_{i=0}^{p-1} |\alpha_i|.(x_{i+1} - x_i)$$

donc $f \in \mathscr{L}^1([a, b], \mathscr{B}([a, b]), \lambda)$ et

$$\int_{[a,b]} f d\lambda = \sum_{i=0}^{p-1} \alpha_i.(x_{i+1} - x_i) = \int_a^b f(t)dt.$$

Si maintenant f est plus généralement continue par morceaux : d'abord, f est encore borélienne (voir le corollaire 1.66 et l'exercice 1.70). Ensuite, par le lemme 5.4 il existe une suite $(f_n)_{n\in\mathbb{N}}$ de fonctions en escalier qui converge uniformément vers f. D'une part, par la proposition 4.11 du Chapitre I on a que la convergence uniforme implique la convergence des intégrales de Riemann :

$$\int_a^b f_n(t)dt \xrightarrow[n\to+\infty]{} \int_a^b f(t)dt.$$

D'autre part, ceci est aussi vrai pour l'intégrale de Lebesgue, avec la même preuve que la proposition 4.11 :

$$\left| \int_{[a,b]} f_n d\lambda - \int_{[a,b]} f d\lambda \right| \leq \int_{[a,b]} |f_n - f| d\lambda \leq \|f_n - f\|_\infty \lambda([a, b]).$$

Comme pour tout $n \in \mathbb{N}$, $\int_{[a,b]} f_n d\lambda = \int_a^b f_n(t)dt$, par passage à la limite, on a l'égalité $\int_{[a,b]} f d\lambda = \int_a^b f(t)dt$. □

Remarque 2.32. La proposition précédente peut être vue via la propriété d'unicité dans le corollaire 2.51 du Chapitre III. En effet on a montré dans la preuve ci-dessus que l'intégrale de Lebesgue et de Riemann coïncident sur l'espace des fonctions en escalier, et comme elles sont continues si on les voit définies de $(C^0_m([a, b]), \|\cdot\|_\infty) \to \mathbb{R}$, et que l'espace des fonctions en escalier est dense dans $(C^0_m([a, b]), \|\cdot\|_\infty)$, on conclut qu'elles sont égales sur cet espace. On a en fait reproduit la preuve d'unicité dans ce cas particulier [53].

3 Théorèmes de convergence

3.1 Conséquences du théorème de convergence monotone

On a déjà vu le théorème de convergence monotone (théorème 2.9) qui est le premier résultat fondamental. Voyons une conséquence importante :

Proposition 3.1 (Interversion série-intégrale, cas positif). *Pour toute suite $(f_n)_{n\in\mathbb{N}}$ de fonctions de $(E, \mathscr{A})$ dans $(\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ mesurables et positives, la fonction $\sum_{n\in\mathbb{N}} f_n$ est bien définie, mesurable et positive, et*

$$\int_E \left(\sum_n f_n \right) d\mu = \sum_n \int_E f_n \, d\mu.$$

53. Ce n'est pas dans l'unicité que réside la difficulté du corollaire 2.51 ou plus généralement du théorème 1.87 du Chapitre III ; il est fréquent (mais pas obligé !) dans ce genre de situation que l'on reproduise la preuve plutôt que d'appliquer directement ces résultats (ça ne "coûte" pas plus cher, au sens où refaire la démonstration n'est pas vraiment plus difficile que de montrer que les hypothèses des-dits résultats sont satisfaites), et on réserve l'application de ces résultats aux situations où on n'a besoin de "toute leur force".

Démonstration. On pose $g_n := \sum_{k=0}^{n} f_k$, on utilise l'additivité de l'intégrale et on applique le théorème de Beppo Levi à la suite croissante $(g_n)_{n\in\mathbb{N}}$. □

Remarque 3.2. Ce résultat est caractéristique de la flexibilité de l'intégrale de Lebesgue quand toutes les fonctions sont supposées positives. On verra à la proposition 3.12 une adaptation de ce résultat au cas où les fonctions ne sont plus supposées à valeurs réelles positives.

Exercice 3.3. Soit $f : (E, \mathscr{A}) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ positive. Montrer que l'application

$$\begin{aligned} \nu : \mathscr{A} &\longrightarrow [0, +\infty] \\ A &\longmapsto \int_E f \mathbb{1}_A \, d\mu \end{aligned}$$

est une mesure sur $(E, \mathscr{A})$. On l'appelle mesure de densité f par rapport à μ.

Exercice 3.4 (Lemme de Borel–Cantelli, partie 1 [54]). Soit (A_n) une suite d'éléments de $\mathscr{A}$ telle que

$$\sum_{n\geq 0} \mu(A_n) < +\infty.$$

1. Montrer que la fonction $f = \sum_{n\in\mathbb{N}} \mathbb{1}_{A_n}$ est intégrable.
2. En déduire que (voir l'exercice 1.47 pour la définition de la limite supérieure d'une suite d'ensembles) :

$$\sum_{n\geq 0} \mu(A_n) < +\infty \Longrightarrow \mu\left(\limsup_{n\to+\infty} A_n\right) = 0.$$

3.2 Lemme de Fatou, théorème de convergence dominée et conséquence

Le second résultat fondamental de convergence, souvent oublié par les étudiants, est le lemme de Fatou ; il est vrai qu'il sert moins souvent, notamment parce qu'il ne donne qu'un résultat partiel sur l'interversion limite/intégrale, mais il est parfois le seul résultat qui peut être appliqué, et il arrive que l'inégalité obtenue suffise pour l'objectif recherché. De plus, c'est l'outil qui permettra de démontrer le résultat le plus important de convergence, à savoir le théorème de convergence dominée qu'on donne juste après.

Lemme 3.5 (Lemme de Fatou). *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonction de $(E, \mathscr{A})$ dans $(\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ mesurables et positives. Alors $\liminf_{n\to\infty} f_n$ est mesurable et positive, et*

$$\int_E \liminf_{n\to+\infty} f_n \, d\mu \leq \liminf_{n\to+\infty} \int_E f_n \, d\mu. \tag{3.1}$$

Remarque 3.6. La présence des $\liminf$ est un moyen de ne pas supposer de convergence. Il est cependant fréquent qu'on l'applique dans des cas où l'une ou l'autre des $\liminf$ est en fait une limite.

Remarque 3.7. Comme tout résultat de ce type, il vaut mieux avoir un moyen mnémotechnique efficace pour se souvenir de quelle est la bonne inégalité. Je propose [55] :

— on peut retenir que ce résultat permet dans certains cas de montrer que la fonction $\liminf_{n\to+\infty} f_n$ est intégrable. En effet si le terme de droite de (3.1) est fini, alors même si on n'obtient qu'une inégalité, on récupère que le terme de gauche est fini. Si l'inégalité était inversée, on ne pourrait pas obtenir ce type de résultat.

— on peut tester sur un exemple simple, qui d'ailleurs convaincra le lecteur qu'il n'y a pas égalité en général : si $(E, \mathscr{A}, \mu) = (\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$ et $f_n = \mathbb{1}_{[n,n+1[}$, alors

$$\text{d'une part,} \quad \liminf_{n\to+\infty} f_n = 0, \quad \text{donc} \quad \int_{\mathbb{R}} \liminf_{n\to+\infty} f_n d\lambda = 0,$$

$$\text{et d'autre part} \quad \forall n \in \mathbb{N}, \int_{\mathbb{R}} f_n d\lambda = 1, \quad \text{donc} \quad \liminf_{n\to+\infty} \left(\int_{\mathbb{R}} f_n d\lambda\right) = 1.$$

54. Il y a une partie 2 qui consiste à prouver une réciproque dans le cas où les évènements $(A_n)_{n\in\mathbb{N}}$ sont indépendants.
55. Mais évidemment, trouvez votre moyen à vous ; tant qu'il fonctionne sans erreur et en un temps court, ça me va.

Démonstration. La proposition 1.69 montre que $\liminf_{n\to\infty} f_n$ est mesurable. Posons pour $n \in \mathbb{N}$, $g_n := \inf_{k\geq n} f_k$ et $g := \liminf_{n\to+\infty} f_n = \lim_{n\to+\infty} g_n$. Comme g est la limite de la suite croissante et positive $(g_n)_{n\in\mathbb{N}}$, le théorème de Beppo Levi assure que

$$\int_E g\, d\mu = \lim_{n\to+\infty} \int_E g_n\, d\mu.$$

D'autre part, pour tout $n \in \mathbb{N}$, $g_n \leq f_n$ donc par croissance de l'intégrale, $\int_E g_n\, d\mu \leq \int_E f_n\, d\mu$ et

$$\liminf_{n\to+\infty} \int_E g_n\, d\mu \leq \liminf_{n\to+\infty} \int_E f_n\, d\mu.$$

Mais d'après ce qui précède, $\liminf_{n\to+\infty} \int_E g_n\, d\mu = \lim_{n\to+\infty} \int_E g_n\, d\mu = \int_E g\, d\mu$, ce qui fournit l'inégalité souhaitée. $\square$

On en vient au résultat le plus célèbre et le plus utilisé de la théorie de Lebesgue :

Théorème 3.8 (Théorème de convergence dominée de Lebesgue). *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions mesurables de $(E, \mathscr{A})$ dans $(\mathbb{K}, \mathscr{B}(\mathbb{K}))$ convergeant μ-p.p.*[56] *vers une fonction f mesurable, et telle qu'il existe $g \in \mathscr{L}^1(E)$ telle que pour tout $n \in \mathbb{N}$, $|f_n| \leq g$ μ-p.p. Alors pour tout $n \in \mathbb{N}$, $f_n \in \mathscr{L}^1(E)$, $f \in \mathscr{L}^1(E)$ et*

$$\lim_{n\to\infty} \int_E |f - f_n|\, d\mu = 0, \tag{3.2}$$

et en particulier

$$\lim_{n\to\infty} \int_E f_n\, d\mu = \int_E f\, d\mu. \tag{3.3}$$

Remarque 3.9. L'hypothèse $\forall n \in \mathbb{N},\ \ |f_n| \leq g$ μ-*p.p* signifie rigoureusement

$$\forall n \in \mathbb{N},\ \ \exists N_n \in \mathscr{A} \text{ tel que } \mu(N_n) = 0,\ \ \forall x \in N_n^c,\ \ |f_n(x)| \leq g(x),$$

c'est-à-dire que l'ensemble négligeable qui apparaît dépend a priori de n. Si on veut par exemple fixer x et faire tendre n vers l'infini, ça n'est pas possible car l'inégalité est valable pour un ensemble de x qui dépend de n. L'astuce très classique dans ce contexte est la suivante : on considère $N_\infty = \bigcup_{n\in\mathbb{N}} N_n$ qui est encore de mesure nulle car comme l'union est dénombrable, $N_\infty \in \mathcal{A}$ et on peut écrire $\mu(N_\infty) \leq \sum_{n\in\mathbb{N}} \mu(N_n) = 0$, et de plus on a cette fois

$$\forall x \in N_\infty^c,\ \forall n \in \mathbb{N},\ \ \ |f_n(x)| \leq g(x).$$

Ce type d'échange de quantificateur faisant intervenir l'expression μ-p.p. fonctionne quand on a un quantificateur sur une variable qui vit dans un ensemble dénombrable (ici $n \in \mathbb{N}$). Si on a un quantificateur sur un ensemble non dénombrable, il faut se méfier, et voir si on pourrait se ramener à un quantificateur dénombrable (par exemple remplacer $\forall \varepsilon > 0$ par $\forall \varepsilon \in \mathbb{Q}_+^*$).

Démonstration. On pose $A := \left\{x \in E : \lim_{n\to\infty} f_n(x) = f(x)\right\} \cap \left(\bigcap_{n\in\mathbb{N}} \{|f_n| \leq g\}\right)$. Alors A est de complémentaire négligeable et pour tout $x \in A$ et tout $n \in \mathbb{N}$, $|\,f_n(x)\,| \leq g(x)$, donc par passage à la limite, $|f(x)| \leq g(x)$. Par conséquent, $\int |f|\, d\mu \leq \int g\, d\mu < \infty$ donc $f \in \mathscr{L}^1(E)$.

Pour obtenir (3.2), on remarque que pour tout $n \in \mathbb{N}$, $|f - f_n| \leq |f| + |f_n| \leq 2g \in \mathscr{L}^1(E)$ et on applique le lemme de Fatou à la suite de fonctions $\big(2g - |f_n - f|\big)_{n\in\mathbb{N}}$ qui est mesurable et presque partout positive. On obtient

$$\int_E \liminf_{n\to+\infty} \left[2g - |f_n - f|\right] d\mu \leq \liminf_{n\to+\infty} \int_E \left[2g - |f_n - f|\right] d\mu$$

56. Cela signifie qu'il existe $A \in \mathscr{A}$ tel que $\mu(A^c) = 0$ et

$$\forall x \in A,\quad f_n(x) \xrightarrow[n\to+\infty]{} f(x).$$

c'est-à-dire
$$\int_E 2g\,d\mu - \int_E \limsup_{n\to+\infty} |f_n - f| d\mu \leq \int_E 2g\,d\mu - \limsup_{n\to+\infty} \int_E |f_n - f| d\mu$$
par linéarité, soit encore, comme g est intégrable [57]
$$\int_E \limsup_{n\to+\infty} |f - f_n|\,d\mu \geq \limsup_{n\to+\infty} \int_E |f - f_n|\,d\mu$$
puisque [58] $g \in \mathscr{L}^1(E)$. Ceci implique $\limsup\limits_{n\to+\infty} \int_E |f - f_n|\,d\mu = 0$, d'où (3.2).

L'équation (3.3) s'obtient avec (2.3) :
$$\left| \int_E f_n\,d\mu - \int_E f\,d\mu \right| \leq \int_E |f_n - f| d\mu \to 0.$$

□

Exercice 3.10. Les théorèmes de la théorie de Lebesgue peuvent être utile dans le cadre de séries qui dépendent d'un paramètre n, car on peut voir ces séries (en la variable m) comme les intégrales par rapport à la mesure de comptage. Voici quelques exemples où on peut étudier la convergence avec les théorèmes de ce paragraphe : étudier la limite quand n tend vers $+\infty$ de

1. $\displaystyle\sum_{k=1}^{+\infty} \frac{n}{k} \sin(\frac{1}{nk})$,
2. $\displaystyle\sum_{k=1}^{+\infty} \frac{n+k}{nk^{3/2} + k^3}$.

Exercice 3.11 (Théorème de convergence monotone pour des suites décroissantes ?)**.** Soient $(E, \mathcal{A}, \mu)$ un espace mesuré et $(f_n)_{n\geq 0}$ une suite décroissante de fonctions mesurables positives qui converge μ-p.p. vers une fonction f.

1. On suppose qu'il existe n_0 tel que $\int_E f_{n_0} d\mu < \infty$. Montrer que le théorème de convergence dominée s'applique [59] et que $\lim\limits_{n\to\infty} \int_E f_n d\mu = \int_E f d\mu$.
2. Montrez un contre-exemple quand on enlève l'hypothèse dans la question précédente.

Le résultat suivant traite de la même question que la proposition 3.1 quand les fonctions ne sont plus supposées positives [60] :

Proposition 3.12 (Interversion série-intégrale)**.** *Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions mesurables de $(E, \mathscr{A})$ et à valeurs dans $\overline{\mathbb{R}}$ ou $\mathbb{C}$. Si la série de terme général $\left(\int_E |f_n|\,d\mu\right)_{n\in\mathbb{N}}$ est convergente, alors la fonction $\sum_{n\in\mathbb{N}} f_n$ est définie μ-p.p., intégrable et*
$$\int_E \left(\sum_{n\in\mathbb{N}} f_n \right) d\mu = \sum_{n\in\mathbb{N}} \left(\int_E f_n\,d\mu \right).$$

57. Je précise cela pour voir que $\int g d\mu$ n'est pas $+\infty$: sinon je ne pourrais pas "simplifier" dans l'inégalité.
58. Oui, attention ! On a supprimé à gauche et à droite de l'inégalité le terme $\int_E g d\mu$, ce qui veut dire qu'on l'a soustrait aux deux côtés de l'inéquation, ce qui n'est possible que parce que ce nombre est fini. Rappelons si besoin que $+\infty - 2 \leq +\infty - 3$, mais si on simplifie $+\infty$ on obtient $-2 \leq -3$...
59. Pour cette raison, on ne parle pas de théorème de convergence monotone pour des suites décroissantes de fonctions positives ; les hypothèses d'un possible résultat seront de toute façon plus fortes que celles du théorème de convergence dominée. Il n'empêche qu'il peut s'avérer une bonne idée occasionnellement de constater la décroissance en le paramètre pour trouver une domination.
60. Attention néanmoins, il ne s'agit pas d'une généralisation, car on va devoir faire plus d'hypothèse dans le cas où on ne sait pas que les fonctions sont positives.

Remarque 3.13. Cet énoncé est caractéristique du bon réflexe à avoir quand on veut justifier une interversion somme/intégrale [61] (on verra que c'est la même chose pour une intégrale double) : on met les valeurs absolues "à l'intérieur" et on calcule

$$\int_E \left(\sum_{n\in\mathbb{N}} |f_n| \right) d\mu. \tag{3.4}$$

Pour ce calcul on peut appliquer la proposition 3.1 : si le résultat de (3.4) est fini, alors les calculs sont justifiables sans le module.

Notez qu'on gagne "en prime" le fait que la fonction $\sum_{n\in\mathbb{N}} f_n$ est bien définie presque partout, ce qui n'a rien d'évident a priori [62].

Démonstration. Posons $g := \sum_n |f_n|$. Alors par la proposition 3.1 et l'hypothèse faite sur la série de terme général $\left(\int_E |f_n|\, d\mu\right)_{n\in\mathbb{N}}$, on obtient $\int_E g\, d\mu < \infty$. Par la proposition 2.19, on en déduit $g < \infty$ μ-p.p.. Autrement dit il existe une partie mesurable A de E, de complémentaire μ-négligeable, telle que $g(x) < \infty$ pour tout $x \in A$. Pour tout $x \in A$, la série de terme général $f_n(x)$ est absolument convergente, donc convergente. La suite $S_n := \sum_{k\leq n} f_k$ converge donc sur A, vers une fonction $S := \sum_{n\in\mathbb{N}} f_n$; et comme les fonctions $|S_n|$ sont dominées par g intégrable, le théorème de convergence dominée 3.8 assure alors que

$$\int_E S\, d\mu = \lim_{n\to\infty} \int_E S_n\, d\mu = \lim_{n\to\infty} \sum_{k=0}^{n} \int_E f_k\, d\mu = \sum_{n\in\mathbb{N}} \int_E f_n\, d\mu,$$

ce qui achève la démonstration. □

Remarque 3.14. Relisez la preuve 2 du théorème 3.42 du Chapitre II où on avait utilisé ce résultat. Voir aussi la question 2 de l'exercice 3.15 pour une situation similaire.

Exercice 3.15.

1. Montrer que : $\displaystyle\int_0^{+\infty} \frac{\sin x}{e^x - 1}\, dx = \sum_{n\geq 1} \frac{1}{n^2+1}$.
2. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction borélienne telle que pour tout $a \in \mathbb{R}$, la fonction $x \mapsto e^{ax} f(x)$ est intégrable. Montrer que pour tout $z \in \mathbb{C}$, $\displaystyle\int_{\mathbb{R}} e^{zx} f(x) dx = \sum_{n\geq 0} \frac{z^n}{n!} \int_{\mathbb{R}} x^n f(x) dx$.

3.3 Intégrales à paramètre

On aborde ici des théorèmes absolument centraux du programme, et qui interviennent dans la quasi-totalité des écrits d'Analyse & Probabilités au concours de l'agrégation. Ils ont également leur place dans de nombreux contextes à l'oral. Ils viennent compléter le théorème 3.42 qu'on a vu au Chapitre II ; on invite le lecteur à relire ce dernier à la lumière des deux énoncés suivants. Cette comparaison est un parfait exemple pour comprendre la spécificité de la théorie des fonctions holomorphes.

Théorème 3.16 (Continuité des intégrales à paramètre)**.** *Soit $f : I \times E \to \mathbb{R}$, où I est un intervalle de $\mathbb{R}$. On suppose que*

1. *pour tout $t \in I$, $f(t, \cdot) : (E, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ est mesurable,*
2. *pour μ-presque tout $x \in E$, $f(\cdot, x)$ est continue sur I,*

61. La faute très répandue étant d'écrire les objets $\sum_{n\in\mathbb{N}} f_n$ ou $\int_E \sum_{n\in\mathbb{N}} f_n d\mu$ sans valeurs absolues, alors qu'on n'a pas justifié que ces objets ont un sens. Ces fautes sont fortement pénalisées dans les épreuves, à l'écrit comme à l'oral, car elles témoignent d'un gros manque de rigueur du candidat qui ne réalise pas qu'en mathématiques, certains objets n'ont pas de sens et donc n'ont pas le droit d'être écrits ; ici les sommes infinies ou les intégrales sont par définition des limites, et il faut justifier de leur convergence pour pouvoir les écrire. Quand on met des valeurs absolues, on se prémunit de tout problème car par les résultats du cours (basés notamment sur le théorème 2.8 du Chapitre I) ces objets existent forcément dans $\mathbb{R}_+ \cup \{+\infty\}$.

62. Notez bien l'idée derrière la preuve de ce fait, qui est un peu inhabituelle et à laquelle il faut penser : pour montrer que la série est convergente presque partout, on considère la somme des valeurs absolues qui est toujours définie mais à valeurs dans $\overline{\mathbb{R}}_+$; on l'intègre, et comme le résultat est fini, on récupère que la somme des valeurs absolues est finie presque partout, et donc la somme sans valeurs absolues est définie presque partout.

3. *Hypothèse de domination : il existe* $g : (E, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *intégrable telle que*

$$\text{pour } \mu\text{-presque tout } x, \text{ pour tout } t \in I, \quad |f(t,x)| \leq g(x),$$

Alors la fonction $h : t \mapsto \int_E f(t,x)\, d\mu(x)$ *est bien définie et continue sur* I.

Remarque 3.17. Il est bien entendu qu'il faut connaître parfaitement les hypothèses d'un tel théorème ; mais cela n'empêche pas de prendre un peu de recul sur celles-ci. La première hypothèse est bien naturelle, puisqu'elle est une hypothèse de régularité nécessaire (non suffisante !) pour donner un sens à la fonction h. De plus, quand on travaille[63] dans l'espace mesuré $(J, \mathscr{B}(J), \lambda)$ où J est un intervalle de $\mathbb{R}$, alors on a déjà remarqué que cette hypothèse ne coûtait pas cher. La seconde hypothèse ne vous échappera pas au moment où vous devez appliquer un tel théorème, elle est à nouveau bien naturelle au vu de la conclusion espérée. C'est la 3ième hypothèse qui est cruciale et qui peut amener son lot de confusions ; il faut d'une part ne pas se tromper dans les variables mises en jeu, c'est bien une fonction qui ne dépend pas du paramètre que l'on cherche. Il n'y a pas d'erreur possible si on comprend les rôles différents de t (le "paramètre") et x (la variable d'intégration) : en effet, on demande que la domination soit une fonction intégrable, il s'agit bien sûr d'une intégrabilité en la variable d'intégration, il est donc naturelle que cette fonction en dépende, et soit indépendante de l'"autre" variable. Ensuite, le lecteur doit garder à l'esprit qu'obtenir une telle hypothèse peut dans certains être une réelle difficulté ; dans d'autres, ce sera une formalité. Pour arriver à améliorer sa capacité à "trouver" de telles dominations, on conseille le lecteur de s'entraîner.

Remarque 3.18. On a bien demandé que $f(t,\cdot)$ soit mesurable, et non intégrable, pour tout $t \in I$. En effet, l'hypothèse de domination implique que $f(t,\cdot)$ est intégrable, il est donc inutile de l'ajouter dans les hypothèses.

Remarque 3.19. On aurait pu remplacer I par un espace métrique quelconque (X, d) où vivra la variable t. On obtient ainsi une fonction $h : t \in X \mapsto \int_E f(t,x) d\mu(x)$ continue sur X. La preuve ci-dessous s'applique sans changement, tout reposant sur la caractérisation séquentielle de la continuité qui est aussi valable dans ce cadre (voir la proposition 1.58 au Chapitre III).

Démonstration. Soit $t \in I$: la domination μ-p.p. de $f(t,\cdot)$ par la fonction intégrable g garantit que $f(t,\cdot)$ est intégrable et donc que h est bien définie. Soit une suite $(t_n)_{n\in\mathbb{N}}$ de I qui converge vers t : on veut montrer que $\lim_{n\to+\infty} h(t_n) = h(t)$. On pose pour $n \in \mathbb{N}$, $f_n(x) := f(t_n, x)$. De la l'hypothèse de continuité de la fonction $f(\cdot, x)$ en t, pour presque tout x, on déduit la convergence presque partout de f_n vers $f(t,\cdot)$. Comme la suite $(f_n)_{n\in\mathbb{N}}$ est dominée presque partout par la fonction intégrable g, le théorème de convergence dominée 3.8 assure que

$$\lim_{n\to+\infty} h(t_n) = \lim_{n\to\infty} \int_E f_n\, d\mu = \int_E f(t,x)\, d\mu(x) = h(t).$$

Par la proposition 3.4 du Chapitre I, on conclut que h est continue sur I. □

Remarque 3.20. Au vu de la preuve, le lecteur se convaincra qu'on aurait pu donner une version "ponctuelle" du théorème : si on suppose les fonctions $t \mapsto f(t,x)$ continues en $t_0 \in I$, pour presque tout $x \in E$, et que les autres hypothèses sont inchangées, on obtient que h est bien définie sur I et continue en t_0.

Dans ce contexte, il suffit d'avoir l'hypothèse de domination pour t sans un voisinage de t_0.

Théorème 3.21 (Dérivabilité des intégrales à paramètre)**.** *Soit* $f : I \times E \to \mathbb{R}$, *où* I *est un intervalle de* $\mathbb{R}$. *On suppose que*

1. *pour tout* $t \in I$, $f(t,\cdot) : (E, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *est mesurable,*
2. *pour* μ*-presque tout* $x \in E$, $t \mapsto f(t,x)$ *est dérivable sur* I,
3. *il existe* $g_0 : (E, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *intégrable telle que*

$$\text{pour } \mu\text{-presque tout } x, \text{ pour tout } t \in I, \quad |f(t,x)| \leq g(x),$$

63. C'est ce qui arrivera une très grande majorité du temps.

4. il existe $g_1 : (E, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ intégrable telle que

$$\text{pour } \mu\text{-presque tout } x, \text{ pour tout } t \in I, \quad \left|\frac{\partial f}{\partial t}(t,x)\right| \leq g_1(x),$$

alors la fonction $h : t \mapsto \int_E f(t,x)\, d\mu(x)$ est bien définie et est dérivable sur I, et

$$\forall t \in I, \quad h'(t) = \int_E \frac{\partial f}{\partial t}(t,x)\, d\mu(x).$$

Remarque 3.22. Si on remplace "$f(\cdot,x)$ dérivable pour presque tout x" par "$f(\cdot,x)$ de classe C^1 pour presque tout x", alors on obtient que h est de classe C^1 sur I. On peut facilement itérer : si $n \in \mathbb{N}$, et si pour tout $k \in [\![0,n]\!]$ il existe g_k intégrable telle que

$$\text{pour } \mu\text{-presque tout } x, \text{ pour tout } t \in I, \quad \left|\frac{\partial^k f}{\partial t^k}(t,x)\right| \leq g_k(x),$$

alors h est de classe C^n sur I.

Remarque 3.23. On a donc demandé une domination de f et de $\frac{\partial f}{\partial t}$. On trouve rarement cette version du théorème dans la littérature. On trouve plus souvent la version qui consiste à remplacer les hypothèses 1 et 3 par

$$1'. \text{ pour tout } t \in I, f(t,\cdot) : (E,\mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R})) \text{ est intégrable.}$$

Cette version est meilleure au sens où ses hypothèses sont un peu plus faibles, car elle évite de trouver une domination de la fonction f : il suffit de vérifier l'intégrabilité de $f(t,\cdot)$ pour tout t, ce qui est plus faible que de trouver une domination. Néanmoins :

1. cette version améliorée me paraît plus difficile à retenir et amène des mélanges entre les théorèmes 3.16 et 3.21 ;
2. dans la majorité des cas, on préfère commencer à montrer que la fonction h est continue, ce qui revient à appliquer le théorème 3.16 ; on trouve donc une domination g_0 de f pour montrer la continuité de h, et pour passer à la dérivabilité, on cherche une domination g_1 de $\frac{\partial f}{\partial t}$.
3. dans de nombreux cas, on veut en fait itérer le résultat précédent à tout ordre, pour montrer que h est infiniment dérivable. Alors on n'échappe par à trouver pour tout $k \in \mathbb{N}$ une fonction g_k intégrable sur E et telle que

$$\text{pour } \mu\text{-presque tout } x, \text{ pour tout } t \in I, \quad \left|\frac{\partial^k f}{\partial t^k}(t,x)\right| \leq g_k(x).$$

Remarquons que la démonstration de cette version "améliorée" est la même, seule la première ligne dans ce qui suit change, le lecteur se convaincra que la domination n'a été utilisée que pour voir que h est bien définie.

Démonstration. D'après l'hypothèse de domination par g_0, la fonction h est bien définie[64].

Soit $t \in I$ et $(s_n)_{n\in\mathbb{N}}$ une suite de I convergeant vers t telle que $s_n \neq t$ pour tout $n \in \mathbb{N}$. Pour $n \in \mathbb{N}$ on pose $f_n : (E, \mathscr{A}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ la fonction intégrable

$$f_n(x) := \frac{f(s_n,x) - f(t,x)}{s_n - t} \qquad x \in E.$$

Alors pour p.t. x, la suite $(f_n)_{n\in\mathbb{N}}$ converge vers $\frac{\partial f}{\partial t}(t,\cdot)$, ce qui fait de cette dérivée partielle une fonction mesurable de x, par la proposition 1.69. L'hypothèse de domination p.p. de cette fonction mesurable par la fonction intégrable g_1 garantit que $\displaystyle\int_E \frac{\partial f}{\partial t}(t,x)\, d\mu(x)$ est bien définie pour tout $t \in I$. De plus,

$$\forall n \in \mathbb{N}, \quad \int_E f_n\, d\mu = \frac{1}{s_n - t}\left(\int_E f(s_n,\cdot)\, d\mu - \int_E f(t,\cdot)\, d\mu\right) = \frac{h(s_n) - h(t)}{s_n - t}.$$

64. Par le théorème 3.16 elle est même continue, mais cela ne servira pas.

Or par l'inégalité des accroissements finis, pour μ-presque tout x et tout $n \in \mathbb{N}$,
$$|f_n(x)| \leq \sup_{s\in I}\left|\frac{\partial f}{\partial t}(s,x)\right| \leq g_1(x).$$
Comme g_1 est intégrable, par le théorème de convergence dominée 3.8 on a
$$\lim_{n\to+\infty}\frac{h(s_n)-h(t)}{s_n-t} = \lim_n \int_E f_n\, d\mu = \int_E \frac{\partial f}{\partial t}(t,x)\, d\mu(x).$$
Par caractérisation séquentielle (proposition 3.4 au Chapitre I), on en déduit que h est dérivable en $t \in I$. □

Remarque 3.24. Comme à l'exercice 3.10, ces résultats peuvent être appliquées à des séries de fonctions en utilisant l'espace mesuré $(\mathbb{N}, \mathcal{P}(\mathbb{N}), m)$ où m est la mesure de comptage. On obtient le résultat suivant (avec la version améliorée de la remarque 3.23) : soit $(u_n)_{n\in\mathbb{N}}$ une suite de fonctions dérivables sur un intervalle I de $\mathbb{R}$ telle que

(i) pour tout $t \in I$, $\sum_{n\in\mathbb{N}} |u_n(t)|$ converge ;

(ii) pour tout $t \in I$ et $n \in \mathbb{N}$, $|u_n'(t)| \leq w_n$ pour une suite $(w_n)_{n\in\mathbb{N}}$ telle que $\sum_{n\in\mathbb{N}} w_n < \infty$.

Alors $S(t) := \sum_{n\in\mathbb{N}} u_n(t)$ est bien définie pour $t \in I$, et est dérivable sur I, avec
$$\forall t \in I, \quad S'(t) = \sum_{n\in\mathbb{N}} u_n'(t).$$
Remarquons que cet énoncé est presque [65] le même que le résultat de dérivabilité des séries de fonctions dans le cadre de la convergence normale, voir l'exercice 5.22 au Chapitre I. Voir aussi le corollaire 3.46.

Exercice 3.25. Soit φ la fonction définie sur $]0,+\infty[$ par : $\varphi(t) = \int_0^{+\infty} e^{-xt}\frac{1-\cos x}{x}\,dx.$

1. Montrer que φ est continue sur $]0,+\infty[$.
2. Montrer que φ est dérivable sur $]0,+\infty[$ et calculer explicitement sa dérivée.
3. Calculer la limite de $\varphi(t)$ quand $t \to +\infty$. En déduire la valeur de $\varphi(t)$.

Exercice 3.26 (Étude de la fonction Γ et formule de Stirling)**.** On définit Γ la fonction définie sur $\mathbb{R}_+^*$ par
$$\Gamma(t) = \int_0^{+\infty} x^{t-1}e^{-x}dx.$$

1. Montrer que Γ est de classe C^∞ sur $\mathbb{R}_+^*$.[66]
2. Montrer que, pour tout $n \in \mathbb{N}^*$, $\Gamma(n+1) = n!$.
3. Montrer que, pour tout $t > 0$, $\Gamma(t+1) = \sqrt{t}\,t^t e^{-t}\int_{-\sqrt{t}}^{+\infty}\left(1+\frac{y}{\sqrt{t}}\right)^t e^{-\sqrt{t}y}dy.$
4. Montrer que, pour tout $y \geq 0$, la fonction $t \mapsto t\ln\left(1+\frac{y}{\sqrt{t}}\right) - y\sqrt{t}$ est décroissante sur $]0,+\infty[$ et que pour tout $y \in \left]-\sqrt{t}, 0\right[$, $t\ln\left(1+\frac{y}{\sqrt{t}}\right) - y\sqrt{t} \leq -\frac{y^2}{2}$.
5. Montrer que
$$\lim_{t\to+\infty}\int_{-\sqrt{t}}^0\left(1+\frac{y}{\sqrt{t}}\right)^t e^{-\sqrt{t}y}dy = \lim_{t\to+\infty}\int_0^{+\infty}\left(1+\frac{y}{\sqrt{t}}\right)^t e^{-\sqrt{t}y}dy = \int_0^{+\infty} e^{-y^2/2}dy.$$
6. En admettant que $\int_0^{+\infty} e^{-y^2/2}dy = \sqrt{\frac{\pi}{2}}$, en déduire la formule de Stirling :
$$\Gamma(t+1) \underset{+\infty}{\sim} \sqrt{2\pi t}\, t^t\, e^{-t}.$$

65. Ici on demande la convergence ponctuelle en tout point de la série, ainsi que la convergence normale de la série des dérivées. Au Chapitre I, on s'est contenté de la convergence en un seul point, et de la convergence uniforme de la série des dérivées. Mais dans la majorité des cas, on vérifie en fait les hypothèses plus forte proposées ici.

66. Notez que ceci coûte plus de montrer que la fonction Γ est C^∞ sur $]0,+\infty[$ que de montrer qu'elle est holomorphe sur $\{z \in \mathbb{C}, \mathrm{Re}(z) > 0\}$, voir l'exercice 3.68 du Chapitre II, alors que ce dernier est un meilleur résultat ; donc adaptez-vous au contexte !

4 Intégrales multiples

4.1 Tribu produit

On se donne $(E_1, \mathscr{A}_1)$ et $(E_2, \mathscr{A}_2)$ deux espaces mesurables.

Définition 4.1. On appelle *tribu produit* sur $E_1 \times E_2$, et l'on note $\mathscr{A}_1 \otimes \mathscr{A}_2$, la plus petite tribu contenant les pavés à côtés mesurables :

$$\mathscr{A}_1 \otimes \mathscr{A}_2 := \sigma\left(\mathscr{A}_1 \times \mathscr{A}_2\right),$$

où l'on a noté

$$\mathscr{A}_1 \times \mathscr{A}_2 := \{A_1 \times A_2 : A_1 \in \mathscr{A}_1, A_2 \in \mathscr{A}_2\}.$$

Le couple $(E_1 \times E_2, \mathscr{A}_1 \otimes \mathscr{A}_2)$ est appelé espace mesurable produit.

Remarque 4.2. La famille $\mathscr{A}_1 \times \mathscr{A}_2$ des pavés à côtés mesurables n'est en général pas une tribu.

Exercice 4.3 (Caractérisation des tribus produits)**.** Montrer que $\mathscr{A}_1 \otimes \mathscr{A}_2$ est la plus petite tribu sur $E_1 \times E_2$ qui rende π_1 et π_2 mesurables, où $\forall (x_1, x_2) \in E_1 \times E_2, \pi_1(x_1, x_2) = x_1,\ \pi_2(x_1, x_2) = x_2$.

Proposition 4.4. *Soit $(X, \mathscr{T})$ un espace mesurable, et*

$$\begin{aligned} f : (X, \mathscr{T}) &\longrightarrow (E_1 \times E_2, \mathscr{A}_1 \otimes \mathscr{A}_2) \\ x &\longmapsto f(x) = (f_1(x), f_2(x)). \end{aligned}$$

Alors la fonction f est mesurable ssi f_1 et f_2 sont mesurables[67].

Remarque 4.5. Ceci est une généralisation de l'énoncé 2. de la proposition 1.67.

Démonstration. $\Rightarrow$: si f est mesurable alors pour tout $i \in \{1, 2\}$, $f_i = \pi_i \circ f$ est mesurable comme composée de fonctions mesurables.

Réciproquement, supposons f_1 et f_2 mesurables. Alors pour tous $A_1 \in \mathscr{A}_1$ et $A_2 \in \mathscr{A}_2$,

$$\begin{aligned} f^{-1}(A_1 \times A_2) &= \{x \in X : f(x) \in A_1 \times A_2\} = \{x \in X : f_1(x) \in A_1, f_2(x) \in A_2\} \\ &= f_1^{-1}(A_1) \cap f_2^{-1}(A_2). \end{aligned}$$

Par hypothèse $f_1^{-1}(A_1) \in \mathscr{T}$ et $f_2^{-1}(A_2) \in \mathscr{T}$, donc $f^{-1}(A_1 \times A_2) \in \mathscr{T}$ par stabilité des tribus par intersection. Par la proposition 1.62, on en déduit que f est mesurable. □

Remarque 4.6. Ce qui précède peut sans problème se généraliser à un produit cartésien fini d'ensembles : si $(E_i, \mathscr{A}_i)_{1 \leq i \leq d}$ sont d espaces mesurables, alors on définit $\mathscr{A}_1 \otimes \cdots \otimes \mathscr{A}_d$ comme la plus petite tribu sur $E_1 \times \cdots \times E_d$ contenant tous les pavés de la forme $A_1 \times \cdots \times A_d$, où $A_i \in \mathscr{A}_i$ pour tous $i = 1, \ldots, d$; c'est aussi la plus petite tribu qui rend continues les projections canoniques.

Exercice 4.7 (Associativité de $\otimes$)**.** Montrer l'égalité suivante :

$$(\mathscr{A}_1 \otimes \cdots \otimes \mathscr{A}_j) \otimes (\mathscr{A}_{j+1} \otimes \cdots \otimes \mathscr{A}_d) = \mathscr{A}_1 \otimes \cdots \otimes \mathscr{A}_d.$$

Ainsi l'opération $\otimes$ est associative, et on peut se passer des parenthèses.

Proposition 4.8 (Cas des tribus boréliennes)**.** *Soit (E_1, d_1) et (E_2, d_2) deux espaces métriques, qu'on munit de leurs tribus boréliennes :*

— *On a toujours l'inclusion*

$$\mathscr{B}(E_1) \otimes \mathscr{B}(E_2) \subset \mathscr{B}(E_1 \times E_2).$$

— *Si $E_1 = \mathbb{R}^d$ et $E_2 = \mathbb{R}^{d'}$, alors l'inclusion précédente devient une égalité*[68].

67. comme fonctions de $(X, \mathscr{T})$ vers $(E_1, \mathscr{A}_1)$ et $(E_2, \mathscr{A}_2)$ respectivement

68. Cette égalité se généralise au cas plus général où E_1 et E_2 sont tous deux des espaces métriques séparables, et même plus généralement à des espaces topologiques à base dénombrable d'ouverts.

Démonstration. Pour le premier point, on constate que les fonctions π_1 et π_2 sont continues[69], et par conséquent sont boréliennes, d'où le résultat avec l'exercice 4.3. Le second point est une conséquence de la proposition 1.31. $\square$

Exemple 4.9. Avec l'associativité de $\otimes$, on en déduit que pour tout entier $d \geq 2$,

$$\mathscr{B}(\mathbb{R})^{\otimes d} = \mathscr{B}(\mathbb{R}^d).$$

Définition 4.10. Si $C \in \mathscr{A}_1 \otimes \mathscr{A}_2$, pour tous $x_1 \in E_1$ et $x_2 \in E_2$, on note

$$C_{x_1} := \{y_2 \in E_2 : (x_1, y_2) \in C\} \quad \text{et} \quad C^{x_2} := \{y_1 \in E_1 : (y_1, x_2) \in C\},$$

que l'on appelle sections de C.

Proposition 4.11.
— *Soit $f : (E_1 \times E_2, \mathscr{A}_1 \otimes \mathscr{A}_2) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mesurable. Alors pour tout $x_1 \in E_1$, l'application partielle*

$$\begin{aligned} f_{x_1} : (E_2, \mathscr{A}_2) &\longrightarrow (\mathbb{R}, \mathscr{B}(\mathbb{R})) \\ x_2 &\longmapsto f(x_1, x_2) \end{aligned}$$

est mesurable[70].

— *Les sections d'éléments de la tribu produit sont mesurables. Autrement dit, pour tout $C \in \mathscr{A}_1 \otimes \mathscr{A}_2$ et pour tous $x_1 \in E_1$ et $x_2 \in E_2$: $C_{x_1} \in \mathscr{A}_2$ et $C^{x_2} \in \mathscr{A}_1$.*

Démonstration. Soit $x_1 \in E_1$. L'application

$$\begin{aligned} g_{x_1} : (E_2, \mathscr{A}_2) &\longrightarrow (E_1 \times E_2, \mathscr{A}_1 \otimes \mathscr{A}_2) \\ x_2 &\longmapsto (x_1, x_2) \end{aligned}$$

est mesurable d'après la proposition 4.4, en effet chacune des applications coordonnées est mesurable de façon évidente. Donc $f_{x_1} = f \circ g_{x_1}$ est mesurable. La seconde partie de l'énoncé se déduit en appliquant la première à la fonction $f = \mathbb{1}_C$, car $f_{x_1} = \mathbb{1}_{C_{x_1}}$, puis en échangeant les rôles de x_1 et x_2. $\square$

4.2 Mesure produit

Théorème 4.12. *Soient μ_1 et μ_2 deux mesures σ-finies, sur $(E_1, \mathscr{A}_1)$ et $(E_2, \mathscr{A}_2)$ respectivement. Alors il existe une unique mesure m sur l'espace produit $(E_1 \times E_2, \mathscr{A}_1 \otimes \mathscr{A}_2)$ vérifiant*

$$\forall A_1 \in \mathscr{A}_1, \ \ \forall A_2 \in \mathscr{A}_2, \qquad m(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2).$$

Cette mesure est σ-finie, est appelée mesure produit, et notée $m = \mu_1 \otimes \mu_2$. De plus, pour tout $C \in \mathscr{A}_1 \otimes \mathscr{A}_2$, les applications

$$\begin{array}{ccc} (E_1, \mathscr{A}_1) & \longrightarrow & (\overline{\mathbb{R}}_+, \mathscr{B}(\overline{\mathbb{R}}_+)) \\ x_1 & \longmapsto & \mu_2(C_{x_1}) \end{array} \quad \textit{et} \quad \begin{array}{ccc} (E_2, \mathscr{A}_2) & \longrightarrow & (\overline{\mathbb{R}}_+, \mathscr{B}(\overline{\mathbb{R}}_+)) \\ x_2 & \longmapsto & \mu_1(C^{x_2}) \end{array}$$

sont mesurables et

$$\int_{E_1} \mu_2(C_{x_1})\, d\mu_1(x_1) = \mu_1 \otimes \mu_2(C) = \int_{E_2} \mu_1(C^{x_2})\, d\mu_2(x_2).$$

Comme précisé dans le programme officiel, ce théorème est admis.

Remarque 4.13. Le théorème 4.12 est faux lorsque μ_1 ou μ_2 n'est pas σ-finie. Soit en effet par exemple la mesure de Lebesgue sur $(E_1, \mathscr{A}_1) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ pour μ_1 (qui est bien σ-finie), et la mesure de comptage sur $(E_2, \mathscr{A}_2) = (\mathbb{R}, \mathcal{P}(\mathbb{R}))$ pour μ_2 (qui n'est pas σ-finie). En prenant par exemple $C = \{(x, x) : x \in \mathbb{R}\}$ la première bissectrice de $\mathbb{R}^2$, alors $C_{x_1} = \{x_1\}$ et $C^{x_2} = \{x_2\}$, donc $\mu_1(C^{x_2}) = 0$, tandis que $\mu_2(C_{x_1}) = 1$. Par conséquent,

$$\int_{E_1} \mu_2(C_{x_1})\, d\mu_1(x_1) = \mu_1(E_1) = +\infty, \qquad \text{mais} \qquad \int_{E_2} \mu_1(C_{x_2})\, d\mu_2(x_2) = 0.$$

Corollaire 4.14. *La mesure de Lebesgue λ_d sur $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ est aussi la mesure produit $\lambda_1^{\otimes d}$.*

69. Par exemple : si O_1 est un ouvert de E_1, $\pi_1^{-1}(O_1) = O_1 \times E_2$ est un ouvert de $E_1 \times E_2$.

70. Attention, la réciproque est fausse : le fait que toutes les applications partielles soient mesurables n'implique pas forcément que f soit mesurable. Un contre-exemple n'est pas pour autant facile à produire : on renvoie par exemple à [Fej, Exercice 13.2].

4.3 Théorèmes de Fubini

Théorème 4.15 (de Fubini-Tonelli)**.** *Soient μ_1 et μ_2 deux mesures σ-finies sur $(E_1, \mathscr{A}_1)$ et $(E_2, \mathscr{A}_2)$ respectivement et $f : (E_1 \times E_2, \mathscr{A}_1 \otimes \mathscr{A}_2) \to (\overline{\mathbb{R}}_+, \mathscr{B}(\overline{\mathbb{R}}_+))$ mesurable. Alors les fonctions ϕ et ψ définies resp. sur E_1 et E_2 par*

$$\phi(x_1) := \int_{E_2} f(x_1, x_2)\, d\mu_2(x_2) \quad \text{et} \quad \psi(x_2) := \int_{E_1} f(x_1, x_2)\, d\mu_1(x_1) \tag{4.1}$$

sont mesurables, et on a l'égalité :

$$\int_{E_1} \phi\, d\mu_1 = \int_{E_1 \times E_2} f\, d(\mu_1 \otimes \mu_2) = \int_{E_2} \psi\, d\mu_2. \tag{4.2}$$

Démonstration. Si $f = \mathbb{1}_C$ pour $C \in \mathscr{A}_1 \otimes \mathscr{A}_2$, alors $\phi(x_1) = \mu_2(C_{x_1})$ et $\psi(x_2) = \mu_1(C_{x_2})$, et donc le résultat découle du théorème 4.12. Cette assertion s'étend aux fonctions étagées positives pas linéarité de l'intégrale, puis aux fonctions mesurables positives par le lemme fondamental d'approximation et le théorème de Beppo Levi. □

Remarque 4.16. Si $(a_{n,m})_{(n,m)\in\mathbb{N}^2}$ est une suite doublement indicée de nombres réels positifs et si μ est la mesure de comptage sur $(\mathbb{N}, \mathscr{P}(\mathbb{N}))$, alors l'interversion suivante

$$\sum_{n\in\mathbb{N}} \sum_{m\in\mathbb{N}} a_{n,m} = \sum_{m\in\mathbb{N}} \sum_{n\in\mathbb{N}} a_{n,m}$$

peut être vue comme une application du théorème de Beppo Levi, car $\sum_{m\in\mathbb{N}} a_{n,m} = \int_{\mathbb{N}} a_{n,m}\, d\mu(m)$, comme du théorème de Fubini-Tonelli, car les termes de l'équation sont tous deux égaux à $\int_{\mathbb{N}\times\mathbb{N}} a_{n,m} d\mu^{\otimes 2}(n, m)$.

Théorème 4.17 (de Fubini)**.** *Soient μ_1 et μ_2 deux mesures σ-finies sur $(E_1, \mathscr{A}_1)$ et $(E_2, \mathscr{A}_2)$ respectivement et $f : (E_1 \times E_2, \mathscr{A}_1 \otimes \mathscr{A}_2) \to (\overline{\mathbb{R}}, \mathscr{B}(\overline{\mathbb{R}}))$ qu'on suppose $\mu_1 \otimes \mu_2$-intégrable, c'est-à-dire*

$$\int_{E_1 \times E_2} |f|\, d(\mu_1 \otimes \mu_2) < +\infty.$$

Alors les fonctions ϕ et ψ définies par (4.1) sont respectivement définies μ_1-p.p. et μ_2-p.p., μ_1-intégrables et μ_2-intégrables, et vérifient (4.2).

Démonstration. On définit

$$\phi_+(x_1) = \int_{E_2} f^+(x_1, x_2)\, d\mu_2(x_2),$$

et de manière similaire ϕ_-, ψ_+ et ψ_-. D'après le théorème de Fubini-Tonelli,

$$\int_{E_1} \phi_+\, d\mu_1 = \int_{E_1 \times E_2} f^+\, d(\mu_1 \otimes \mu_2) = \int_{E_2} \psi_+\, d\mu_2,$$

qui est un nombre réel fini par hypothèse d'intégrabilité de f. Par conséquent, ϕ_+ est finie μ_1-p.p. et ψ_+ est finie μ_2-p.p., ainsi que ϕ_- et ψ_- respectivement. Donc la fonction ϕ est définie μ_1-p.p. (comme différence de deux fonctions finies p.p.) et l'intégrale de $|\phi|$ est finie car égale à la somme des intégrales de ϕ_+ et de ϕ_-, qui sont toutes deux finies. Le résultat analogue se démontre de la même manière pour ψ, et ainsi l'égalité (4.2) s'obtient en faisant la différence de deux quantités finies. □

Remarque 4.18. Si f est positive, le théorème de Fubini-Tonelli assure que l'intégrale de f par rapport à $\mu_1 \otimes \mu_2$ peut toujours se calculer en faisant deux intégrales « simples » successives dans l'ordre que l'on souhaite. Si f est de signe quelconque, il faut d'abord vérifier l'intégrabilité de $|f|$ par rapport à $\mu_1 \otimes \mu_2$, et pour ce faire on utilise en général le théorème de Fubini-Tonelli.

Remarque 4.19. Quand l'un des théorèmes de Fubini s'applique, on notera donc

$$\int_{E_1} \int_{E_2} f(x_1, x_2)\, d\mu_1(x_1) d\mu_2(x_2) = \int_{E_1 \times E_2} f\, d(\mu_1 \otimes \mu_2).$$

4.4 Changement de variable

4.4.1 Mesure image

Soient $(E_1, \mathscr{A}_1)$ et $(E_2, \mathscr{A}_2)$ deux espaces mesurables, μ une mesure sur $(E_1, \mathscr{A}_1)$ et $h : (E_1, \mathscr{A}_1) \to (E_2, \mathscr{A}_2)$ mesurable.

Définition-Proposition 4.20. L'égalité

$$\forall A_2 \in \mathscr{A}_2, \quad \nu(A_2) := \mu\left(h^{-1}(A_2)\right)$$

définit une mesure ν sur $(E_2, \mathscr{A}_2)$ appelée mesure image de μ par h, notée $\mu \circ h^{-1}$, $h\sharp\mu$, $h(\mu)$, ou encore μ_h.

Démonstration. i) $\nu(\varnothing) = \mu(h^{-1}(\varnothing)) = \mu(\varnothing) = 0$.
ii) Soit $(B_n)_{n\in\mathbb{N}}$ une famille dénombrable d'éléments de $\mathscr{A}_2$ deux à deux disjoints ; alors

$$\forall i \neq j, \quad h^{-1}(B_i) \cap h^{-1}(B_j) = h^{-1}(B_i \cap B_j) = h^{-1}(\varnothing) = \varnothing,$$

et ainsi

$$\nu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \mu\left(h^{-1}\left(\bigcup_{n\in\mathbb{N}} B_n\right)\right) = \mu\left(\bigcup_{n\in\mathbb{N}} h^{-1}(B_n)\right) = \sum_{n\in\mathbb{N}} \mu\left(h^{-1}(B_n)\right) = \sum_{n\in\mathbb{N}} \nu(B_n),$$

par σ-additivité de μ. □

Théorème 4.21 (de transfert)**.** *Soit $f : (E_2, \mathscr{A}_2) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ mesurable.*

— *Si f est positive, alors on a l'égalité (dans $\overline{\mathbb{R}}_+$) :*

$$\int_{E_2} f \, d\mu_h = \int_{E_1} f \circ h \, d\mu. \tag{4.3}$$

— *f est μ_h-intégrable ssi $f \circ h$ est μ-intégrable, et on a alors l'égalité (4.3) dans $\mathbb{R}$.*

Démonstration. Si $f = \mathbb{1}_B$, où $B \in \mathscr{A}_2$, alors

$$\int_{E_1} f \circ h \, d\mu = \int_{E_1} \mathbb{1}_B \circ h \, d\mu = \int_{E_1} \mathbb{1}_{h^{-1}(B)} \, d\mu = \mu\left(h^{-1}(B)\right) = \mu_h(B) = \int_{E_2} f \, d\mu_h.$$

La linéarité de l'intégrale, le lemme fondamental d'approximation et le théorème de Beppo Levi impliquent (4.3) dès que f est positive. L'extension aux fonctions mesurables de signe quelconque et l'équivalence des intégrabilités se déduisent de la décomposition $f = f^+ - f^-$. □

Exemple 4.22. Soit $h : (E, \mathscr{A}) \to (E, \mathscr{A})$ mesurable et telle que $\mu_h = \mu$. Alors

$$\int_E f \, d\mu = \int_E f \circ h \, d\mu$$

pour toute fonction positive μ-p.p. ou μ-intégrable. En particulier, si μ est la mesure de Lebesgue sur $\mathbb{R}^d$ et $h = \tau_a$ est la translation de vecteur $a \in \mathbb{R}^d$, alors $\mu_h = \mu$ et donc

$$\int_{\mathbb{R}^d} f(x+a) \, d\lambda(x) = \int_{\mathbb{R}^d} f \, d\lambda.$$

Exercice 4.23 (*Changement de variable linéaire)**.** On note λ la mesure de Lebesgue sur $\mathbb{R}^d$ avec $d \geq 1$. Soit $A \in GL_d(\mathbb{R})$ et $b \in \mathbb{R}^d$ et h l'application affine définie par $h(x) = Ax + b$. On souhaite montrer [71]

$$\lambda_h = |\det A|^{-1} \lambda.$$

1. Montrer qu'on peut se ramener au cas $b = 0$.

71. Ceci est un cas particulier du théorème 4.24 ; on l'utilise en fait dans la preuve de ce dernier.

2. On pose $\nu := \lambda \circ A^{-1}$.
   (a) Montrer que ν est invariante par translation.
   (b) On pose C_d le cube $[0,1]^d$. Montrer que $\nu(C_d) \in]0, +\infty[$, et en déduire que
   $$\nu = c(A)\lambda, \qquad \text{où} \quad c(A) = \lambda \circ A^{-1}\left([0,1]^d\right).$$
   (c) Montrer que c est un morphisme de groupe sur $GL_d(\mathbb{R})$.
   (d) Montrer que[72] $c(A) = |\det A|^{-1}$ et conclure.
3. En déduire que pour toute fonction f positive mesurable, ou λ-intégrable, on a
$$\int_{\mathbb{R}^d} f \circ h \, d\lambda = \frac{1}{|\det A|} \int_{\mathbb{R}^d} f \, d\lambda.$$

4.4.2 Formule du changement de variable

Soient U et V deux ouverts de $\mathbb{R}^d$. On dit que ϕ est un C^1-difféomorphisme entre U et V si ϕ est une bijection de U sur V telle que ϕ est de classe C^1 sur U et ϕ^{-1} est de classe C^1 sur V. Pour tout $u \in U$, on note $\phi'(u)$ la matrice carrée $d \times d$ des dérivées partielles de ϕ évaluées en u, autrement dit la matrice représentative de la différentielle de ϕ en u, appelée matrice jacobienne de ϕ en u, et $J_\phi(u)$ le déterminant de $\phi'(u)$, appelé jacobien de ϕ en u :
$$\phi'(u) = \left(\frac{\partial \phi_i}{\partial x_j}\right)_{1 \le i,j \le d}, \qquad J_\phi(u) = \det(\phi'(u)).$$

Théorème 4.24 (Formule de changement de variable)**.** *Soit f une fonction borélienne sur V. Si f est positive ou λ-intégrable, alors*
$$\int_V f \, d\lambda = \int_U f \circ \phi \, |J_\phi| \, d\lambda.$$
De manière équivalente, si f est positive ou si $f \circ \phi$ est λ-intégrable, alors
$$\int_U f \circ \phi \, d\lambda = \int_V f \, |J_{\phi^{-1}}| \, d\lambda.$$

La démonstration de la formule du changement de variable est admise.

Remarque 4.25. En dimension 1, on invite le lecteur à comparer avec la proposition 5.18 au Chapitre I et à lire la remarque 5.19. Si $\phi :]\alpha, \beta[\to [a, b]$ est un C^1-difféomorphisme, alors ϕ' ne s'annule pas et est de signe constant : si $\phi' > 0$, alors
$$\int_a^b f(x) \, dx = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f \circ \phi(u) \, \phi'(u) \, du = \int_\alpha^\beta f \circ \phi(u) \, |\phi'(u)| \, du,$$
et si $\phi' < 0$, alors
$$\int_a^b f(x) \, dx = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f \circ \phi(u) \, \phi'(u) \, du = -\int_\alpha^\beta f \circ \phi(u) \, \phi'(u) \, du = \int_\alpha^\beta f \circ \phi(u) \, |\phi'(u)| \, du.$$

Remarque 4.26. Pour vérifier que ϕ est un C^1-difféomorphisme, on pourra appliquer le théorème d'inversion globale qui sera montré au Chapitre VII.

72. On pourra utiliser que toute matrice inversible peut se décomposer comme produit de matrice de la forme (en notant $(e_i)_{i \in [\![1,d]\!]}$ la base canonique de $\mathbb{R}^d$) :
$$\forall i \in [\![1,d]\!], \quad B_1.e_i = e_{\sigma(i)} \qquad \text{pour } \sigma \text{ une permutation de } \{1, \dots, d\},$$
$$B_2.e_1 = \alpha e_1 \quad \text{et } \forall j \in [\![2,d]\!], \quad B_2.e_j = e_j \text{ (avec } \alpha \neq 0),$$
$$B_3.e_1 = e_1 + e_2, \quad \text{et } \forall j \in [\![2,d]\!], \quad B_3.e_j = e_j.$$

Exercice 4.27. Soit μ la mesure de densité f par rapport à λ et $\phi : \mathbb{R}^d \to \mathbb{R}^d$ un C^1-difféomorphisme. Montrer que la mesure image de μ par ϕ admet une densité g par rapport à λ, et que g est donnée par

$$g(x) = f \circ \phi^{-1}(x)\, |J_{\phi^{-1}}(x)| \qquad x \in \mathbb{R}^d.$$

Exemple 4.28 (coordonnées polaires)**.** Par le théorème d'inversion globale, la fonction

$$\begin{aligned} \phi :]0,\infty[\times]0,2\pi[&\longrightarrow \mathbb{R}^2 \setminus ([0,\infty[\times\{0\}) \\ (\rho,\theta) &\longmapsto (\rho\cos\theta, \rho\sin\theta) \end{aligned}$$

est un C^1-difféomorphisme, avec

$$\phi'(\rho,\theta) = \begin{pmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{pmatrix}, \qquad \text{et donc } J_\phi(\rho,\theta) = \rho$$

. Ainsi pour toute fonction borélienne f λ_2-intégrable,

$$\begin{aligned} \int_{\mathbb{R}^2} f(x,y)\,dx\,dy &= \int_{\mathbb{R}^2\setminus([0,\infty[\times\{0\})} f(x,y)\,dx\,dy \\ &= \int_{]0,\infty[\times]0,2\pi[} f\circ\phi(\rho,\theta)\, J_\phi(\rho,\theta)\,d\rho\,d\theta = \int_{[0,\infty[\times[0,2\pi]} f\circ\phi(\rho,\theta)\,\rho\,d\rho\,d\theta. \end{aligned}$$

Exercice 4.29 (Coordonnées sphériques)**.**

1. Montrer que l'application

$$\begin{aligned} \Phi : \quad]0,+\infty[\times]-\pi,\pi[\times]-\tfrac{\pi}{2},\tfrac{\pi}{2}[&\longrightarrow \mathbb{R}^3 \setminus (\mathbb{R}_- \times \{0\} \times \mathbb{R}) \\ (r,\theta,\varphi) &\longmapsto \begin{pmatrix} r\cos\theta\cos\varphi \\ r\sin\theta\cos\varphi \\ r\sin\varphi \end{pmatrix} \end{aligned}$$

   est un C^1-difféomorphisme.
2. Montrer que

$$J_\Phi(r,\theta,\varphi) = r^2\cos\varphi.$$

3. En déduire la valeur du volume d'une boule de rayon R dans $\mathbb{R}^3$.

Exercice 4.30 (Calcul de l'intégrale de Gauss (hyper classique))**.** Posons $I = \int_{\mathbb{R}} e^{-x^2}\,dx$.

1. Montrer avec le théorème de Fubini que

$$I^2 = \int_{\mathbb{R}^2} e^{-(x^2+y^2)}dx\,dy.$$

2. Par un changement de variable en coordonnées polaires, montrer

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)}dx\,dy = \pi,$$

   et en déduire la valeur de I.
3. En déduire la valeur de $I_\alpha = \int_{\mathbb{R}} e^{-\alpha x^2}\,dx$ pour $\alpha \in]0,+\infty[$.

Exercice 4.31 (Calcul du volume de la boule unité de $\mathbb{R}^n$.)**.** On note $B_n(r)$ la boule (centrée sur l'origine) de rayon r dans $\mathbb{R}^n$ muni de la norme euclidienne.

1. Montrer

$$\lambda_n(B_n(r)) = r^n\,\lambda_n(B_n(1)).$$

2. On pose $c_n := \lambda_n(B_n(1))$ pour $n \geq 1$. Montrer

$$\forall n \in \mathbb{N}, \quad c_{n+1} = c_n I_n, \qquad \text{où} \quad I_n := \int_{-1}^{1} \left(1-x^2\right)^{n/2}\,dx.$$

3. Montrer que pour $n \geq 2$, $I_n = nI_{n-2}/(n+1)$.
4. Conclure que

$$c_{2k} = \frac{\pi^k}{k!}, \qquad c_{2k+1} = \frac{2^{k+1}\pi^k}{1\cdot 3\cdot 5\cdots(2k+1)}.$$

   On retrouve ainsi $c_1 = 2$, $c_2 = \pi$ et $c_3 = 4\pi/3$.

5 Espaces L^p

Nous présentons les résultats dans un espace mesuré $(E, \mathcal{A}, \mu)$ où E est un ensemble, $\mathcal{A}$ est une tribu sur E et μ une mesure positive sur $\mathcal{A}$. En pratique, on utilise très souvent le cas de $(I, \mathscr{B}(I), \lambda)$ où I est un intervalle de $\mathbb{R}$, et on conseille le lecteur de commencer par se familiariser avec ce cas. Le cas de l'espace mesuré $(\mathbb{N}, \mathcal{P}(\mathbb{N}), m)$ où m est la mesure de comptage est aussi intéressant : certains aspects purement "analyse fonctionnelle" possèdent déjà la richesse attendue des "espaces L^p"[73], mais certains aspects sont simplifiés[74]. Le cas général peut s'avérer utile, notamment dans le cadre des probabilités, et aussi les espaces à poids[75] ; si à l'oral vous décidez de ne pas aborder ces applications, il peut être judicieux de se restreindre au cas $(I, \mathscr{B}(I), \lambda)$.

5.1 Définitions et premières propriétés

Définition 5.1 (Espaces $\mathcal{L}^p$)**.** Soit $1 \leq p \leq +\infty$[76]. On définit

$$\mathcal{L}^p_{\mathbb{K}}(E, \mathscr{A}, \mu) = \mathcal{L}^p(E) = \left\{ f\colon E \to \mathbb{K} \text{ mesurable}^{77} \text{ et telle que } \|f\|_{\mathcal{L}^p(E)} < +\infty \right\},$$

où on a posé

$$\|f\|_{\mathcal{L}^p(E)} = \begin{cases} \left(\displaystyle\int_E |f|^p \, d\mu(x) \right)^{1/p} & \text{si } 1 \leq p < +\infty, \\ \operatorname*{ess\,sup}_{x \in E} |f(x)| := \inf \left\{ C \geq 0,\ \lambda(\{|f| > C\}) = 0 \right\} & \text{si } p = +\infty. \end{cases}$$

Remarque 5.2. Attention donc à toujours distinguer dans les calculs le cas $p < +\infty$ et le cas $p = +\infty$, car seulement le premier cas vient d'une intégrale.

Remarque 5.3. Attention, on utilise la notation $\|\cdot\|_{\mathcal{L}^p(E)}$, mais ces objets ne sont pas des normes en général. On verra comment résoudre ce problème un peu plus loin.

Le nombre $\operatorname{ess\,sup}_{x \in E} g(x) := \inf\{C \in \mathbb{R},\ \ \lambda(\{g > C\}) = 0\}$ s'appelle le sup-essentiel de g (appliqué ici à $|f|$). Afin de justifier la notation $\|\cdot\|_{\mathcal{L}^\infty(E)}$ qui rappelle la norme uniforme $\|\cdot\|_\infty$, on donne le lemme suivant :

Lemme 5.4. — *Si f est mesurable*[78] *alors pour presque tout $x \in E$,*[79]

$$|f(x)| \leq \|f\|_{\mathcal{L}^\infty(E)}.$$

— *Si $(E, \mathscr{A}, \mu) = (I, \mathscr{B}(I), \lambda)$ où I est un intervalle de $\mathbb{R}$, et si $f \in C^0(I)$ alors*

$$\operatorname*{ess\,sup}_{x \in I} |f(x)| = \sup_{x \in I} |f(x)|.$$

La deuxième partie de ce lemme permet d'expliquer que si une fonction f est continue, on peut manipuler sup et sup-essentiel indifféremment (et donc quand on est sur un intervalle compact, utiliser sans distinction les notations $\|f\|_\infty$ ou $\|f\|_{\mathcal{L}^\infty(E)}$).

73. Notamment ça fournit des exemples d'espaces de Banach (voire de Hilbert dans le cas $p = 2$).

74. Pas besoin de passer au quotient par l'égalité presque partout ; aussi, il est facile de voir que l'espace des suites à support fini est dense. Il n'y a pas de notion d'espace des fonctions continues sur $\mathbb{N}$, donc ces aspects disparaissent. Enfin, on peut se poser la question de la dualité dans ce cas particulier, qui est beaucoup plus simple que le cas général.

75. C'est-à-dire quand la mesure μ est de la forme $f(x)d\lambda$ où f est une fonction intégrable.

76. Attention, on suppose bien $p \geq 1$. Rien n'interdit de définir $\|\cdot\|_p$, $\mathcal{L}^p(E)$ et $L^p(E)$ si $p \in]0,1[$; mais alors $\|\cdot\|_p$ ne satisfait pas l'inégalité triangulaire (qui s'appelle inégalité de Minkowski si $p \geq 1$, voir la proposition 5.11 ; la raison est qu'on a utilisé la convexité de la fonction $x \in \mathbb{R}_+ \mapsto x^p$ dans la preuve de l'inégalité de Hölder, elle-même utilisée dans la preuve de l'inégalité de Minkowski). On peut quand même montrer que $\mathcal{L}^p(E)$ et $L^p(E)$ sont des espaces vectoriels. On peut les munir de $(f, g) \mapsto \|f - g\|_p^p$ qui est une métrique ; mais cette métrique n'est pas issue d'une norme.

78. On ne suppose pas $f \in \mathcal{L}^\infty(E)$, car sinon $\|f\|_{\mathcal{L}^\infty} = +\infty$ et l'inégalité est triviale mais vraie.

79. On invite le lecteur à se convaincre que l'inégalité donnée n'est pas triviale au vu de la définition : en effet on montre en fait que l'infimum dans la définition de $\|\cdot\|_{\mathcal{L}^\infty}$ est un plus petit élément, et non seulement une borne inférieure.

Démonstration. Par définition de borne inférieure, il existe une suite $(\alpha_n)_{n\in\mathbb{N}}$ qui converge vers $\operatorname{ess\,sup}_E |f|$ et telle que pour tout $n \in \mathbb{N}$, $\Big[|f| \leq \alpha_n \text{ presque partout}\Big]$. En d'autres termes, pour tout $n \in \mathbb{N}$, il existe $N_n \in \mathscr{A}$ tel que $\mu(N_n) = 0$ et pour tout $x \in E \setminus N_n$, $|f(x)| \leq \alpha_n$. Par conséquent, posant $N = \bigcup_{n\in\mathbb{N}} N_n \in \mathscr{A}$, on voit que pour tout $x \in E \setminus N$, $|f(x)| \leq \alpha_n$ pour tout $n \in \mathbb{N}$. On peut donc passer à la limite et obtenir que $|f(x)| \leq \operatorname{ess\,sup}_E |f|$ pour tout $x \in E \setminus N$. Pour conclure, on note que N est une réunion dénombrable d'ensembles de mesure nulle, donc $\mu(N) = 0$.

Soit maintenant $f \in C^0(I)$. Comme f est continue, l'ensemble $\{x \in I : |f(x)| > \operatorname{ess\,sup}_I |f|\}$ est un ouvert. Par le point précédent, il est de mesure de Lebesgue nulle. Il est donc vide [80], c'est-à-dire que $|f(x)| \leq \operatorname{ess\,sup}_I |f|$ partout, ce qui implique que $\sup_I |f| \leq \operatorname{ess\,sup}_I |f|$. L'inégalité dans l'autre sens est évidente par la définition du sup essentiel. □

Si la fonction f n'est pas continue, il faut être un peu plus vigilant, comme le montre l'exercice suivant :

Exercice 5.5. Calculer le sup et le ess sup des fonctions suivantes (définies sur $\mathbb{R}$) :

$$\mathbb{1}_{\mathbb{Q}}, \qquad x \mapsto x\mathbb{1}_{\mathbb{Q}}(x), \qquad \mathbb{1}_{\mathbb{Q}} + 2\mathbb{1}_{\mathbb{R}\setminus\mathbb{Q}}, \qquad x \mapsto \mathbb{1}_{\{0\}}(x) + \frac{1}{1+x^2}.$$

Exercice 5.6 (Justification de la notation ∞ dans $\|\cdot\|_\infty$)**.** Montrer que si $f \in L^{p_0}(E)$ pour un certain $p_0 \in [1, +\infty]$, alors

$$\lim_{p\to+\infty} \|f\|_p = \|f\|_\infty.$$

On établit maintenant deux inégalités fondamentales en théorie de l'intégration.

Proposition 5.7 (Inégalité de Hölder). *Soient $1 \leq p \leq +\infty$ et $p' \geq 1$ son exposant conjugué, défini comme l'unique élément de $[1, +\infty]$ tel que $1/p + 1/p' = 1$ [81]. Alors pour toutes fonctions $f, g : E \to \mathbb{K}$ mesurables,*

$$\|fg\|_{\mathcal{L}^1(E)} \leq \|f\|_{\mathcal{L}^p(E)} \|g\|_{\mathcal{L}^{p'}(E)}. \tag{5.1}$$

En particulier, si $f \in \mathcal{L}^p(E)$ et $g \in \mathcal{L}^{p'}(E)$, alors $fg \in \mathcal{L}^1(E)$.

Remarque 5.8. On a volontairement donné la formule (5.1) sans hypothèse d'intégrabilité ; cela est permis car les éléments en jeu peuvent valoir $+\infty$. Donc si f n'est pas dans $\mathcal{L}^p$, on a $\|f\|_{\mathcal{L}^p} = +\infty$ et l'inégalité n'est pas intéressante [82], mais correcte.

Remarque 5.9. Quand $p = 2$, cette inégalité est l'inégalité de Cauchy-Schwarz qu'on reverra dans le cadre des espaces préhilbertien.

Démonstration. Dans les cas $p = 1$ ou $p = +\infty$, le résultat est immédiat avec le lemme 5.4, première partie. Supposons donc maintenant $1 < p < +\infty$. Par concavité du logarithme sur $]0, +\infty[$, pour $a, b > 0$, on a

$$\log(ab) = \log(a) + \log(b) = \frac{1}{p}\log(a^p) + \frac{1}{p'}\log(a^{p'}) \leq \log\left(\frac{1}{p}a^p + \frac{1}{p'}b^{p'}\right).$$

En prenant l'exponentielle, on obtient l'inégalité suivante [83] :

$$ab \leq \frac{1}{p}a^p + \frac{1}{p'}b^{p'}, \tag{5.2}$$

qui reste valable si a ou b est nul.

80. Rappelons ce lemme classique et facile à redémontrer sur le lien topologie/mesure : pour la mesure de Lebesgue, un ouvert non vide est de mesure non nulle.

81. On retiendra $(+\infty)' = 1$, $2' = 2$, et $1' = +\infty$. On peut aussi apprendre la formule $p' = \frac{p}{p-1}$, mais celle-ci peut se retrouver facilement.

82. Sauf éventuellement si $\|g\|_{\mathcal{L}^{p'}} = 0$, et alors on utilise la convention $+\infty \times 0 = 0$, et l'inégalité (5.1) donne $\|fg\|_{\mathcal{L}^1} = 0$, ce qui est effectivement intéressant, mais pas difficile à anticiper car g doit être nulle μ-presque partout pour être telle que $\|g\|_{\mathcal{L}^p} = 0$.

83. Parfois appelée inégalité de Young.

Remarquons aussi que les cas où $\|f\|_{\mathcal{L}^p(E)}$ ou $\|g\|_{\mathcal{L}^{p'}(E)}$ est nul sont triviaux car alors f ou g est nulle presque partout (par la proposition 2.15) et l'inégalité est alors claire, voir aussi la note 82. On suppose donc $\|f\|_{\mathcal{L}^p(E)} > 0$ et $\|g\|_{\mathcal{L}^{p'}(E)} > 0$: on applique l'inégalité (5.2) avec $a = |f(x)|/\|f\|_{\mathcal{L}^p(E)}$ et $b = |g(x)|/\|g\|_{\mathcal{L}^{p'}(E)}$ et en intégrant sur E, on obtient l'inégalité voulue. □

Exercice 5.10 (Inégalité de Hölder généralisée)**.** Soit $(E, \mathscr{A}, \mu)$ un espace mesuré.

1. Soient $(p, q, r) \in \mathbb{R}^3$ tels que $p \geq 1$, $q \geq 1$ et $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. Soient f et g deux fonctions mesurables de $(E, \mathscr{A})$ dans $(\mathbb{K}, \mathscr{B}(\mathbb{K}))$. Montrer que
$$\|fg\|_{\mathcal{L}^r(E)} \leq \|f\|_{\mathcal{L}^p(E)} \|g\|_{\mathcal{L}^q(E)}.$$
2. Soit f mesurable. Montrer que l'ensemble des $p \in [1, +\infty]$ tels que $f \in \mathcal{L}^p(E)$ est un intervalle.

Proposition 5.11 (Inégalité de Minkowski). *Pour $1 \leq p \leq +\infty$ et pour tout f, $g \in \mathcal{L}^p(E)$, on a*
$$\|f + g\|_{\mathcal{L}^p(E)} \leq \|f\|_{\mathcal{L}^p(E)} + \|g\|_{\mathcal{L}^p(E)}.$$

Démonstration. On commence par le cas $p = +\infty$. Par la première partie du lemme 5.4, $|f(x)| \leq \|f\|_{\mathcal{L}^\infty(E)}$ presque partout et $|g(x)| \leq \|g\|_{\mathcal{L}^\infty(E)}$ presque partout, d'où [84]
$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_{\mathcal{L}^\infty(E)} + \|g\|_{\mathcal{L}^\infty(E)},$$
pour presque tout $x \in E$, et donc $\|f + g\|_{\mathcal{L}^\infty(E)} \leq \|f\|_{\mathcal{L}^\infty(E)} + \|g\|_{\mathcal{L}^\infty(E)}$.
Si $p = 1$, on a
$$\|f + g\|_{\mathcal{L}^1(E)} = \int_E |f + g| \, d\mu(x) \leq \int_E (|f| + |g|) \, d\mu(x) \leq \|f\|_{\mathcal{L}^1(E)} + \|g\|_{\mathcal{L}^1(E)}.$$

Enfin, si $1 < p < +\infty$, si $\|f+g\|_{\mathcal{L}^p(E)} = 0$, il n'y a rien à montrer. Sinon, $\|f+g\|_{\mathcal{L}^p(E)} > 0$ et l'inégalité de Hölder implique que
$$\begin{aligned}
\|f + g\|^p_{\mathcal{L}^p(E)} &= \int_E |f + g|^p \, d\mu(x) = \int_E |f + g||f + g|^{p-1} \, d\mu(x) \\
&\leq \int_E |f||f + g|^{p-1} \, d\mu(x) + \int_E |g||f + g|^{p-1} \, d\mu(x) \\
&\leq (\|f\|_{\mathcal{L}^p(E)} + \|g\|_{\mathcal{L}^p(E)}) \left(\int_E (|f + g|^{p-1})^{p'} \right)^{\frac{1}{p'}} \\
&= (\|f\|_{\mathcal{L}^p(E)} + \|g\|_{\mathcal{L}^p(E)}) \|f + g\|^{p-1}_{\mathcal{L}^p(E)},
\end{aligned}$$
en effet, l'égalité $\frac{1}{p} + \frac{1}{p'} = 1$ se réécrit $p'(p-1) = p$, ce qui conclut la preuve de ce résultat en divisant par $\|f + g\|^{p-1}_{\mathcal{L}^p(E)}$. □

L'inégalité de Minkowski montre que l'application $\mathcal{L}^p(E) \to \mathbb{R}_+$, $f \mapsto \|f\|_{\mathcal{L}^p(E)}$, vérifie l'inégalité triangulaire. Par ailleurs, $\|\alpha f\|_p = |\alpha| \|f\|_p$ pour tout $\alpha \in \mathbb{K}$ et $f \in \mathcal{L}^p(E)$ et bien sûr $\|0\|_{\mathcal{L}^p(E)} = 0$. Malheureusement, $\|f\|_{\mathcal{L}^p(E)} = 0$ n'implique pas que $f = 0$, ce qui veut dire que cette application ne définit pas une norme sur $\mathcal{L}^p(E)$ [85]. Néanmoins, on peut constater que pour $1 \leq p \leq +\infty$, $\|f\|_{\mathcal{L}^p(E)} = 0$ est équivalent à $f = 0$ presque partout [86].

84. On n'insiste plus, mais on utilise ici qu'on peut "additionner" deux inégalités valables presque partout car on peut écarter l'union des deux ensembles négligeables qui "posent problèmes", qui est encore un ensemble négligeable. Dorénavant on ne préisera plus ces manipulations.
85. On dit que c'est une semi-norme.
86. Distinguer les cas $p < +\infty$ pour lesquels on utilise la proposition 2.15 et le cas $p = +\infty$ pour lequel on utilise le lemme 5.4.

Définition-Proposition 5.12. Soit $p \in [1, +\infty]$. L'ensemble $\mathcal{N}$ des fonctions nulles presque partout sur E est un sous-espace vectoriel de $\mathcal{L}^p(E)$, et donc on peut définir l'espace vectoriel quotient [87] $L^p(E) = \mathcal{L}^p(E)/\mathcal{N}$.

Si $[f]$ désigne la classe d'équivalence de f, on peut définir [88]

$$\|[f]\|_{L^p(E)} := \|g\|_{\mathcal{L}^p(E)} \text{ pour tout } g \in [f]. \tag{5.3}$$

L'espace obtenu $(L^p(E), \|\cdot\|_{L^p(E)})$ est un espace vectoriel normé. S'il n'y a pas d'ambiguïté, on pourra utiliser la notation $\|\cdot\|_p$ à la place de $\|\cdot\|_{L^p(E)}$.

Remarque 5.13. Dans la pratique, on traitera ces classes d'équivalence comme des fonctions ordinaires, et on écrira f à la place de $[f]$. On parlera donc de "fonction de L^p", de son intégrale, etc. On fera juste attention à ce que les opérations utilisées sur ces "fonctions" sont bien "légitimes", au sens où elles passent au quotient pour la relation d'égalité presque partout. Voir l'exercice 5.14 pour des exemples. En particulier, les inégalités de Hölder et Minkowski sont valables pour ces "fonctions classes d'équivalences" et pour les normes définies par (5.3).

Démonstration. Il est facile de voir que la relation d'égalité presque partout (voir la note 87) est une relation d'équivalence [89]. C'est un résultat classique d'algèbre que l'espace quotient peut alors être muni d'une structure d'espace vectoriel, en considérant les lois : si $[f], [g] \in L^p(E)$ et $\alpha \in \mathbb{K}$, on pose

$$[f] + [g] := [f + g], \qquad \alpha.[f] = [\alpha.f],$$

pour lesquelles $[0]$ sera l'élément neutre ; ces définitions sont légitimes car ne dépendent pas des représentants (f, g) choisis.

La formule (5.3) est aussi légitime car si $[f] \in L^p(E)$, et si (f, g) sont deux représentants de $[f]$ (autrement dit $f \in [f]$ et $g \in [f]$), alors $f(x) = g(x)$ pour presque tout $x \in E$ et donc $\|f\|_{\mathcal{L}^p(E)} = \|g\|_{\mathcal{L}^p(E)}$. La fonction $\|\cdot\|_{L^p(E)}$ est homogène, elle satisfait l'inégalité triangulaire d'après la proposition 5.11, et elle est définie positive, car si $\|[f]\|_{L^p(E)} = 0$, alors $f = 0$ presque partout, i.e. $[f] = [0]$; autrement dit, le seul élément de $L^p(E)$ dont la norme est nul est l'élément neutre de $L^p(E)$, ce qui conclut la démonstration. $\square$

Exercice 5.14. Dans cet exercice, on manipule des exemples d'opérations simples sur les fonctions de $\mathcal{L}^p(E)$, et on cherche à voir si ces opérations sont aussi légitimes pour les fonctions de $L^p(E)$; cela revient à dire que les résultats obtenus ne dépendent pas des représentants. D'un point de vue algébrique, on dit que ces opérations "passent au quotient". On fixe $p \in [1, +\infty]$ [90]

1. Afin de fixer les idées, on choisit $(E, \mathscr{A}, \mu) = (I, \mathscr{B}(I), \lambda)$ dans cette question. Pour chacune des opérations suivantes qu'on peut définir sur $\mathcal{L}^p(E)$, dire si on peut les définir sur $L^p(E)$:
   (a) l'évaluation en un point $f \mapsto f(x_0)$ où $x_0 \in I$,
   (b) le supremum $f \mapsto \sup_{x \in I} f(x)$,
   (c) une fonction de type intégrale : $f \mapsto \int_I |f(x)| e^{-x^2} dx$,
   (d) (ici $I = \mathbb{R}$) la translation d'une fonction $\tau_a :\mapsto \tau_a f$ où $a \in \mathbb{R}$.

87. Il s'agit de l'espace obtenu en quotientant par la relation d'équivalence

$$\forall (f, g) \in \mathcal{L}^p(E), \qquad f \sim g \quad \Leftrightarrow \quad f - g \in N \Leftrightarrow \quad f = g \quad \mu\text{-presque partout.}$$

On dit aussi qu'on passe au quotient pour "l'égalité presque partout".

88. Généralement, quand on a un espace X muni d'une semi-norme $\|\cdot\|_X$, et F un s.e.v. de X, on peut poser

$$\forall [f] \in X/F, \quad \|[f]\|_{X/F} = \inf_{g \in [f]} \|g\|_X$$

mais ici l'infimum n'est pas utile car la valeur de $\|g\|_X$ est la même pour tout $g \in [f]$.

89. Cela vient du fait que $\mathcal{N}$ est un sous-espace vectoriel de $\mathcal{L}^p$.

90. Le paramètre p n'a pas d'importance dans cet exercice, on aurait pu travailler dans l'espace vectoriel des fonctions mesurables, qu'on quotiente par $\mathcal{N}$; on obtient ainsi un espace qui contient tous les espaces $L^p(E)$, mais qui n'est pas un espace vectoriel normé a priori.

2. Dans cette question, on regarde $(E, \mathscr{A}, \mu) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), m)$ où m est la mesure de comptage. Montrer que

$$\mathcal{L}^p(\mathbb{N}) = L^p(\mathbb{N}) = \Big\{(u_n)_{n\in\mathbb{N}},\ \ \|u\|_{L^p(E)} < +\infty\Big\},$$

$$\text{et} \quad \forall u \in \mathbb{K}^{\mathbb{N}},\ \ \|u\|_{L^p(\mathbb{N})} = \begin{cases} \left(\displaystyle\sum_{n\in\mathbb{N}} |u_n|^p\right)^{1/p} & \text{si} \quad 1 \leq p < +\infty, \\ \displaystyle\sup_{n\in\mathbb{N}} |u_n| & \text{si} \quad p = +\infty. \end{cases}$$

Dans ce cas on note habituellement $\ell^p(\mathbb{N}) = L^p(\mathbb{N}, \mathcal{P}(\mathbb{N}), m)$.

3. [*] Dans cette question, on considère E un ensemble quelconque, $a \in E$, et $(E, \mathscr{A}, \mu) = (E, \mathcal{P}(E), \delta_a)$ où δ_a est la mesure de Dirac en a. Montrer que l'espace $L^p(E)$ est isomorphe à $\mathbb{K}$.

Exercice 5.15 (Support d'une classe de fonctions)**.** On travaille dans $(E, \mathscr{B}(E), \lambda_d)$ où E est un ensemble borélien de $\mathbb{R}^d$.

1. Soit $f : E \to \mathbb{K}$ borélienne. Justifier que l'ensemble $\{f \neq 0\}$ ne peut pas être défini si f est une classe de fonction.
2. On pose

$$Supp(f) = \bigcap_{F\in A_f} F, \qquad \text{où} \quad A_f = \Big\{F \text{ fermé de } E \text{ tel que } f = 0 \text{ presque partout sur } F^c\Big\}$$

   (a) Montrer que

$$Supp(f)^c = \bigcup_{O\in N_f} O \qquad \text{où} \quad N_f = \Big\{O \text{ ouvert de } E \text{ tel que } f = 0 \text{ presque partout sur } O\Big\}.$$

   On dit que $Supp(f)^c$ est le plus grand ouvert de nullité de f.

   (b) Soit $f : E \to \mathbb{K}$ borélienne. Montrer que $f = 0$ presque partout sur $Supp(f)^c$.

   (c) Montrer que si $f : E \to \mathbb{K}$ est continue, alors

$$Supp(f) = \overline{\{f \neq 0\}}.$$

   (d) Montrer que si (f, g) sont deux fonctions boréliennes de E dans $\mathbb{K}$ et si $f = g$ presque partout, alors $Supp(f) = Supp(g)$.

   Ainsi on peut définir $Supp(f)$ quand f est une classe de fonctions. On l'appelle support de f.

Remarque 5.16. On a donc en main des espaces vectoriels normés $(L^p(E), \|\cdot\|_p)$ pour $p \in [1, +\infty]$. On peut donc utiliser les notions du Chapitre III. Par exemple on peut parler de convergence $L^p(E)$ pour une suite $(f_n)_{n\in\mathbb{N}}$ d'éléments de L^p : on dira que $(f_n)_{n\in\mathbb{N}}$ converge vers f si $\|f_n - f\|_{L^p(E)} \xrightarrow[n\to+\infty]{} 0$. Remarquons que si $p < +\infty$, cela équivaut à

$$\int_E |f_n - f|^p d\mu \xrightarrow[n\to+\infty]{} 0.$$

Remarque 5.17. Pour $p = 2$, l'application

$$(f, g) \in L^2(E) \times L^2(E) \mapsto \langle f|g\rangle = \int_E f\overline{g}\, d\mu(x).$$

définit un produit scalaire sur $L^2(E)$ dont la norme associée est la norme $\|\cdot\|_2$, voir le Chapitre IX.

Exercice 5.18 (Inclusion des espaces L^p dans un espace de mesure finie)**.** Soit $(E, \mathscr{A}, \mu)$ un espace mesuré tel que μ est finie.

1. Montrer que si $(p,q) \in \mathbb{R}^2$ sont tels que $1 \leq p \leq q \leq +\infty$, alors[91]
$$L^\infty(E) \subset L^q(E) \subset L^p(E) \subset L^1(E).$$
2. Montrer que si f est mesurable, alors[92]
$$\|f\|_p \leq (\mu(E))^{\frac{1}{p}-\frac{1}{q}} \|f\|_q.$$
Remarquez que si μ est une probabilité, on obtient donc la croissance de la fonction $p \mapsto \|f\|_p$.
3. On suppose que $(E, \mathscr{A}, \mu) = (I, \mathscr{B}(I), \lambda)$ où I est un intervalle de $\mathbb{R}$ et λ est la mesure de Lebesgue.
   (a) À quelle condition a-t-on que λ est une mesure finie sur I ?
   (b) Montrer qu'aucune inclusion entre $L^p(I)$ et $L^q(I)$ (pour $p \neq q$) n'est satisfaite si I est de mesure infinie.
4. On suppose que $(E, \mathscr{A}, \mu) = (\Omega, \mathscr{B}(\Omega), \lambda_d)$ où Ω est un ouvert de $\mathbb{R}^d$.
   (a) Montrer que si Ω est borné, alors Ω est de mesure de Lebesgue finie. Existe-t-il des ouverts non bornés de $\mathbb{R}^d$ qui sont de mesure finie ?
   (b) On ne suppose pas Ω de mesure finie : montrer que pour tout $p \in [1, +\infty]$,
$$L^p(\Omega) \subset L^1_{loc}(\Omega) := \left\{ f : \Omega \to \mathbb{K},\ \forall K \text{ compact inclus dans } \Omega,\ \int_K |f| d\lambda_d < +\infty \right\}.$$

Exercice 5.19 (Espaces $\ell^p(\mathbb{N})$)**.** Pour $p \in [1, \infty[$, on pose
$$\ell^p(\mathbb{N}) = \left\{ a = (a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}},\ \ \|a\|_p = \left(\sum_{n\in\mathbb{N}} |a_n|^p \right)^{\frac{1}{p}} < \infty \right\}$$
et pour $p = \infty$ on pose
$$\ell^\infty(\mathbb{N}) = \left\{ a = (a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}, \|a\|_\infty = \sup_{n\in\mathbb{N}} |a_n| < \infty \right\}.$$
1. Montrer que pour tout $1 \leq p \leq q \leq \infty$, $\ell^1(\mathbb{N}) \subset \ell^p(\mathbb{N}) \subset \ell^q(\mathbb{N}) \subset \ell^\infty(\mathbb{N})$ et que
$$\forall a \in \mathbb{R}^{\mathbb{N}}, \|a\|_\infty \leq \|a\|_q \leq \|a\|_p \leq \|a\|_1.$$
2. Montrer que si $1 < p < q < \infty$, ces inclusions $\ell^1(\mathbb{N}) \subset \ell^p(\mathbb{N}) \subset \ell^q(\mathbb{N}) \subset \ell^\infty(\mathbb{N})$ sont strictes.

Exercice 5.20. On travaille dans $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$. On pose $\forall x \in \mathbb{R}, \forall n \in \mathbb{N}^*,\ \ \alpha_n(x) = \frac{1}{2n} \mathbb{1}_{[-n,n]}(x)$.
1. Montrer que pour tout $n \in \mathbb{N}$, $\alpha_n \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, et que $\alpha_n \xrightarrow[n\to+\infty]{} 0$ dans $L^2(\mathbb{R})$.
2. Montrer que la suite $(\alpha_n)_{n\in\mathbb{N}}$ n'a pas de limite dans $L^1(\mathbb{R})$.
3. Soit $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. On pose $\forall n \in \mathbb{N}$, $\forall x \in \mathbb{R}$, $\ f_n(x) = f(x) - I\alpha_n(x)$ où $I = \int_{\mathbb{R}} f(y)dy$. Montrer que $f_n \xrightarrow[n\to\infty]{} f$ dans $L^2(\mathbb{R})$.
4. On pose
$$A = \left\{ f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}),\ \int_{\mathbb{R}} f(x)dx = 0 \right\}.$$
Montrer que A est dense dans $L^2(\mathbb{R})$.
5. Montrer que A n'est pas dense dans $L^1(\mathbb{R})$.

91. On peut voir que ces inclusions sont une conséquence de la question suivante ; mais on peut aussi faire une preuve à la main qui n'utilise pas l'inégalité de Hölder.

92. Notez que ces inégalités sont plus intéressantes que de juste savoir les inclusions de la question précédentes. En effet, elles disent en fait que l'inclusion canonique
$$\begin{array}{ccc} L^q(E) & \longrightarrow & L^p(E) \\ f & \longmapsto & f \end{array}$$
sont continues ; autrement dit, elles montrent l'inclusion de $L^q(E)$ dans $L^p(E)$, mais aussi le fait que si $(f_n)_{n\in\mathbb{N}}$ converge vers f dans $L^q(E)$, alors $(f_n)_{n\in\mathbb{N}}$ converge vers f dans $L^p(E)$.

5.2 Complétude

On va voir dans ce paragraphe un résultat important qui justifie a posteriori l'importance de l'intégrale de Lebesgue par rapport à l'intégrale de Riemann, du point de vue de l'analyse fonctionnelle.

Commençons par donner une version L^p du théorème de convergence dominée, qui se déduit du théorème standard 3.8 :

Lemme 5.21 (**Théorème de convergence dominée version** L^p). *Soit* $p \in [1, +\infty[$, $(f_n)_{n\in\mathbb{N}}$ *une suite de fonctions mesurables et* f *une fonction mesurable tels que*

— *la suite* $(f_n)_{n\in\mathbb{N}}$ *converge* μ*-presque partout vers* f,

— *il existe* $g \in L^p(E)$ *telle que*

$$\forall n \in \mathbb{N}, \quad |f_n| \leq g \quad \mu\text{-presque partout..}$$

Alors $f \in L^p(E)$ *et* $\|f_n - f\|_{L^p(E)} \xrightarrow[n\to+\infty]{} 0$, *c'est-à-dire que* $(f_n)_{n\in\mathbb{N}}$ *converge vers* f *dans* $L^p(E)$.

Démonstration. On peut passer à la limite dans l'hypothèse de domination, ce qui donne que $f \in L^p(E)$. Pour la convergence, on applique le théorème 3.8 à la suite $\forall n \in \mathbb{N}, \quad h_n := |f_n - f|^p$ qui est mesurable[93], et satisfait la domination[94]

$$\forall n \in \mathbb{N}, |h_n| \leq 2^{p-1}|g|^p \in L^1(E) \quad \mu\text{-presque partout.}$$

On en déduit $\displaystyle\int_E h_n d\mu \xrightarrow[n\to+\infty]{} 0$, ce qui prouve bien la convergence L^p. □

Théorème 5.22 (**Théorème de Riesz-Fischer**). *Pour tout* $1 \leq p \leq +\infty$, *l'espace* $L^p(E)$ *est complet.*

Démonstration. Supposons d'abord que $1 \leq p < +\infty$. Soit $(u_n)_{n\in\mathbb{N}}$ une suite de $L^p(E)$ telle que $\sum_{n\in\mathbb{N}} \|u_n\|_p < +\infty$ et montrons que $\sum_{n\in\mathbb{N}} u_n$ converge dans $L^p(E)$. D'après la proposition 2.18 du Chapitre III, cela montrera que $L^p(E)$ est complet.

Posons $S_n(x) = \displaystyle\sum_{k=0}^{n} |u_k(x)|$. La suite $(S_n)_{n\in\mathbb{N}}$ est croissante, donc elle converge ponctuellement vers S où $S(x) = \displaystyle\lim_{n\to+\infty} S_n(x) = \sup_{n\in\mathbb{N}} S_n(x) \in [0, +\infty]$. De plus, par le théorème de convergence monotone 2.9,

$$\int_E S_n^p \, d\mu(x) \xrightarrow[n\to+\infty]{} \int_E S^p \, d\mu(x).$$

Par ailleurs, par l'inégalité de Minkowski (proposition 5.11),

$$\|S_n\|_p \leq \sum_{k=1}^{n} \|u_k\|_p \leq \sum_{k=1}^{+\infty} \|u_k\|_p.$$

Ceci montre que $\int_E |S|^p \, d\mu(x) \leq \sum_{k=1}^{+\infty} \|u_k\|_p < +\infty$, et donc en particulier par la proposition 2.19 $S(x) < +\infty$ presque pour tout $x \in E$. Ainsi, pour presque tout $x \in E$, la série numérique $\sum_{k\in\mathbb{N}} u_k(x)$ est absolument convergente dans $\mathbb{R}$, donc est convergente. Appelons sa somme $f(x)$. On pose $f(x) = 0$ sur l'ensemble négligeable sur lequel la série précédente ne converge pas. La fonction f ainsi définie

93. Si on souhaite détailller, on peut dire que h_n s'obtient par composition d'une fonction mesurable $(f_n - f)$ avec une fonction continue $(x \mapsto |x|^p)$ donc borélienne

94. On utilise l'inégalité

$$\forall (a, b) \in \mathbb{R}_+, \quad (a + b)^p \leq 2^{p-1}(a^p + b^p)$$

qui peut se déduire de la convexité de la fonction $x \in \mathbb{R}_+ \mapsto x^p$ qui donne $\left(\frac{a+b}{2}\right)^p \leq \frac{1}{2}a^p + \frac{1}{2}b^p$. Notez qu'on peut montrer de manière plus élémentaire une inégalité similaire en écrivant plus simplement :

$$\forall (a, b) \in \mathbb{R}_+, \quad (a + b)^p \leq (2 \max\{a, b\})^p \leq 2^p(a^p + b^p).$$

La constante 2^p est moins bonne car plus grande que 2^{p-1}, mais ici ça n'a pas d'importance et la preuve est plus élémentaire.

est mesurable comme limite presque partout d'une suite de fonctions mesurables. De plus, on a la domination

$$\left|\sum_{k=0}^{n} u_k(x)\right| \leq S(x)$$

pour tout $x \in E$ et pour tout $n \in \mathbb{N}$ par l'inégalité triangulaire dans $\mathbb{R}$, or on vient de voir que $S \in L^p(E)$. Par le lemme 5.21, on en déduit que $f \in L^p(E)$ et que $\sum_{k=0}^{n} u_k$ converge vers f au sens de la norme $\|\cdot\|_p$, ce qui conclut le cas $p < +\infty$.

Pour $p = +\infty$ nous allons faire une preuve plus classique. Soit $(f_n)_{n\in\mathbb{N}}$ est une suite de Cauchy dans $L^\infty(E)$: par le Lemme 5.4, il existe un ensemble $N \in \mathscr{A}$ de mesure nulle tel que

$$|f_n(x)| \leq \|f_n\|_\infty, \quad \text{et} \quad |f_n(x) - f_m(x)| \leq \|f_n - f_m\|_\infty \text{ pour tout } x \in E \setminus N \text{ et tous } m, n \in \mathbb{N}. \quad (5.4)$$

En effet, ces inégalités ont lieu a priori en dehors d'ensembles négligeables indexés par n et m, qui sont en nombre dénombrable, et dont la réunion N est également négligeable. Par conséquent, si $x \in E \setminus N$, $(f_n(x))_{n\in\mathbb{N}}$ est une suite de Cauchy dans $\mathbb{R}$, elle admet donc une limite qu'on note $f(x)$. Par ailleurs, on pose $f(x) = 0$ si $x \in N$. La fonction f ainsi définie est mesurable comme limite presque partout d'une suite de fonctions mesurables. De plus, comme $(f_n)_{n\in\mathbb{N}}$ est de Cauchy dans $L^\infty(E)$, elle y est bornée, c'est-à-dire qu'il existe $M > 0$ tel que $\|f_n\|_\infty \leq M$ pour tout $n \in \mathbb{N}$. En passant à la limite dans la première inégalité de (5.4), il vient $|f(x)| \leq M$ pour presque tout $x \in E$, soit $f \in L^\infty(E)$. Enfin, par hypothèse, pour tout $\varepsilon > 0$, il existe $n_0 \in \mathbb{N}$ tel que pour tous m, $n \geq n_0$, $\|f_n - f_m\|_\infty \leq \varepsilon$. En passant à la limite quand $m \to +\infty$ dans la deuxième inégalité de (5.4), il vient $|f_n(x) - f(x)| \leq \varepsilon$ pour tout $n \geq n_0$ et presque tout $x \in E$, soit $\|f_n - f\|_\infty \leq \varepsilon$ pour tout $n \geq n_0$, ce qui montre bien que $f_n \xrightarrow[n\to+\infty]{} f$ dans $L^\infty(E)$. □

Corollaire 5.23. *L'espace $L^2(E)$ est un espace de Hilbert.*

Remarquons que si on reprend la démonstration du théorème de Riesz-Fischer (et de la proposition 2.18 du Chapitre III), on obtient le résultat suivant qui peut être vu comme une réciproque partielle du théorème de convergence dominée :

Corollaire 5.24. *[Réciproque partielle du théorème de convergence dominée L^p] Soit $p \in [1, +\infty]$. Soit $(f_n)_{n\in\mathbb{N}}$ une suite de $L^p(E)$ telle que $f_n \xrightarrow[n\to+\infty]{} f$ dans $L^p(E)$. Alors il existe une suite extraite $(f_{\sigma(n)})_{n\in\mathbb{N}}$ et une fonction $g \in L^p(E)$ telles que*

— *$f_{\sigma(n)} \xrightarrow[n\to+\infty]{} f$ presque partout quand $n \to +\infty$,*

— *pour tout $n \in \mathbb{N}$, $|f_{\sigma(n)}| \leq g$ μ-presque partout dans E.*

Remarque 5.25. On retient donc que la convergence L^p implique la convergence presque partout à extraction près (l'aspect domination de la réciproque ci-dessous sert moins souvent). Cette extraction est en général nécessaire (sauf si $p = +\infty$), comme on peut le voir avec l'exemple suivant : si on se place dans l'espace mesuré $([0,1], \mathscr{B}([0,1]), \lambda)$, on pose

$$\forall m \in \mathbb{N}, \forall k \in [\![0, 2^m - 1]\!], \quad f_{2^m+k} = \mathbb{1}_{[\frac{k}{2^m}, \frac{k+1}{2^m}[}.$$

On a bien ainsi défini une suite $(f_n)_{n\in\mathbb{N}}$ (car tout entier $n \in \mathbb{N}$ s'écrit de manière unique sous la forme $2^m + k$ pour un certain $m \in \mathbb{N}$ et $k \in [\![0, 2^m - 1]\!]$) ; on vérifie que $\|f_{2^m+k}\|_p = \dfrac{1}{2^{m/p}} \xrightarrow[m\to+\infty]{} 0$ donc $\|f_n\|_p \xrightarrow[n\to+\infty]{} 0$. Pourtant, pour tout $x \in [0,1[$, il existe une suite $(k_m^x)_{m\in\mathbb{N}}$ telle que $\frac{k_m^x}{2^m} \leq x < \frac{k_m^x+1}{2^m}$ et alors $\forall m \in \mathbb{N}$, $f_{2^m+k_m^x}(x) = 1$, ce qui montre que la suite $(f_n(x))_{n\in\mathbb{N}}$ ne tend pas vers 0.

Démonstration. Le cas $p = +\infty$ est évident par définition du sup-essentiel. Dans ce cas, il n'y a même pas besoin d'extraire de sous-suite. Supposons $1 \leq p < +\infty$: comme $(f_n)_{n\in\mathbb{N}}$ est convergente, elle est de Cauchy dans $L^p(E)$. On reprend alors la démonstration de la proposition 2.18 du Chapitre III qui montre l'existence d'une sous-suite $(f_{\sigma(n)})_{n\in\mathbb{N}}$ telle que $\|f_{\sigma(n+1)} - f_{\sigma(n)}\|_p \leq \frac{1}{2^n}$. Ainsi en posant $S(x) = \sum_{n\in\mathbb{N}} |f_{\sigma(n+1)}(x) - f_{\sigma(n)}(x)|$, on a $S \in L^p(E)$, d'où $S < +\infty$ presque partout : on peut définir

$g(x) = \sum_{j\in\mathbb{N}}(f_{\sigma(n+1)}(x) - f_{\sigma(n)}(x))$ presque partout, et $(f_{\sigma(n)})_{n\in\mathbb{N}}$ converge p.p. vers $\tilde{f} = g + f_{n_0}$. De plus

$$\forall j \in \mathbb{N},\quad |f_{\sigma(n)}(x)| \leq |f_{\sigma(0)}(x)| + \sum_{k=0}^{n-1} |f_{\sigma(k+1)}(x) - f_{\sigma(k)}(x)| \leq |f_{\sigma(0)}(x)| + S(x),$$

est bien une domination puisque $|f_{\sigma(0)}| + S \in L^p(E)$. Par le théorème de convergence dominée (lemme 5.21), on a que $f_{\sigma(n)}$ converge en norme L^p vers $\tilde{f}$ donc par unicité de la limite $f = \tilde{f}$ presque partout, ce qui termine la démonstration. □

5.3 Convolution

Dans ce paragraphe, on se place dans l'espace mesuré $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d), \lambda_d)$ [95].

Définition 5.26. Soit (f, g) deux fonctions de $\mathbb{R}^d$ à valeurs dans $\mathbb{K}$ ou $\overline{\mathbb{R}}$ boréliennes. Si

$$\int_{\mathbb{R}^d} |f(t)g(x-t)|dt < +\infty \qquad \text{pour un certain } x \in \mathbb{R}^d \tag{5.5}$$

alors on posera

$$f * g(x) = \int_{\mathbb{R}^d} f(t)g(x-t)dt,$$

qu'on appelle convolée de f et g au point x.

Proposition 5.27. *Soit (f, g) deux fonctions de $\mathbb{R}^d$ à valeurs dans $\mathbb{K}$ ou $\overline{\mathbb{R}}$ boréliennes.*

- *Soit $x \in \mathbb{R}^d$. Si $f * g(x)$ est bien défini (i.e. (5.5) est vérifiée), alors $g * f(x)$ aussi et $f * g(x) = g * f(x)$.*
- *On suppose que (5.5) est vérifiée pour tout $x \in \mathbb{R}^d$. Alors on $f * g$ est une fonction borélienne.*

Démonstration. La commutativité vient directement d'un changement de variable $y = x - t$ dans l'intégrale qui définit $f * g(x)$.

Pour la mesurabilité de $f * g$, on commence par remarquer que $(u, v) \mapsto f(u)$ et $(u, v) \mapsto g(v)$ sont mesurables de $(\mathbb{R}^d \times \mathbb{R}^d, \mathscr{B}(\mathbb{R}^d) \otimes \mathscr{B}(\mathbb{R}^d))$ à valeurs dans $(K, \mathscr{B}(\mathbb{K}))$, donc par la proposition 1.67, la fonction $(u, v) \mapsto f(u)g(v)$ l'est aussi. Or $\mathscr{B}(\mathbb{R}^d) \otimes \mathscr{B}(\mathbb{R}^d) = \mathscr{B}(\mathbb{R}^d \times \mathbb{R}^d)$, donc par composition la fonction $(x, t) \mapsto f(t)g(x-t)$ est mesurable de $(\mathbb{R}^d \times \mathbb{R}^d, \mathscr{B}(\mathbb{R}^d) \otimes \mathscr{B}(\mathbb{R}^d))$ à valeurs dans $(K, \mathscr{B}(\mathbb{K}))$. D'après le théorème de Fubini appliqué à la mesure $\lambda_d \otimes \lambda_d$ [96], la fonction $x \mapsto \int_{\mathbb{R}^d} f(t)g(x-t)dt$ est bien borélienne. □

Exercice 5.28. Attention, la loi $*$ n'est pas associative dans le cas général. On laisse le lecteur se convaincre en comparant $f * (g * h)$ et $(f * g) * h$ où

$$f = \mathbb{1}_{\mathbb{R}_+}, \quad g = \mathbb{1}_{[-1,0]} - \mathbb{1}_{[0,1]}, \quad h = 1.$$

Exercice 5.29. On suppose que (5.5) est vérifiée pour tout $x \in \mathbb{R}^d$. Montrer que [97]

$$\{f * g \neq 0\} \subset \{f \neq 0\} + \{g \neq 0\}.$$

On donne désormais des conditions suffisantes pour que la convolution des fonctions f et g soit bien définie :

95. La convolution n'a en effet pas de sens dans n'importe quel espace mesuré ; il faut que l'espace de travail (ici $\mathbb{R}^d$) soit invariant par translation, et que la mesure (ici λ_d) le soit aussi. On définira au Chapitre VIII un autre exemple dans lequel on peut définir une convolution, à savoir $\mathbb{R}/2\pi\mathbb{Z}$ (aussi muni des boréliens et de la mesure de Lebesgue.

96. On doit dire ici que la mesure λ_d est σ-finie.

97. Rappelons que la somme (dite de Minkowski) de deux sous-ensembles A et B de $\mathbb{R}^d$ est définie comme

$$A + B = \{a + b, a \in A, b \in B\}.$$

Proposition 5.30 (Convolution $L^1 - L^1$)**.** [98] *Soient $(f,g) \in L^1(\mathbb{R}^d)^2$. Alors pour presque tout $x \in \mathbb{R}^d$, la fonction $f(\cdot)g(x-\cdot)$ est intégrable sur $\mathbb{R}^d$, et donc on peut définir $f * g$ presque partout. De plus $f * g \in L^1(\mathbb{R}^d)$ et*

$$\|f * g\|_{L^1(\mathbb{R}^d)} \leq \|f\|_{L^1(\mathbb{R}^d)} \|g\|_{L^1(\mathbb{R}^d)}.$$

Remarque 5.31. On peut donc considérer l'espace $(L^1(\mathbb{R}^d), +, \cdot, *)$ qui est une algèbre de Banach (voir aussi la remarque 2.46 du Chapitre III). Pour le voir, il faut montrer que la loi $*$ est associative dans ce contexte, c'est-à-dire que si $(f,g,h) \in L^1(\mathbb{R}^d)$, alors $f * (g * h) = (f * g) * h$. Voir [BP12, page 268].

Remarquons aussi que cette algèbre n'a pas d'unité pour la loi $*$: voir aussi [BP12, page 268] pour une preuve, et nous en donnerons une autre au Chapitre VIII.

Démonstration. On peut appliquer le théorème de Fubini-Tonelli qui donne :

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(t)g(x-t)| dt dx = \int_{\mathbb{R}^d} |f(t)| \underbrace{\int_{\mathbb{R}^d} |g(x-t)| dx}_{= \int_{\mathbb{R}^d} |g(x)| dx} dt = \|f\|_1 \|g\|_1 < +\infty.$$

Avec la proposition 2.19 on en déduit que $\int_{\mathbb{R}^d} |f(t)g(x-t)| dt < +\infty$ presque partout. De plus, le théorème de Fubini permet d'écrire :

$$\|f * g\|_1 = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(t)g(x-t) dt \right| dx \leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(t)||g(x-t)| dx dt = \|f\|_1 \|g\|_1.$$

□

Proposition 5.32 (Convolution $L^p - L^{p'}$)**.** *Soient $(f,g) \in L^p(\mathbb{R}^d) \times L^{p'}(\mathbb{R}^d)$ où $p \in [1, +\infty]$ et $\frac{1}{p} + \frac{1}{p'} = 1$. Alors pour tout $x \in \mathbb{R}^d$, la fonction $f(\cdot)g(x-\cdot)$ est intégrable sur $\mathbb{R}^d$, et donc on peut définir $f * g$. De plus $f * g$ est une fonction continue*[99]*, bornée et*

$$\|f * g\|_\infty \leq \|f\|_{L^p(\mathbb{R}^d)} \|g\|_{L^{p'}(\mathbb{R}^d)}. \tag{5.6}$$

Démonstration. Par l'inégalité de Hölder,

$$\int_{\mathbb{R}^d} |f(t)g(x-t)| dt \leq \|f\|_p \|g\|_{p'} < +\infty$$

donc la fonction $t \mapsto f(t)g(x-t)$ est bien intégrable sur $\mathbb{R}^d$. Ainsi $f * g$ est bien définie, et on a

$$\forall x \in \mathbb{R}^d, \quad |f * g(x)| \leq \int_{\mathbb{R}^d} |f(t)g(x-t)| dt \leq \|f\|_p \|g\|_{p'}.$$

Il reste à montrer la continuité de $f * g$: si $x \in \mathbb{R}^d$ et $h \in \mathbb{R}^d$, on peut écrire

$$f*g(x+h) - f*g(x) = \int_{\mathbb{R}^d} f(t) \left[g(x+h-t) - g(x-t)\right] dt = \int_{\mathbb{R}^d} f(t)(\tau_{-h}g - g)(x-t) dt = f*(\tau_{-h}g - g)(x)$$

donc l'inégalité (5.6) appliquée à f et $\tau_{-h}g - g$ donne

$$|f * g(x+h) - f * g(x)| \leq \|f\|_p \|\tau_{-h}g - g\|_{p'}.$$

Par continuité des translations dans L^p (corollaire 5.44), valable si $p' < +\infty$ c'est-à-dire si $p > 1$, on obtient la continuité de $f * g$.

Si $p = 1$, on peut échanger les rôles de f et g par commutativité. □

98. Remarquons que l'opération $*$ peut bien se définir sur les classes de fonctions, car si $f_1 = f_2$ et $g_1 = g_2$ presque partout, on constate que $f_1 * g_1 = f_2 * g_2$ presque partout.

99. Et même uniformément continue en fait.

Exercice 5.33. Montrer que si $p \in]1,+\infty[$ et $(f,g) \in L^p(\mathbb{R}^d) \times L^{p'}(\mathbb{R}^d)$, alors

$$f * g(x) \xrightarrow[|x|\to+\infty]{} 0.$$

(On pourra utiliser le théorème 5.38)

Exercice 5.34 (Convolution L^1_{loc} avec les fonctions bornées à support compact [100])**.** Soit $f \in L^1_{loc}(\mathbb{R}^d)$ et $g \in L^\infty(\mathbb{R}^d)$ à support compact. Montrer que $f * g$ est définie en tout point.

Exercice 5.35 (Convolution $L^1 - L^p$)**.** Soit $f \in L^p(\mathbb{R}^d)$ avec $p \in [1,+\infty]$, et $g \in L^1(\mathbb{R}^d)$. Montrer que $f * g(x)$ est bien définie pour presque tout $x \in \mathbb{R}^d$, et que

$$\|f * g\|_p \leq \|f\|_p \|g\|_1.$$

Remarque 5.36 (**)**.** On peut énoncer un résultat général qui inclut à la fois la convolution $L^1 - L^1$, la convolution $L^p - L^{p'}$, et la convolution $L^1 - L^p$: il s'agit de l'inégalité de Young, qui dit que si $f \in L^p(\mathbb{R}^d)$, $g \in L^q(\mathbb{R}^d)$, et $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ avec $(p,q,r) \in [1,+\infty]^3$, alors $f * g(x)$ est bien défini pour presque tout $x \in \mathbb{R}^d$, et

$$\|f * g\|_r \leq \|f\|_p \|g\|_q.$$

La preuve est un peu plus longue que pour les cas particuliers ; on peut trouver une preuve dans [Amr08, page 84]. Il y a une preuve très différente dans [QZ13, page 481] ou il est montré par la théorie de l'interpolation (via le théorème de Riez-Thorin) que l'inégalité dans des cas particuliers peut impliquer le cas général.

5.4 Résultats de densité

Soit $(E, \mathscr{A}, \mu)$ un espace mesuré général. Alors on a le premier résultat de densité élémentaire suivant :

Théorème 5.37. *Pour tout $1 \leq p \leq +\infty$, l'espace $\{f \in L^p(E), f \text{ étagée}\}$ est dense dans $L^p(E)$.*

Démonstration. Soit $f \in L^p(E)$, en décomposant $f = f_+ - f_-$, on peut supposer sans restreindre la généralité que $f \geq 0$. On applique le lemme 1.80, qui donne l'existence d'une suite $(f_n)_{n\in\mathbb{N}}$ de fonctions étagées qui converge en croissant vers f.

Si $p = +\infty$, la fonction f est égale presque partout à une fonction bornée, et le lemme 1.80 montrait que la convergence de $(f_n)_{n\in\mathbb{N}}$ vers f est uniforme, ce qui correspond bien à la convergence $L^\infty(E)$.

Si $p \in [1,+\infty[$, on constate d'abord que $0 \leq f_n \leq f$ pour tout $n \in \mathbb{N}$, ce qui donne d'une part que $f_n \in L^p(E)$ pour tout $n \in \mathbb{N}$, et d'autre part qu'on peut appliquer le lemme 5.21, ce qui montre que $(f_n)_{n\in\mathbb{N}}$ converge vers f en norme $L^p(E)$. □

Pour la suite de ce paragraphe, on se place dans le cas $(E, \mathscr{A}, \mu) = (\Omega, \mathscr{B}(\Omega), \lambda_d)$ où Ω est un ouvert de $\mathbb{R}^d$:

Théorème 5.38 (Densité des fonctions continues à support compact)**.** *Soit $p \in [1,+\infty[$. Alors l'espace*

$$C^0_c(\Omega) := \left\{ f : \Omega \to \mathbb{K} \quad \textit{continue à support compact dans } \Omega \right\}$$

est dense dans $L^p(\Omega)$.

Remarque 5.39. Notons l'importance de considérer des fonctions à support compact : les fonctions continues sur un ouvert Ω ne sont en effet a priori pas dans $L^p(\Omega)$, car leur comportement au voisinage du bord de Ω [101] peut être quelconque. Alors qu'en demandant aux fonctions à être à support compact dans Ω, on les force à être nulles au voisinage de $\partial\Omega$.

100. Voir l'exercice 5.15 pour une définition du support. Remarquons qu'on peut plus simplement caractériser le fait qu'une fonction g soit à support compact dans $\mathbb{R}^d$ par le fait qu'il existe $R > 0$ tel que $f = 0$ presque partout sur $B(0,R)^c$.

101. On invite le lecteur à penser à $\Omega = \mathbb{R}$ ou $\Omega =]0,1[$.

Remarque 5.40. Attention, l'énoncé peut être considéré un peu ambigu car on affirme la densité d'un espace de fonctions (les fonctions continues à support compact) dans un espace de classes de fonctions. En fait, l'inclusion

$$C_c^0(\Omega) \subset L^p(\Omega)$$

est une notation abusive pour le fait que l'application suivante

$$\begin{array}{ccc} C_c^0(\Omega) & \longrightarrow & L^p(\Omega) \\ f & \longmapsto & [f] \end{array}$$

est une injection : en effet, si $f = g$ presque partout sur Ω et si f et g sont des fonctions continues sur Ω, alors $f = g$ partout car l'ensemble $\{f \neq g\}$ est un ouvert de mesure nulle, et donc est vide.

Remarque 5.41. [*] Combiné au théorème 5.22, on en déduit que $L^p(\Omega)$ est le plus petit espace complet qui contient $(C_c^0(\Omega), \|\cdot\|_p)$. Ainsi ce dernier (qui pouvait être défini seulement en utilisant l'intégrale de Riemann) n'est pas complet, et son complété est $(L^p(\Omega), \|\cdot\|_p)$. Ceci est une justification a posteriori de l'importance de l'intégrale de Lebesgue pour obtenir des espaces de Banach pour les normes de type intégrales.

Pour montrer le théorème 5.38, nous avons besoin d'une propriété de régularité intérieure et extérieure de la mesure de Lebesgue :

Proposition 5.42. *Soit $E \subset \mathbb{R}^d$ un ensemble mesurable. Alors*

$$\lambda(E) = \inf\{\lambda(U) : U \text{ ouvert avec } E \subset U\} = \sup\{\lambda(K) : K \text{ compact avec } K \subset E\}.$$

On admet ce résultat, dont la démonstration est un peu technique, voir par exemple [BP12, page 102] [102].

Démonstration du théorème 5.38 : Soit $f \in L^p(\Omega)$. D'après le théorème 5.37, pour tout $\varepsilon > 0$, il existe une fonction étagée g telle que $\|f - g\|_p \leq \varepsilon$. De plus, on peut écrire $\Omega = \bigcup_{n\in\mathbb{N}} K_n$ où $(K_n)_{n\in\mathbb{N}}$ est une suite exhaustive de compacts (voir la note 53 au Chapitre III) ; alors

$$g\mathbb{1}_{K_n} \xrightarrow[n\to+\infty]{} g \text{ simplement, et } \forall n \in \mathbb{N}, \quad |g\mathbb{1}_{K_n}| \leq |g| \in L^p(\Omega)$$

donc par le lemme 5.21, $g\mathbb{1}_{K_n} \xrightarrow[n\to+\infty]{} g$ en norme $L^p(\Omega)$, donc on peut approcher g dans $L^p(\Omega)$ par une fonction $\widetilde{g}$ étagée à support compact, c'est-à-dire identiquement nulle en dehors d'un compact $K_{n_0} \subset \Omega$.

La fonction $\widetilde{g}$ est combinaison linéaire de fonctions caractéristiques d'ensembles mesurables. Par linéarité, il suffit donc d'approcher $\mathbb{1}_E$ où $E \subset K_{n_0}$. Comme $\lambda(E) \leq \lambda(K_{n_0}) < +\infty$, la proposition 5.42 donne l'existence d'un ouvert U et d'un compact K tels que $K \subset E \subset U$ et $\lambda(U \setminus K) \leq \varepsilon$. Quitte à remplacer U par $U \cap K_{n_0+1}$, on peut supposer $\overline{U}$ compact et inclus dans Ω. On définit alors la fonction $h : \Omega \to [0, 1]$ [103],

$$h(x) = \frac{d(x, \Omega \setminus U)}{d(x, \Omega \setminus U) + d(x, K)}.$$

La fonction h est bien définie puisque $K \cap (\Omega \setminus U) = \emptyset$; elle est continue sur Ω comme somme et quotient de fonctions continues dont le dénominateur ne s'annule pas, elle est à support compact car nulle en dehors de $\overline{U}$, et identiquement égale à 1 sur K. Enfin, si $x \in K$ alors $h(x) = \mathbb{1}_E(x) = 1$, si $x \in \Omega \setminus U$ alors $h(x) = \mathbb{1}_E(x) = 0$ et si $x \in U \setminus K$ alors $|h(x) - \mathbb{1}_E(x)| \leq 1$. Il vient donc

$$\int_I |h - \mathbb{1}_E|^p \, dx = \int_{U\setminus K} |h - \mathbb{1}_E|^p \, dx \leq \lambda(U \setminus K) \leq \varepsilon,$$

ce qui conclut la preuve du résultat. □

102. ** Si (E, d) est un espace métrique, la régularité d'une mesure μ sur $\mathscr{B}(E)$ n'est pas systématique : on trouve dans [BP12, 6.6.3] une preuve que si E est localement compact (i.e. tout point de E admet un voisinage compact) et séparable, et si μ est une mesure de Borel (c'est-à-dire finie sur les compacts), alors μ est régulière. Le cas de λ_d sur $\mathbb{R}^d$ est un cas particulier de ce résultat.

103. On redémontre ici le lemme d'Urysohn, voir la question 5 de l'exercice 1.47 du Chapitre III.

Exercice 5.43. Attention, le résultat est faux pour $p = +\infty$. Quelle est d'ailleurs l'adhérence de $C_c^0(I)$ dans $L^\infty(\Omega)$?

Corollaire 5.44 (Continuité des translations dans $L^p(\mathbb{R}^d)$)**.** *Soit $p \in [1, +\infty[$, et $f \in L^p(\mathbb{R}^d)$. Alors*

$$\lim_{a \to 0} \|\tau_a f - f\|_p = 0.$$

Autrement dit, la fonction

$$\begin{array}{ccc} \mathbb{R} & \longrightarrow & L^p(\mathbb{R}^d) \\ a & \longmapsto & \tau_a f \end{array}$$

est continue.

Démonstration. On suppose $f \in C_c^0(\mathbb{R}^d)$. Alors avec le théorème de Heine, on a que f est uniformément continue sur $\mathbb{R}^d$, donc si $\varepsilon > 0$, il existe $\eta > 0$ tel que $\forall (x, y) \in \mathbb{R}^d$, $|x - y| \leq \eta \Rightarrow |f(x) - f(y)| \leq \varepsilon$. Comme f est à support compact, il existe $K \subset \mathbb{R}^d$ compact tel que $f = 0$ presque partout sur K^c. Alors

$$\forall |a| \leq \min\{1, \eta\},\ \|\tau_a f - f\|_p^p = \int_{K + \overline{B}(0,1)} |f(x + a) - f(x)|^p dx \leq \varepsilon^p \lambda_d(K + \overline{B}(0, 1)),$$

d'où le résultat.

Si maintenant $f \in L^p(\mathbb{R}^d)$, par le théorème 5.38, il existe $g \in C_c^0(\mathbb{R}^d)$ tel que $\|f - g\|_p \leq \varepsilon$. D'après l'étape précédente, si $\varepsilon > 0$, il existe $a_0 > 0$ tel que

$$\forall |a| \leq a_0,\ \ \|\tau_a g - g\|_p \leq \varepsilon.$$

Alors

$$\forall |a| \leq a_0,\ \ \|\tau_a f - f\|_p \leq \|\tau_a f - \tau_a g\|_p + \|\tau_a g - g\|_p + \|g - f\|_p \leq 3\varepsilon.$$

□

Nous allons à présent améliorer le résultat précédent en montrant que les fonctions régulières et à support compact sont denses dans les espaces de Lebesgue pour $p < +\infty$. Si Ω est un ouvert de $\mathbb{R}^d$, on considère $C_c^\infty(\Omega)$ l'ensemble des fonctions $f : \Omega \to \mathbb{K}$ indéfiniment dérivables et à support compact dans Ω.

Théorème 5.45. *Soit $\Omega \subset \mathbb{R}^d$ un ouvert et $1 \leq p < +\infty$. Alors l'espace $C_c^\infty(\Omega)$ est dense dans $L^p(\Omega)$.*

Ce résultat utilise le procédé de régularisation par convolution, qui est important en soit.

Définition 5.46 (Identité approchée, approximation de l'identité, noyau de sommabilité positif)**.** La famille $(\rho_\varepsilon)_{\varepsilon > 0}$ de fonctions mesurables de $\mathbb{R}^d$ dans $\mathbb{R}$ est appelée identité approchée positive si elle satisait :

1. $\forall \varepsilon > 0,\ \ \rho_\varepsilon \geq 0$,
2. $\forall \varepsilon > 0,\ \ \displaystyle\int_{\mathbb{R}^d} \rho_\varepsilon(x) dx = 1$,
3. $\forall \eta > 0,\ \ \displaystyle\lim_{\varepsilon \to 0} \int_{\mathbb{R}^d \setminus B(0,\eta)} \rho_\varepsilon(x) dx = 0$.

Remarque 5.47. Intuitivement, la dernière propriété montre que $(\rho_\varepsilon)_{\varepsilon > 0}$ contient très peu de masse sur $\mathbb{R}^d \setminus B(0, \eta)$ quand $\varepsilon > 0$ est petit ; néanmoins la masse totale reste constante égale à 1, donc la famille $(\rho_\varepsilon)_{\varepsilon > 0}$ se concentre autour de 0.
[*] En fait, on verra au Chapitre VIII que la famille $(\rho_\varepsilon)_{\varepsilon > 0}$ converge vers δ_0 au sens des distributions (en fait, au sens des mesures), qui est intuitivement une fonction nulle sur $\mathbb{R}^d$ mais qui vaut $+\infty$ en 0 ; attention, ceci n'est qu'une intuition, car rigoureusement une telle fonction n'est autre qu'une fonction nulle presque partout. Il faut abandonner l'idée de fonction, et voir δ_0 comme une mesure ou une distribution pour avoir le bon objet.

Remarque 5.48. On aurait pu indicer la famille $(\rho_\varepsilon)_{\varepsilon > 0}$ avec $\varepsilon \to 0$ par $n \in \mathbb{N}$ au lieu de ε, et alors $n \to +\infty$.

Remarque 5.49. Il existe une notion d'identité approchée non nécessairement positive ; dans ce cas, il faut remplacer les hypothèses par

1. $\exists M \in \mathbb{R},\ \ \forall \varepsilon > 0,\ \ \displaystyle\int_{\mathbb{R}^d} |\rho_\varepsilon(x)| dx \leq M,$
2. $\forall \varepsilon > 0,\ \ \displaystyle\int_{\mathbb{R}^d} \rho_\varepsilon(x) dx = 1,$
3. $\forall \eta > 0,\ \ \displaystyle\lim_{\varepsilon \to 0} \int_{\mathbb{R}^d \setminus B(0,\eta)} |\rho_\varepsilon(x)| dx = 0.$

On laisse le lecteur se convaincre que la définition 5.46 implique ces hypothèses. Certaines des propriétés qui suivent restent valables pour cette définition plus générale (les preuves ne sont pas beaucoup plus compliquées), mais pas toutes (par exemple l'équation (5.7) utilise la positivité de $(\rho_\varepsilon)_{\varepsilon>0}$). Ces identités interviennent plus rarement.

Exemple 5.50. L'exemple qu'on décrit ici est un moyen important de créer des identités approchées. Notamment, il peut arriver qu'on n'ait juste besoin de l'existence d'une telle identité approchée, et dans ce cas on peut procéder comme suit : si $\rho : \mathbb{R}^d \to \mathbb{R}$ est une fonction mesurable, positive et d'intégrale 1 (et une telle fonction existe !), alors

$$\forall \varepsilon > 0,\ \ \forall x \in \mathbb{R}^d,\ \ \ \rho_\varepsilon(x) := \frac{1}{\varepsilon^n} \rho\left(\frac{x}{\varepsilon}\right)$$

définit une identité approchée. En effet, $(\rho_\varepsilon)_{\varepsilon>0}$ est bien une famille de fonctions mesurables positives, et par changement de variable

$$\int_{\mathbb{R}^d} \rho_\varepsilon(x) dx = \int_{\mathbb{R}^d} \rho\left(\frac{x}{\varepsilon}\right) \frac{dx}{\varepsilon^n} = \int_{\mathbb{R}^d} \rho(y) dy = 1$$

et pour tout $\eta > 0$,

$$\int_{\mathbb{R}^d \setminus B(0,\eta)} \rho_\varepsilon(x) dx = \int_{\mathbb{R}^d \setminus B(0,\frac{\eta}{\varepsilon})} \rho(y) dy \xrightarrow[\varepsilon \to 0]{} 0$$

où la convergence se justifie avec le théorème de convergence dominée 3.8, car

$$\rho \mathbb{1}_{\mathbb{R}^d \setminus B(0,\frac{\eta}{\varepsilon})} \xrightarrow[\varepsilon \to 0]{} 0 \ \text{ simplement, et } \quad \forall \varepsilon > 0,\ \ |\rho \mathbb{1}_{\mathbb{R}^d B(0,\frac{\eta}{\varepsilon})}| \leq \rho \in L^1(\mathbb{R}^d).$$

Exemple 5.51. L'exemple précédent appliqué à $\alpha(x) = \frac{1}{(2\pi)^{d/2}} e^{-\frac{\|x\|^2}{2}}$ donne

$$\forall \varepsilon > 0,\ \ \forall x \in \mathbb{R}^d,\ \ \alpha_\varepsilon(x) = \frac{1}{(2\pi)^{d/2} \varepsilon^d} e^{-\frac{\|x\|^2}{2\varepsilon^2}}$$

qu'on appelle noyau de Gauss ou noyau de la chaleur.

On va justement utiliser ce procédé dans la définition suivante, mais en ajoutant plus d'hypothèses sur la fonction ρ :

Définition 5.52. [Famille régularisante] Soit $\rho \in C_c^\infty(\mathbb{R}^d)$ telle que $\rho \geq 0$, $\operatorname{supp}(\rho) \subset \overline{B}(0,1)$ et $\displaystyle\int_{\mathbb{R}^d} \rho(x)\, dx = 1$. Pour tout $\varepsilon > 0$, on pose $\rho_\varepsilon(x) = \frac{1}{\varepsilon^d} \rho\left(\frac{x}{\varepsilon}\right)$ de sorte que $\int_{\mathbb{R}^d} \rho_\varepsilon(x)\, dx = 1$ et $\operatorname{supp}(\rho_\varepsilon) \subset \overline{B}(0,\varepsilon)$. On dit que $(\rho_\varepsilon)_{\varepsilon>0}$ est une famille régularisante.

Remarque 5.53. D'après l'exemple 5.50, on voit qu'une famille régularisante est en particulier une identité approchée.

Exemple 5.54. À titre d'exemple, on peut vérifier que la fonction $\rho \colon \mathbb{R}^d \to \mathbb{R}$ définie par [104]

$$\rho(x) = \begin{cases} c e^{-\frac{1}{1-\|x\|^2}} & \text{si } \|x\| < 1, \\ 0 & \text{si } \|x\| \geq 1, \end{cases}$$

avec $c = \left(\displaystyle\int_{B(0,1)} e^{-\frac{1}{1-\|x\|^2}}\, dx\right)^{-1}$, satisfait les propriétés requises ci-dessus.

104. $\|\cdot\|$ désigne ici la norme euclidienne de $\mathbb{R}^d$.

Remarque 5.55. Comme suggéré à la remarque 5.48, on pouvait aussi poser par exemple[105] $\forall n \in \mathbb{N},\ \ \rho_n(x) = n^d \rho(nx)$, et on parle alors de suite régularisante.

Commençons pas un premier lemme qui justifie l'expression d'"identité approchée" :

Lemme 5.56. *Soit* $1 \leq p < +\infty$[106] *et* $f \in L^p(\mathbb{R}^d)$. *Alors*

$$\forall \varepsilon > 0, \quad \|\rho_\varepsilon * f\|_p \leq \|f\|_p, \qquad \text{et} \quad f * \rho_\varepsilon \xrightarrow[\varepsilon \to 0]{} f \ \text{dans } L^p(\mathbb{R}^d)$$

Démonstration. Comme $f \in L^p(\mathbb{R}^d)$, par l'exercice 5.35, $f * \rho_\varepsilon$ est bien définie pour tout $\varepsilon > 0$, et de plus

$$\|f * \rho_\varepsilon\|_p \leq \|f\|_p \|\rho_\varepsilon\|_1 = \|f\|_p. \tag{5.7}$$

Soit $\varepsilon > 0$ et $x \in \mathbb{R}^d$: alors

$$\begin{aligned}
|f * \rho_\varepsilon(x) - f(x)|^p &= \left| \int_{\mathbb{R}^d} \rho_\varepsilon(x) f(x-t)dt - \underbrace{\left(\int_{\mathbb{R}^d} \rho_\varepsilon(t)dt\right)}_{=1} f(x) \right|^p \\
&\leq \left[\int_{\mathbb{R}^d} \rho_\varepsilon(t) \left| f(x-t) - f(x) \right| dt \right]^p \\
&\leq \left[\int_{\mathbb{R}^d} \underbrace{\rho_\varepsilon(t)^{1/p'} \rho_\varepsilon(t)^{1/p}}_{=\rho_\varepsilon(t)} \left| f(x-t) - f(x) \right| dt \right]^p \quad \text{et avec Hölder,} \\
&\leq \left[\int_{\mathbb{R}^d} \rho_\varepsilon(t)^{p'/p'} dx \right]^{p/p'} \left[\int_{\mathbb{R}^d} \rho_\varepsilon(t)^{p/p} \left| f(x-t) - f(x) \right|^p \right]^{p/p} \\
&\leq \int_{\mathbb{R}^d} \rho_\varepsilon(t) \left| f(x-t) - f(x) \right|^p dt.
\end{aligned}$$

[107] On intègre et on applique le théorème de Fubini-Tonelli, pour obtenir :

$$\int_{\mathbb{R}^d} |f * \rho_\varepsilon(x) - f(x)|^p dx \leq \int_{\mathbb{R}^d} \rho_\varepsilon(t) \|\tau_t f - f\|_p^p dt.$$

Soit $\varepsilon' > 0$[108]. Par continuité des translations dans L^p (corollaire 5.44, valable car $p < +\infty$), il existe $\eta > 0$ tel que

$$\forall t \in B(0,\eta), \ \ \|\tau_t f - f\|_p^p \leq \varepsilon'$$

Puis par propriété des identités approchées, il existe ε_0 tel que

$$\forall \varepsilon < \varepsilon_0, \ \ \int_{\mathbb{R}^d \setminus B(0,\eta)} \rho_\varepsilon(t) dt \leq \varepsilon'.$$

Ainsi, on peut écrire :

$$\begin{aligned}
\int_{\mathbb{R}^d} \rho_\varepsilon(t) \|\tau_t f - f\|_p^p dt &= \int_{B(0,\eta)} \rho_\varepsilon(t) \|\tau_t f - f\|_p^p dt + \int_{\mathbb{R}^d \setminus B(0,\eta)} \rho_\varepsilon(t) \|\tau_t f - f\|_p^p dt \\
&\leq \varepsilon' \underbrace{\int_{B(0,\eta)} \rho_\varepsilon(t) dt}_{\leq \int_{\mathbb{R}^d} \rho_\varepsilon(t)dt = 1} + \int_{\mathbb{R}^d \setminus B(0,\eta)} \rho_\varepsilon(t) (\underbrace{\|\tau_t f\| + \|f\|_p}_{=2\|f\|_p})^p \\
&\leq (1 + 2^p \|f\|_p^p) \varepsilon'.
\end{aligned}$$

d'où le résultat. □

105. On peut remplacer dans l'expression de ρ_n le nombre $n \in \mathbb{N}$ par toute suite $(\alpha_n)_{n \in \mathbb{N}}$ qui tend vers $+\infty$.
106. Attention, l'énoncé n'est pas valable pour $p = \infty$; voir l'exercice 5.57 pour une condition suffisante sur f pour avoir la convergence uniforme.
107. Dans le calcul qui précède, au lieu d'utiliser l'inégalité de Hölder, on peut utiliser l'inégalité de Jensen pour la mesure de probabilité $\rho_\varepsilon(x)dx$ et la fonction convexe $\psi : s \mapsto s^p$, ce qui donne

$$\psi\left(\int_{\mathbb{R}^d} |f(x-t) - f(x)| \rho_\varepsilon(t) dt \right) \leq \int_{\mathbb{R}^d} \psi(|f(x-t) - f(t)|) \rho_\varepsilon(t) dt.$$

108. La notation ε' vient du fait que la lettre ε est déjà prise pour paramétrer la famille $(\rho_\varepsilon)_{\varepsilon > 0}$.

Exercice 5.57 (Autres résultats d'approximation)**.** Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction bornée, et $(\rho_\varepsilon)_{\varepsilon>0}$ une identité approchée.

1. Montrer que $(f * \rho_\varepsilon)_{\varepsilon>0}$ est bien définie, et continue sur $\mathbb{R}^d$ [109].
2. Montrer que si f est continue en $x_0 \in \mathbb{R}^d$, alors
$$f * \rho_\varepsilon(x_0) \xrightarrow[\varepsilon\to 0]{} f(x_0).$$
3. (a) On suppose que f est uniformément continue sur $\mathbb{R}^d$: montrer
$$\|f * \rho_\varepsilon - f\|_\infty \xrightarrow[\varepsilon\to 0]{} 0.$$
   (b) Montrer que si $f \in C_c^0(\mathbb{R}^d)$, alors f est uniformément continue sur $\mathbb{R}^d$, et donc le résultat de la question précédente s'applique.

Maintenant que nous avons l'aspect "approximation de l'identité", nous voyons dans le lemme suivant l'aspect régularisation :

Lemme 5.58. *Soit $f \in L^1_{\text{loc}}(\mathbb{R}^d)$, et $(\rho_\varepsilon)_{\varepsilon>0}$ une famille régularisante. Alors pour tout $\varepsilon > 0$, $f * \rho_\varepsilon \in C^\infty(\mathbb{R}^d)$. Si de plus f est à support compact, alors $f * \rho_\varepsilon \in C_c^\infty(\mathbb{R}^d)$.*

Démonstration. Soit $\varepsilon > 0$. D'après l'exercice 5.34, $f * \rho_\varepsilon$ est bien définie. Alors

— $t \mapsto \rho_\varepsilon(x-t)f(t)$ est mesurable pour tout $x \in \mathbb{R}^d$,

— $x \mapsto \rho_\varepsilon(x-t)f(t)$ est C^∞ sur $\mathbb{R}^d$,

— pour tout $x_0 \in \mathbb{R}^d$, tout $k = (k_1, \ldots, k_d) \in \mathbb{N}^d$, et tout $x \in B(x_0, 1)$,
$$\left|\frac{\partial^k}{\partial x^k}(\rho_\varepsilon(x-t)f(t))\right| \leq \left\|\frac{\partial^k \rho_\varepsilon}{\partial x^k}\right\|_{L^\infty(B(x_0,1+\varepsilon))} |f(t)|\mathbb{1}_{B(x_0,1+\varepsilon)} \in L^1(\mathbb{R}^d)$$

donc par le théorème de régularité des intégrales à paramètre (théorème 3.21 et remarque 3.22), on en déduit que $f * \rho_\varepsilon$ est de classe C^∞ sur $B(x_0, 1)$, et comme ceci a été prouvé pour tout $x_0 \in \mathbb{R}^d$, finalement $f * \rho_\varepsilon$ est C^∞ sur $\mathbb{R}^d$.

[110]Si f est à support compact, alors il existe $R \in \mathbb{R}$ tel que $\text{supp}(f) \subset \overline{B}(0,R)$. Si $x \notin \overline{B}(0, R+\varepsilon)$, alors pour tout $t \in B(0,\varepsilon)$, on a $\|x-t\| \geq \|x\| - \|t\| \geq R+\varepsilon-\varepsilon = R$ si bien que $f(x-t) = 0$, et par conséquent,
$$f * \rho_\varepsilon(x) = \int_{\mathbb{R}^d} f(x-t)\rho_\varepsilon(y)\,dy = \int_{B(0,\varepsilon)} f(x-t)\rho_\varepsilon(t)\,dt = 0.$$
Ceci montre que le support de $f*\rho_\varepsilon$ est inclus dans $\overline{B}(0, R+\varepsilon)$, qui est compact. Donc $f \in C_c^\infty(\mathbb{R}^d)$. □

Démonstration du théorème 5.45 : Soit $f \in L^p(\Omega)$ et $\varepsilon' > 0$. On peut écrire $\Omega = \bigcup_{n\in\mathbb{N}} K_n$ où $(K_n)_{n\in\mathbb{N}}$ est une suite exhaustive de compacts. Par convergence dominée (lemme 5.21), on a
$$f\mathbb{1}_{K_n} \xrightarrow[n\to+\infty]{L^p} f$$
donc il existe $n_0 \in \mathbb{N}$ tel que $\|f\mathbb{1}_{K_{n_0}} - f\|_p \leq \varepsilon'$. On note g la fonction $f\mathbb{1}_{K_{n_0}}$ prolongée par 0 sur $\mathbb{R}^d\backslash\Omega$. Par le lemme 5.58, les fonctions $(\rho_\varepsilon * g)_{\varepsilon>0}$ sont C^∞ à support compact (dans $\mathbb{R}^d$). Plus précisément, il existe $\varepsilon_0 > 0$ tel que $K_{n_0} + \overline{B}(0,\varepsilon_0) \subset \Omega$, et alors le support de $g * \rho_\varepsilon$ est compact dans Ω pour $\varepsilon < \varepsilon_0$, par la formule de l'exercice 5.29. De plus, par le lemme 5.56,
$$g * \rho_\varepsilon \xrightarrow[\varepsilon\to 0]{L^p(\mathbb{R}^d)} g$$
donc il existe ε_0 tel que $\|g * \rho_{\varepsilon_0} - g\|_p \leq \varepsilon'$. Par l'inégalité triangulaire, on récupère
$$\|f - g * \rho_{\varepsilon_0}\|_p \leq 2\varepsilon'$$
avec $g * \rho_{\varepsilon_0} \in C_c^\infty(\Omega)$, d'où le résultat. □

109. Ceci explique qu'on ne peut pas avoir en général convergence uniforme de $(f * \rho_\varepsilon)_{\varepsilon>0}$ vers f : en effet si tel était le cas, f serait nécessairement continue comme limite uniforme de fonctions continues.
110. On refait l'exercice 5.29 dans le cas particulier qui nous intéresse ici.

Remarque 5.59. En fait, on peut donner une preuve en apparence différente du théorème 5.45, plus proche de la preuve du théorème 5.38, qui consiste à donner une version C^∞ du lemme d'Urysohn dans ce contexte :

Lemme 5.60 (Urysohn C^∞)**.** *Soit O est un ouvert de $\mathbb{R}^d$, et $K \subset O$ un compact. Alors il existe $\varphi : \mathbb{R}^d \to \mathbb{R}$ de classe C^∞ telle que $\varphi = 1$ sur K, et $\varphi = 0$ sur O^c.*

Voir par exemple [BP12, Théorème 13.9]. On a dit “en apparence” car la preuve de ce lemme repose sur la régularisation par convolution.

Une application fondamentale du théorème 5.45 (ou plutôt de sa preuve) est la suivante :

Corollaire 5.61. *Soit $\Omega \subset \mathbb{R}^d$ un ouvert et $f \in L^1_{\mathrm{loc}}(\Omega)$ telle que*

$$\int_\Omega f\varphi\, dx = 0 \text{ pour tout } \varphi \in C_c^\infty(\Omega). \tag{5.8}$$

Alors $f = 0$ presque partout sur Ω.

Remarque 5.62. Ce résultat est essentiel pour la théorie des distributions : il dit que l'on peut injecter l'espace $L^1_{\mathrm{loc}}(\Omega)$ dans l'espace des distributions $\mathcal{D}'(\Omega)$, autrement dit que la connaissance des formes linéaires $C_c^\infty(\Omega) \ni \varphi \mapsto \int_\Omega f\varphi\, dx$ suffit à retrouver la fonction f presque partout. C'est ce qui justifie le mot “fonction généralisée” quand on parle de distributions, voir le Chapitre VIII.

Démonstration. Soit $K \subset \Omega$ un ensemble compact. Alors on pose g le prolongement de $f\mathbb{1}_K$ par 0 sur $\mathbb{R}^d \setminus \Omega$. Alors $g \in L^1(\mathbb{R}^d)$ et d'après le lemme 5.58, $g * \rho_\varepsilon \xrightarrow[\varepsilon\to 0]{L^1} g$ dans $L^1(\mathbb{R}^d)$. Par conséquent, la réciproque partielle du théorème de convergence dominée (corollaire 5.24) donne l'existence d'une suite $(\varepsilon_n)_{n\in\mathbb{N}}$ qui converge vers 0 et telle que $g * \rho_{\varepsilon_n} \xrightarrow[n\to+\infty]{} g$ presque partout. Soit $x \in \mathring{K}$ tel que cette convergence a lieu : alors d'une part $g * \rho_{\varepsilon_n}(x) \xrightarrow[n\to+\infty]{} g(x) = f(x)$. D'autre part, pour ε assez petit,

$$g * \rho_\varepsilon(x) = \int_{\mathbb{R}^d} g(t)\rho_\varepsilon(x-t)dt = \int_\Omega f(t)\rho_\varepsilon(x-t)dt = 0$$

où on a utilisé que le support de $t \mapsto \rho_\varepsilon(x-t)$ est dans $\overline{B}(x,\varepsilon) \subset K$ si ε petit, et aussi que c'est une fonction C^∞ à support compact dans Ω, et donc qu'on peut l'utiliser comme fonction test dans (5.8). Ainsi $f = 0$ presque partout sur $\mathring{K}$.
Si on considère $(K_n)_{n\in\mathbb{N}}$ une suite exhaustive de compacts pour Ω, alors on a bien $\Omega = \bigcup_{n\in\mathbb{N}} \mathring{K}_n$, et $f = 0$ presque partout sur chaque K_n d'après ce qui précède, ce qui donne le résultat puisqu'une union dénombrable d'ensembles de mesure nulle est de mesure nulle. □

Exercice 5.63. Soit Ω un ouvert de $\mathbb{R}^d$.

1. Pour $1 \le p < +\infty$, montrer que l'espace $L^p(I)$ est séparable.
2. [*] On suppose Ω non vide. Montrer que l'espace $L^\infty(I)$ n'est pas séparable.

6 **Compléments de théorie de la mesure

On propose ici quelques compléments sur la théorie de la mesure :

6.1 *Théorème de la classe monotone et résultats d'unicité

Définition 6.1. Une classe $\mathscr{M}$ de parties d'un ensemble E est appelée *classe monotone* si

(i) $E \in \mathscr{M}$;

(ii) elle est stable par *différence propre* : pour tout $(A,B) \in \mathscr{M}^2$, $A \subset B \Rightarrow B \setminus A \in \mathscr{M}$;

(iii) elle est stable par *réunion dénombrable croissante* : si $(A_n)_{n\in\mathbb{N}}$ est une suite croissante d'éléments de $\mathscr{M}$, alors $\cup_{n\in\mathbb{N}} A_n \in \mathscr{M}$.

Proposition 6.2. *a) L'intersection d'une collection quelconque non vide de classes monotones est une classe monotone.*

b) Pour toute classe $\mathscr{C}$ de parties de E, l'intersection[111] *de toutes les classes monotones contenant tous les éléments de $\mathscr{C}$ est donc une classe monotone, noté $\mathscr{M}(\mathscr{C})$, et appelée classe monotone engendrée par $\mathscr{C}$ ou plus petite classe monotone contenant $\mathscr{C}$.*

Les preuves sont laissées en exercice.

Proposition 6.3. *a) Une tribu est une classe monotone.*

b) Une classe monotone stable par intersections finies est une tribu.

Démonstration. On laisse a) en exercice, et on montre b) : soit $\mathscr{M}$ une telle classe monotone et vérifions les trois propriétés caractéristiques des tribus.

(i) $E \in \mathscr{M}$ puisque $\mathscr{M}$ est une classe monotone.

(ii) Comme $E \in \mathscr{M}$, pour tout $A \in \mathscr{M}$, le complémentaire de A est la différence propre $E \setminus A$, qui appartient donc à $\mathscr{M}$.

(iii) Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathscr{M}$. Pour tout $n \in \mathbb{N}$, posons $B_n := \cup_{k=0}^n A_k$. Comme $\cup_{n\in\mathbb{N}} A_n = \cup_{n\in\mathbb{N}} B_n$ et que $(B_n)_{n\in\mathbb{N}}$ est une suite croissante, la propriété (iii) des classes monotones assure qu'il suffit de montrer que $B_n \in \mathscr{M}$ pour tout $n \in \mathbb{N}$. Autrement dit, il suffit de montrer que $\mathscr{M}$ est stable par réunion finie. Or $\mathscr{M}$ est stable par passage au complémentaire d'après (ii) et, par hypothèse, stable par intersection finie : ainsi $A \cup B = (A^c \cap B^c)^c \in \mathscr{M}$ pour tous $A, B \in \mathscr{M}$, ce qui assure que $\mathscr{M}$ est stable par réunion finie. □

Théorème 6.4 (Théorème de la classe monotone)**.** *Si $\mathscr{C} \subset \mathscr{P}(E)$ est stable par intersection finie, alors $\mathscr{M}(\mathscr{C}) = \sigma(\mathscr{C})$.*

Démonstration. De manière générale, comme $\sigma(\mathscr{C})$ est une tribu contenant $\mathscr{C}$, c'est une classe monotone contenant $\mathscr{C}$, et donc contenant $\mathscr{M}(\mathscr{C})$ puisque $\mathscr{M}(\mathscr{C})$ est la plus petite classe monotone contenant $\mathscr{C}$.

Supposons à présent avoir montré que $\mathscr{M}(\mathscr{C})$ est une tribu. Alors $\mathscr{M}(\mathscr{C})$ est une tribu contenant $\mathscr{C}$, donc contenant $\sigma(\mathscr{C})$ puisque $\sigma(\mathscr{C})$ est la plus petite tribu contenant $\mathscr{C}$.

Il reste donc à montrer que $\mathscr{M}(\mathscr{C})$ est une tribu. D'après la proposition qui précède, il suffit de montrer que $\mathscr{M}(\mathscr{C})$ est stable par intersection finie, en utilisant le fait que $\mathscr{C}$ l'est.

1. Soit $\mathscr{M}_1 := \{A \in \mathscr{M}(\mathscr{C}); A \cap C \in \mathscr{M}(\mathscr{C})\ \forall C \in \mathscr{C}\}$ et montrons que $\mathscr{M}_1 = \mathscr{M}(\mathscr{C})$. C'est une conséquence des trois points suivants :

1.1. $\mathscr{M}_1 \subset \mathscr{M}(\mathscr{C})$ par définition.

1.2. $\mathscr{M}_1$ est une classe monotone, comme on le voit en vérifiant les trois propriétés caractéristiques :

(i) $E \in \mathscr{M}_1$ puisque pour tout $C \in \mathscr{C}$ on a $E \cap C = C \in \mathscr{C} \subset \mathscr{M}(\mathscr{C})$;

(ii) soit $A, B \in \mathscr{M}_1$ avec $A \subset B$ et montrons que $B \setminus A \in \mathscr{M}_1$. En effet, pour tout $C \in \mathscr{C}$, $(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C)$ appartient à la classe monotone $\mathscr{M}(\mathscr{C})$ comme différence de deux éléments de $\mathscr{M}(\mathscr{C})$, inclus l'un dans l'autre ;

(iii) soit $(A_n) \in \mathscr{M}_1$ croissante et montrons que $\cup_n A_n \in \mathscr{M}_1$. En effet, pour tout $C \in \mathscr{C}$, $(\cup_n A_n) \cap C = \cup_n (A_n \cap C)$ appartient à la classe monotone $\mathscr{M}(\mathscr{C})$ comme union croissante d'éléments de $\mathscr{M}(\mathscr{C})$.

1.3. $\mathscr{M}_1$ contient $\mathscr{C}$. En effet, si $A \in \mathscr{C}$, alors pour tout $C \in \mathscr{C}$ on a $A \cap C \in \mathscr{C}$ car $\mathscr{C}$ est stable par intersections finies. Donc $A \cap C \in \mathscr{M}(\mathscr{C})$, puis $A \in \mathscr{M}_1$.

2. On montre de même que $\mathscr{M}_2 = \{A \in \mathscr{M}(\mathscr{C}); A \cap C \in \mathscr{M}(\mathscr{C})\ \forall C \in \mathscr{M}(\mathscr{C})\}$ est une classe monotone, puis est égale à $\mathscr{M}(\mathscr{C})$.

3. Autrement dit, pour tout $A \in \mathscr{M}(\mathscr{C})$, on a $A \in \mathscr{M}_2$, c'est-à-dire que pour tout $C \in \mathscr{M}(\mathscr{C})$ on a $A \cap C \in \mathscr{M}(\mathscr{C})$. Ainsi $\mathscr{M}(\mathscr{C})$ est stable par intersection de deux éléments, puis par intersections finies, ce que l'on cherchait à montrer. □

On en déduit les deux résultats d'unicité suivants :

111. non vide puisque $\mathscr{P}(E)$ est une classe monotone

Corollaire 6.5. *Soient μ et ν deux mesures finies sur un espace mesurable $(E, \mathscr{A})$, telles que $\mu(E) = \nu(E)$ et qui coïncident*[112] *sur une classe $\mathscr{C} \subset \mathscr{A}$ stable par intersections finies et engendrant*[113] *$\mathscr{A}$. Alors μ et ν coïncident sur $\mathscr{A}$.*

Corollaire 6.6. *Soient μ et ν deux mesures σ-finies sur un espace mesurable $(E, \mathscr{A})$ telles que :*

a) il existe une suite croissante (E_n) d'ensembles mesurables telle que $\cup_n E_n = E$;

b) pour tout entier n, $\mu(E_n) = \nu(E_n) < \infty$;

c) μ et ν coïncident sur une classe $\mathscr{C} \subset \mathscr{A}$ stable par intersections finies, engendrant $\mathscr{A}$ et contenant chaque E_n.

Alors μ et ν coïncident sur $\mathscr{A}$.

Remarque 6.7. Le fait que μ et ν sont σ-finies est une conséquence des conditions a) et b).

Démonstration du Corollaire 6.5. Posons $\mathscr{M} := \{A \in \mathscr{A} : \mu(A) = \nu(A)\}$ et montrons que $\mathscr{M} = \mathscr{A}$.

$\mathscr{M}$ contient $\mathscr{C}$ par hypothèse et est une classe monotone comme on le voit en vérifiant les trois propriétés caractéristiques. Par conséquent $\mathscr{M}$ contient la classe monotone engendrée $\mathscr{M}(\mathscr{C})$. Or $\mathscr{C}$ est stable par intersections finies, donc $\mathscr{M}(\mathscr{C}) = \sigma(\mathscr{C})$ par le théorème de la classe monotone, qui est $\mathscr{A}$ par hypothèse. Ainsi $\mathscr{M}$ contient $\mathscr{A}$, et est donc égal à $\mathscr{A}$. □

Démonstration du Corollaire 6.6. On applique le corollaire 6.5 aux mesures traces $\mu_n := \mu(\cdot \cap E_n)$ et $\nu_n := \nu(\cdot \cap E_n)$ qui sont finies gr,ce à l'hypothèse (b). Elles coïncident bien sur $\mathscr{C}$ par l'hypothèse (c) : en effet, pour tout $C \in \mathscr{C}$, on a $C \cap E_n \in \mathscr{C}$ car $E_n \in \mathscr{C}$ et $\mathscr{C}$ est stable par intersection, puis

$$\mu_n(C) = \mu(C \cap E_n) = \nu(C \cap E_n) = \nu_n(C).$$

Donc μ_n et ν_n coïncident sur $\mathscr{A}$ d'après le corollaire 6.5 . Maintenant l'hypothèse (a) permet de conclure en utilisant la continuité à gauche de la mesure, car pour tout $A \in \mathscr{A}$,

$$\mu(A) = \mu\Big(\cup_n (E_n \cap A)\Big) = \mu\Big(\lim_{n\to\infty} (E_n \cap A)\Big) = \lim_{n\to\infty} \mu(E_n \cap A) = \lim_{n\to\infty} \mu_n(A),$$

et de même pour ν. Or $\mu_n(A) = \nu_n(A)$, donc

$$\mu(A) = \lim_{n\to\infty} \mu_n(A) = \lim_{n\to\infty} \nu_n(A) = \nu(A),$$

ce qui montre que μ et ν coïncident sur $\mathscr{A}$. □

Unicité de la mesure de Lebesgue : Supposons qu'il existe deux mesures μ et ν sur $\mathscr{B}(\mathbb{R}^d)$ telles que pour tout pavé ouvert $R = \prod_{k=1}^d]a_k, b_k[$, avec $-\infty \le a_k \le b_k \le +\infty$ on ait

$$\mu(R) = \prod_{k=1}^d (b_k - a_k) = \nu(R),$$

avec la convention habituelle $0 \times \infty = 0$. Montrons qu'alors μ et ν coïncident sur $\mathscr{B}(\mathbb{R}^d)$. Ceci prouvera l'unicité de la mesure de Lebesgue (dont nous montrerons l'existence à la section suivante).

Soit $\mathscr{C}$ l'ensemble des pavés ouverts de $\mathbb{R}^d$ (produits d'intervalles ouverts pouvant être infinis, donc en particulier pouvant être égaux à $\mathbb{R}$ tout entier). Alors $\mathscr{C}$ contient $\mathbb{R}^d$, est stable par intersections finies et est de tribu engendrée égale à $\mathscr{B}(\mathbb{R}^d)$. Soit E_n le produit des intervalles $]-n, n[$, c'est-à-dire $E_n := \prod_{k=1}^d]-n, n[$. Alors les propriétés du corollaire 6.6 sont bien vérifiées : a) $\cup_n E_n = \mathbb{R}^d$; b) pour tout entier n, $\mu(E_n) = \nu(E_n) = (2n)^d < \infty$; c) $E_n \in \mathscr{C}$, et $\sigma(\mathscr{C}) = \mathscr{B}(\mathbb{R}^d)$. On peut donc conclure que μ et ν coïncident sur $\mathscr{B}(\mathbb{R}^d)$.

112. c'est-à-dire que pour tout $A \in \mathscr{C}$, $\mu(A) = \nu(A)$
113. c'est-à-dire que $\mathscr{A} = \sigma(\mathscr{C})$

Caractérisation d'une mesure par sa fonction de répartition :

Définition 6.8. Si μ est une mesure finie sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, on appelle *fonction de répartition* de μ la fonction $F : \mathbb{R} \to \mathbb{R}_+$ définie par $F(x) := \mu(]-\infty, x])$.

Proposition 6.9. *La fonction de répartition F d'une mesure finie est continue à droite, croissante, et vérifie*

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{et} \quad \lim_{x\to+\infty} F(x) = \mu(\mathbb{R}).$$

De plus, pour tous réels $a < b$

$$\begin{array}{llll} i) & \mu(]a,b]) = F(b) - F(a) & ii) & \mu([a,b]) = F(b) - F(a-) \\ iii) & \mu(]a,b[) = F(b-) - F(a) & iv) & \mu([a,b[) = F(b-) - F(a-). \end{array}$$

Dans cette proposition on a noté $F(b-) = \lim_{x\to b, x<b} F(x)$.

Exemple 6.10. Si $\mu = \delta_a$, alors $F = \mathbb{1}_{[a,+\infty[}$. De manière générale, si $\mu = \sum_{n\in\mathbb{N}} \alpha_n \delta_{x_n}$, alors F est discontinue en tout point x_n tel que $\alpha_n > 0$ et continue partout ailleurs

$$F(x) = \sum_n \alpha_n \mathbb{1}_{[x_n,+\infty[}.$$

Théorème 6.11. *Deux mesures finies sur $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ de même fonction de répartition sont égales.*

Démonstration. Soit μ et ν ces deux mesures, et soit

$$\mathscr{C} := \{]-\infty, x] : x \in \mathbb{R}\} \cup \{\mathbb{R}\}.$$

Alors $\mathscr{C}$ est stable par intersection finie et engendre $\mathscr{B}(\mathbb{R})$. De plus μ et ν coïncident sur $\mathscr{C}$ car $\mu(]-\infty, x]) = \nu(]-\infty, x])$ pour tout x, l'égalité $\mu(\mathbb{R}) = \nu(\mathbb{R})$ s'obtenant par passage à la limite. Le corollaire 6.5 permet de conclure que μ et ν coïncident sur $\mathscr{B}(\mathbb{R})$. □

6.2 **Construction de mesures : le théorème de Carathéodory

Définition 6.12. Une classe $\mathscr{B}$ de parties d'un ensemble E est appelée *algèbre* ou *algèbre de Boole* si

(i) elle contient E : $E \in \mathscr{B}$;

(ii) elle est stable par *passage au complémentaire* : $A^c \in \mathscr{B}$ pour tout $A \subset E, A \in \mathscr{B}$;

(iii) elle est stable par *réunions finies* : $A \cup B \in \mathscr{B}$ pour tous $A, B \in \mathscr{B}$.

Remarque 6.13. Une tribu est une algèbre de Boole stable par réunion dénombrable, d'où le nom de σ-algèbre.

Remarque 6.14. Dans $\mathbb{R}^d$, l'ensemble des réunions finies de pavés forment une algèbre, ainsi que l'ensemble des réunions finies de pavés disjoints.

Définition 6.15. Soit $\mathscr{B}$ une algèbre de Boole sur un ensemble E. Une *mesure d'algèbre* sur $(E, \mathscr{B})$ est une application $m : \mathscr{B} \to [0, +\infty]$ qui :

(i) associe la valeur 0 à l'ensemble vide : $m(\emptyset) = 0$;

(ii) est *finiment additive* : pour tous $A, B \in \mathscr{B}$ tels que $A \cap B = \emptyset$, $m(A \cup B) = m(A) + m(B)$;

(iii) satisfait la popriété suivante : il existe une suite croissante (E_n) d'éléments de $\mathscr{B}$ convergeant vers E telle que $m(E_n) < \infty$ pour chaque entier n et telle que pour tout $A \in \mathscr{B}$, $\lim_{n\to\infty} m(A \cap E_n) = m(A)$;

(iv) satisfait la *propriété de Carathéodory* : pour toute suite décroissante (A_n) d'éléments de $\mathscr{B}$ convergeant vers $\emptyset$ et telle que $m(A_0) < \infty$, $\lim_n \downarrow m(A_n) = 0$.

Proposition 6.16. *Une mesure d'algèbre m sur $(E, \mathscr{B})$ vérifie pour tous $A, B \in \mathscr{B}$:*

(i) Additivité finie : $m(A) = m(A \setminus B) + m(A \cap B)$;

(ii) Additivité forte : $m(A \cup B) + m(A \cap B) = m(A) + m(B)$ *;*

(iii) Sous-additivité : $m(A \cup B) \leq m(A) + m(B)$ *;*

(iv) Croissance : si $A \subset B$, $m(A) \leq m(B)$.

Théorème 6.17 (de prolongement de Carathéodory)**.** *Soit* $\mathscr{B}$ *une algèbre de Boole sur un ensemble* E *et* m *une mesure d'algèbre sur* $(E, \mathscr{B})$*. Alors il existe une mesure* μ *sur la tribu* $\sigma(\mathscr{B})$ *qui coïncide*[114] *avec* m *sur* $\mathscr{B}$*.*

On dit alors que μ est un prolongement de la mesure (d'algèbre) m, qui elle est seulement définie sur l'algèbre $\mathscr{B}$, à la tribu $\sigma(\mathscr{B})$.

Voir par exemple [BP12, Section 6.5] pour une démonstration.

Remarque 6.18. Le prolongement construit dans le théorème est en fait *unique*. En effet, si μ et ν sont deux prolongements d'une même mesure d'algèbre $\mathscr{B}$, alors μ et ν coïncident sur $\mathscr{B}$, qui contient E et est stable par intersections finies. Comme de plus il existe une suite mesurable (E_n) convergeant vers E telle que $\mu(E_n) = \nu(E_n) < \infty$, alors μ et ν coïncident sur $\sigma(\mathscr{B})$ d'après le corollaire 6.6.

Application : existence de la mesure de Lebesgue Montrons l'existence d'une mesure sur $\mathscr{B}(\mathbb{R}^d)$ telle que la mesure d'un pavé $R = \prod_{k=1}^d I_k$, où chaque I_k est un intervalle de $\mathbb{R}$ d'extrémité gauche $a_k \geq -\infty$ et d'extrémité droite $b_k \leq +\infty$ (les extrémités pouvant être fermées ou ouvertes), vaut $\prod_{k=1}^d (b_k - a_k)$. On définit $\mathscr{B}$ l'ensemble des réunions finies de pavés deux à deux disjoints. Alors pour tout $A \in \mathscr{B}$, A s'écrit de manière unique[115] sous la forme $A = \cup_{i=1}^j R_i$, où les (R_i) sont des pavés deux à deux disjoints, et l'on peut définir sans ambiguïté la mesure d'algèbre m sur $\mathscr{B}$ par $m(A) = \sum_{i=1}^j m(R_i)$, où la mesure d'un pavé a été définie précédemment. On peut alors vérifier que $\mathscr{B}$ est une algèbre et que m est une mesure d'algèbre sur $(E, \mathscr{B})$ avec E_n définie comme le produit des intervalles $]-n, n[$. Comme $\sigma(\mathscr{B}) = \mathscr{B}(\mathbb{R}^d)$, le théorème de Carathéodory permet bien de déduire l'existence d'une mesure, appelée mesure de Lebesgue, prolongeant la mesure m à tous les boréliens de $\mathbb{R}^d$.

Remarque 6.19. Si $F : \mathbb{R} \to \mathbb{R}_+$ est une fonction croissante, bornée, continue à droite et telle que $\lim_{x \to -\infty} F(x) = 0$, on peut déduire du théorème de Carathéodory l'existence d'une (unique) mesure μ sur $\mathscr{B}(\mathbb{R})$ qui admet F pour fonction de répartition.

114. c'est-à-dire que pour tout $B \in \mathscr{B}$, $\mu(B) = m(B)$

115. Cette écriture n'est unique que si elle est supposée minimale, c'est-à-dire utilisant un minimum de pavés ; de plus, si les pavés ne sont plus supposés disjoints, il n'existe même plus de décomposition minimale unique, et il faut donc montrer que la définition de m ne dépend pas de la décomposition choisie

Chapitre V

Probabilités

Préambule : quelques références

Cette partie traite de la théorie des probabilités. Les résultats de théorie de la mesure ou d'intégration seront rappelés au moment de leur utilisation. Voici les trois références sur lesquelles se basent essentiellement ce cours :

[App15] W. Appel, Probabilités pour les non-probabilistes, H&K (2015).
Très clair, beaucoup d'exemples intéressants, extrêmement complet malgré ce que le titre pourrait laisser penser.

[GK19] O. Garet, A. Kurtzman, De l'intégration aux probabilités 2nd édition, Ellipses (2019).
Très bien écrit, extrêmement complet, précision remarquable.

[Ouv09] J.-Y. Ouvrard, Probabilités, Tome 2, Cassini (2009).
Un peu moins clair que les deux références précédentes (plus aride, moins d'exemples intéressants à mon goût), mais contient des résultats qui ne sont pas ailleurs.

Autres références qui peuvent être utiles, pour compléter les précédentes.

[BCD21] Q. Berger, F. Caravenna, P. Dai Pra, Introduction aux probabilités : modèles et applications, Dunod (2021).

[Bil12] P. Billingsley, Probability and Measure (anniversary edition), Willey (2012). *Un grand classique, à raison.*

[FFF12] D. Foata, J. Franchi, A. Fuchs, Calcul des probabilités, Dunod (2012).

1 Espaces probabilisés

On souhaite donner un formalisme mathématique à l'idée de procéder à une *expérience aléatoire*, c'est-à-dire à une expérience dont on ne peut pas prédire à coup sûr le résultat. En voici quelques exemples classiques : on lance un dé ordinaire à 6 faces ; on lance une pièce équilibrée jusqu'à ce que l'on obtienne Pile ; on tire au hasard un nombre dans $[0, 1]$; on effectue une suite *infinie* de Pile ou Face.

1.1 Définitions, premières propriétés

Pour décrire une expérience aléatoire, le formalisme moderne consiste à utiliser l'axiomatique de Kolmogorov, introduite dans les années 1930 et basée sur un espace mesuré $(\Omega, \mathcal{F}, \mathbb{P})$. Le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ est appelé *espace probabilisé* — il s'agit d'un espace mesuré dont la mesure est de masse totale égale à 1 — et n'est pas forcément canonique : il relève d'un *choix de modélisation*. Les trois ingrédients $\Omega, \mathcal{F}, \mathbb{P}$ ont l'interprétation suivante :

L'espace d'états, Ω. Aussi appelé *espace des issues* ou *univers*. C'est un ensemble dont les éléments représentent tous les résultats possibles de l'expérience.

L'ensemble des *événements* $\mathcal{F}$. de manière informelle, correspondent à des affirmations sur le résultat de l'expérience aléatoire. De manière générale, on considérera un ensemble d'événements $\mathcal{F} \subset \mathcal{P}(\Omega)$ *non vide*, qui a la propriété d'être une *tribu* de sous-ensembles de Ω.

Définition 1.1 (Tribu). On dit que $\mathcal{F}$ est une *tribu* sur Ω (ou une σ-algèbre), si $\mathcal{F}$ est une partie *non vide* de $\mathcal{P}(\Omega)$ vérifiant :

(i) $\Omega \in \mathcal{F}$;

(ii) si $A \in \mathcal{F}$ alors $A^c \in \mathcal{F}$;

(iii) si $(A_i)_{i\in\mathbb{N}}$ est une suite d'événements de $\mathcal{F}$, alors $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$.

Remarque 1.2. Si Ω est dénombrable, on prendra en général $\mathcal{F} = \mathcal{P}(\Omega)$. Si Ω est infini non dénombrable (comme dans les troisièmes et quatrièmes exemple un peu plus haut), alors on pourra considérer des événements *de base*, à partir desquels on construit l'ensemble $\mathcal{F}$ (la tribu engendrée par des événements de base).

Une mesure de probabilité $\mathbb{P}$. À chaque événement (c'est-à-dire un élément de $\mathcal{F}$), associe un nombre dans $[0,1]$, correspondant à son degré de confiance. C'est une *mesure*, de masse totale égale à 1.

Définition-Proposition 1.3 (Axiomes des probabilités). Soit Ω un ensemble (non vide) muni d'une tribu $\mathcal{F}$ (autrement dit $(\Omega, \mathcal{F})$ est un espace mesurable). Une fonction $\mathbb{P} : \mathcal{F} \to [0,1]$ est appelée *probabilité* si elle vérifie les propriétés suivantes :

(A1) $\mathbb{P}(\Omega) = 1$.

(A2) (*σ-additivité*) Pour toute suite $(A_n)_{n\geq 1}$ d'événements de $\mathcal{F}$ disjoints (c'est-à-dire $A_n \cap A_m = \emptyset$ pour $n \neq m$) on a

$$\mathbb{P}\Big(\bigcup_{n=1}^{+\infty} A_n\Big) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n)\,.$$

Remarque 1.4. Si Ω est dénombrable, une probabilité $\mathbb{P}$ sur $(\Omega, \mathcal{F})$ est caractérisée par ses valeurs sur les singletons, c'est-à-dire par $(\mathbb{P}(\{\omega\}))_{\omega\in\Omega}$: il suffit de voir que par σ-additivité, on doit avoir

$$\mathbb{P}(A) = \sum_{\omega\in A} \mathbb{P}(\{\omega\})\,.$$

La condition pour que $(\mathbb{P}(\{\omega\}))_{\omega\in\Omega}$ définisse une probabilité est que l'on doit avoir $\sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) = 1$.

Exemple 1.5. Reprenons les exemples cités plus haut :

(a) Lancer d'un dé ordinaire à 6 faces.
Choix standard : $\Omega = \{1,2,3,4,5,6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}$ est la mesure uniforme : $\mathbb{P}(A) = \frac{|A|}{6}$. Exemple d'événement : $A =$« le dé donne un nombre impair »$= \{1,3,5\}$.

(b) Lancer d'une pièce équilibrée jusqu'à ce que l'on obtienne Pile.
Choix 1 : $\Omega = \mathbb{N}^*$ (i représente l'indice du lancer qui donne Pile), $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}(\{i\}) = 2^{-i}$. Exemple d'événement : $A =$« on n'obtient pas Pile lors des n premiers lancers »$= \{n+1, n+2, \dots\}$.
Choix 2 : $\Omega = \mathbb{N}$ (i représente le nombre de lancers avant d'obtenir Pile), $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}(\{i\}) = 2^{-(i+1)}$. Exemple d'événement : $A =$« on n'obtient pas Pile lors des n premiers lancers »$= \{n, n+1, \dots\}$.
Choix 3 : l'espace probabilisé de l'exemple (d) ci-dessous.

(c) tirage d'un nombre dans $[0,1]$;
Choix standard : $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}\mathrm{or}([0,1])$ (la tribu engendrée par les intervalles $]a,b[$), $\mathbb{P}$ la mesure de Lebesgue : $\mathbb{P}(]a,b[) = b-a$. Exemple d'événement : $A =$« le nombre est supérieur à e^{-1} »$= [e^{-1}, 1]$.

(d) tirage d'une suite *infinie* de Pile ou Face ;
$\Omega = \{0,1\}^{\mathbb{N}}$, $\mathcal{F} = \sigma(A_i, i \in \mathbb{N})$ la tribu engendrée par les événements $A_i =$« le i-ème lancer est Pile »$= \{\omega \in \Omega, \omega_i = 1\}$, $\mathbb{P}$ la probabilité telle que les événements $(\bigcap_{i\in I} A_i) \cap (\bigcap_{i\in J} A_i^c)$ soient de probabilité $2^{-|I|+|J|}$ si les ensembles d'indices I, J sont disjoints (puis étendue à tout événement de $\mathcal{F}$...) On verra une autre construction plus loin, page 307.

De la Définition 1.3, on déduit facilement les propriétés suivantes des mesures de probabilité.

Proposition 1.6. *Soit* $(\Omega, \mathcal{F}, \mathbb{P})$ *un espace probabilisé. On a les propriétés de base, pour* $A, B \in \mathcal{F}$ *:*

(i) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, $\mathbb{P}(\emptyset) = 0$ *;*

(ii) Si $A \subseteq B$*, alors* $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ *et en particulier* $\mathbb{P}(A) \leq \mathbb{P}(B)$ *;*

(iii) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ *; en particulier* $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.

Démonstration. Laissée en exercice. □

Une propriété qui peut se déduire (par récurrence – exercice) de (iii) ci-dessus est la *formule d'inclusion-exclusion*, aussi appelée *formule de Poincaré*.

Proposition 1.7 (formule d'inclusion-exclusion). *Soit* $n \in \mathbb{N}^*$ *et soient* $A_1, \ldots, A_n$ *des événements de* $\mathcal{F}$*. On a la formule suivante :*

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{I \subset \{1,\ldots,n\},\, |I|=k} \mathbb{P}\Big(\bigcap_{i \in I} A_i\Big).$$

Cette formule est en réalité souvent difficile à manipuler et donne tout de suite un côté technique aux calculs qu'il n'est pas toujours facile de maîtriser. On lui préféra en général d'autres techniques, mais donnons trois exemples marquants où cette formule produit des résultats intéressants.

Exemple 1.8. La formule d'inclusion-exclusion permet de calculer la probabilité p_n que deux entiers pris au hasard dans $\{1, \ldots, n\}$ soient premiers entre eux (références : [BCD21, p. 20]). En effet, on peut écrire l'événement « les deux entiers ne sont pas premiers entre eux » comme l'union des événements (non disjoints) $A_q =$« q divise les deux entiers », pour $q \leq n$ premier. On finit par obtenir la formule exacte

$$p_n = \sum_{d=1}^{n} \mu(d) \frac{\lfloor n/d \rfloor^2}{n^2},$$

où $\lfloor x \rfloor$ est la partie entière et $\mu(d)$ est la fonction de Möbius en arithmétique : $\mu(d) = (-1)^k$ si d possède k facteurs premiers distincts et $\mu(d) = 0$ si d possède un facteur premier de multiplicité supérieure à 2. Ce qui est frappant est que l'on peut ensuite calculer $\lim_{n \to +\infty} p_n = \frac{6}{\pi^2}$.

Exemple 1.9. Problème des chapeaux ou du père Noël secret ou des dérangements ou du point fixe d'une permutation (références : [GK19, p. 57], [BCD21, p. 74]). Un groupe de n personnes font un père Noël secret : chacune écrit son nom sur un bout de papier, on met les bouts de papier dans un bol et chacune tire un nom au hasard — cela correspond à tirer une permutation σ aléatoire, $\sigma(i)$ étant le numéro tiré par la personne i. Autrement dit $\Omega = \mathfrak{S}_n$, $\mathcal{F} = \mathcal{P}(\Omega)$, $\mathbb{P}$ est la probabilité uniforme sur Ω. On note p_n la probabilité de l'événement $A =$« une personne au moins tire son propre nom », que l'on peut écrire comme une union des événements (non disjoints) $A_i =$« la personne i tire son propre nom ». La formule d'inclusion-exclusion permet d'obtenir

$$p_n = \sum_{k=1}^{n} (-1)^{k+1} \frac{1}{k!},$$

et notamment $\lim_{n \to +\infty} p_n = 1 - e^{-1}$.

Exemple 1.10. Problème du collectionneur de vignettes (références : [App15, p. 163] [BCD21, p. 194]). Un collectionneur essaie de compléter une collection de N vignettes : il achète chaque semaine un paquet, la vignette récupérée possédant une chance $\frac{1}{N}$ d'être la vignette k, pour $k \in \{1, \ldots, N\}$. La question est de savoir combien de temps le collectionneur mettra pour compléter sa collection. On peut en fait calculer explicitement la probabilité $p(m)$ que la collection ne soit pas finie après m semaines, en écrivant l'événement en question comme l'union des événements $A_i =$« la vignette i n'a pas été récupérée après m semaines ». La formule exacte est donnée par

$$p(m) = \sum_{k=1}^{N} (-1)^{k+1} \binom{N}{k} \Big(1 - \frac{k}{N}\Big)^m.$$

On peut alors calculer explicitement la limite suivante : pour tout $x \in \mathbb{R}$, $\lim_{N\to+\infty} p(N\ln N + xN) = 1 - e^{-e^{-x}}$. Ainsi, la probabilité que la collection ne soit pas finie après $N \ln N + xN$ semaines est proche de 1 si x est très négatif ($x \to -\infty$) et proche de 0 si x est très grand ($x \to +\infty$). Autrement dit, selon toute vraissemblance, la collection sera finie après un nombre de semaines de l'ordre de $N\ln N + O(N)$ (on peut donner un sens plus précis à cette phrase, voir l'Exemple 5.10).

Voici maintenant une propriété fondamentale, qui utilise la σ-additivité (et pas simplement l'additivité finie).

Proposition 1.11 (**Continuité par le haut/par le bas**). *(i) Si $(A_k)_{k\geq 0}$ est une suite* croissante *d'événements de $\mathcal{F}$, c'est-à-dire telle que $A_{k+1} \supseteq A_k$ pour tout $k \geq 0$, alors*

$$\mathbb{P}\Big(\bigcup_{k=0}^{\infty} A_k\Big) = \lim_{n\to+\infty} \mathbb{P}(A_n)\,.$$

(ii) Si $(A_k)_{k\geq 0}$ est une suite décroissante *d'événements de $\mathcal{F}$, c'est-à-dire telle que $A_{k+1} \subseteq A_k$ pour tout $k \geq 0$, alors*

$$\mathbb{P}\Big(\bigcap_{k=0}^{\infty} A_k\Big) = \lim_{n\to+\infty} \mathbb{P}(A_n)\,.$$

Démonstration. On peut montrer le (ii) à partir du (i) en passant au complémentaire. Concentrons-nous sur le (i).

Posons $B_0 = A_0$ et pour $k \geq 1$ $B_k = A_k \setminus A_{k-1}$. Ainsi les $(B_k)_{k\in\mathbb{N}}$ sont des événements disjoints et vérifient $A_n = \bigcup_{k=0}^n B_k$ et aussi $\bigcup_{k=0}^\infty A_k = \bigcup_{k=0}^\infty B_k$. Ainsi, par σ-additivité, on a

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \mathbb{P}\Big(\bigcup_{n=1}^{\infty} B_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{n\to+\infty} \sum_{k=1}^{n} \mathbb{P}(B_k) = \lim_{n\to+\infty} \mathbb{P}\Big(\bigcup_{k=1}^{n} B_k\Big) = \lim_{n\to+\infty} \mathbb{P}(A_n)\,. \qquad \square$$

Corollaire 1.12. *Soit $(A_k)_{k\geq 0}$ une suite quelconque d'événements de $\mathcal{F}$, alors*

$$\mathbb{P}\Big(\bigcup_{k=0}^{\infty} A_k\Big) = \lim_{n\to+\infty} \mathbb{P}\Big(\bigcup_{k=0}^{n} A_k\Big)\,, \qquad \mathbb{P}\Big(\bigcap_{k=0}^{\infty} A_k\Big) = \lim_{n\to+\infty} \mathbb{P}\Big(\bigcap_{k=0}^{n} A_k\Big)\,.$$

Démonstration. Il suffit d'appliquer le résultat précédent aux suites d'événements $(B_n)_{n\in\mathbb{N}}$ suivantes :
— $B_n := \bigcup_{k=0}^n A_k$, qui donne une suite croissante ;
— $B_n := \bigcap_{k=0}^n A_k$, qui donne une suite décroissante. $\square$

Proposition 1.13 (**Sous-additivité**). *Si $(A_k)_{k\geq 0}$ est une suite d'événements de $\mathcal{F}$, alors*

$$\mathbb{P}\Big(\bigcup_{k=0}^{\infty} A_k\Big) \leq \sum_{k=0}^{\infty} \mathbb{P}(A_k)\,.$$

Démonstration. En utilisant que $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$, on montre facilement par récurrence que pour tout $n \geq 1$

$$\mathbb{P}\Big(\bigcup_{k=0}^{n} A_k\Big) \leq \sum_{k=0}^{n} \mathbb{P}(A_k)\,.$$

On peut alors prendre la limite quand $n \to +\infty$: grâce au corollaire précédent, on obtient la conclusion voulue. $\square$

Exercice 1.14 (À retenir !). Soit $(A_i)_{i\in I}$ une famille dénombrable d'événements. Montrer que

1. si $\mathbb{P}(A_i) = 0$ pour tout $i \in I$, alors $\mathbb{P}\big(\bigcup_{i\in I} A_i\big) = 0$;
2. si $\mathbb{P}(A_i) = 1$ pour tout $i \in I$, alors $\mathbb{P}\big(\bigcap_{i\in I} A_i\big) = 1$.

Une dernière propriété, extrêmement utile, est celle de la formule de décomposition des probabilités.

Proposition 1.15 (**Formule de décomposition**). *Soit $(B_i)_{i\in I}$ une partition dénombrable de Ω : les $B_i \in \mathcal{F}$ sont deux à deux disjoints et $\bigcup_{i\in I} B_i = \Omega$. Alors, pour tout événement $A \in \mathcal{F}$, on a*

$$\mathbb{P}(A) = \sum_{i\in I} \mathbb{P}(A \cap B_i)\,.$$

On utilise souvent cette formule dans le cadre suivant : $\mathbb{P}(A) = \mathbb{P}(A\cap B) + \mathbb{P}(A \cap B^c)$.

1.2 Probabilités conditionnelles

Définition 1.16. Soit $B \in \mathcal{F}$ un événement tel que $\mathbb{P}(B > 0)$. On définit, pour tout $A \in \mathcal{F}$ la *probabilité conditionnelle de A sachant B* par

$$\mathbb{P}_B(A) = \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \,.$$

Remarque 1.17. Soit $B \in \mathcal{F}$ tel que $\mathbb{P}(B > 0)$. Alors la fonction $A \mapsto \mathbb{P}(A \mid B)$ définie sur $\mathcal{F}$ est une mesure de probabilité sur Ω. (En exercice)

On peut voir $\mathbb{P}(\cdot \mid B)$ comme la probabilité $\mathbb{P}$ que l'on n'applique qu'à des événements inclus dans B (et renormalisée par $\mathbb{P}(B)$ pour garder une mesure de probabilité).

Attention, l'application $B \mapsto \mathbb{P}(A \mid B)$ ne jouit pas de bonnes propriétés ! En particulier, ce n'est pas une probabilité.

Proposition 1.18 (Probabilités conditionnelles en chaîne). *Soit $n \geq 2$ et $A_1, \ldots, A_n$ des événements tels que $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) > 0$. Alors*

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_3 \mid A_2 \cap A_3) \cdots \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \,.$$

Démonstration. Facile, par récurrence. □

Exemple 1.19 (Le paradoxe des anniversaires)**.** Considérons n personnes : quelle est la probabilité p_n qu'au moins deux d'entre elles partagent le même anniversaire ? (Supposons qu'une année a toujours 365 jours.) À quel point n doit-il être grand pour que cette probabilité soit supérieure à $1/2$.
Solution. Il est plus facile de calculer la probabilité de $A =$« toutes les dates d'anniversaires sont différentes ». Pour $i \geq 2$ on note $A_i =$« l'anniversaire de la i-ème personne est différent de celui des $i-1$ premières », de sorte que $A = A_2 \cap \cdots \cap A_n$. Par la formule des probabilités conditionnelles en chaîne,

$$\mathbb{P}(A) = \mathbb{P}(A_2)\mathbb{P}(A_3 \mid A_2) \cdots \mathbb{P}(A_n \mid A_2 \cap \cdots \cap A_{n-1}) = \prod_{i=2}^{n} \Big(1 - \frac{i-1}{365}\Big)$$

car $\mathbb{P}(A_i \mid A_2 \cap \cdots \cap A_{i-1}) = 1 - \frac{i-1}{365}$ (l'anniversaire de la i-ème personne ne doit pas être parmi les $i-1$ dates d'anniversaire des personnes précédentes). La probabilité recherchée vaut donc (après un changement d'indice dans le produit)

$$p_n = 1 - \prod_{i=1}^{n-1} \Big(1 - \frac{i}{365}\Big) \,.$$

On peut voir à l'aide d'un ordinateur que cette probabilité dépasse $1/2$ pour $n \geq 23$ (ce qui est relativement faible comparé à ce que notre intuition suggère). Pour $n = 50$, on a $p_n \approx 97\%$. Une autre manière d'estimer cette probabilité est d'utiliser l'inégalité $1 - x \leq e^{-x}$ pour tout $x \in \mathbb{R}$: cela donne

$$p_n \geq 1 - \prod_{i=1}^{n-1} e^{-\frac{i}{365}} = 1 - \exp\Big(-\frac{(n-1)n}{2 \times 365}\Big) \,.$$

Cela permet de montrer que $p_n > 1/2$ pour $n \geq 23$.

Proposition 1.20 (Formule des probabilités totales). *Soit $(B_i)_{i \in I}$ une partition dénombrable de Ω : les $B_i \in \mathcal{F}$ sont deux à deux disjoints et $\bigcup_{i \in I} B_i = \Omega$. On suppose que $\mathbb{P}(B_i) > 0$ pour tout $i \in I$. Alors, pour tout événement $A \in \mathcal{F}$, on a la* formule des probabilités totales *:*

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(B_i)\mathbb{P}(A \mid B_i) \,.$$

Démonstration. Utiliser la formule de décomposition, puis le fait que $\mathbb{P}(A \cap B_i) = \mathbb{P}(B_i)\mathbb{P}(A \mid B_i)$. □

Proposition 1.21 (**Formule de Bayes**). *Soit $A, B \in \mathcal{F}$ tels que $\mathbb{P}(A), \mathbb{P}(B) > 0$. Alors*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A)\mathbb{P}(B \mid A)}{\mathbb{P}(B)}\,.$$

Démonstration. Il suffit d'écrire $P(B)\mathbb{P}(A \mid B) = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B \mid A)$. □

Exercice 1.22 (Un exercice classique)**.** Une maladie M affecte une personne sur 1000 dans une population donnée. Un test sanguin permet de détecter cette maladie avec une fiabilité de 99% (lorsque cette maladie est effectivement présente). En revanche, pour un individu sain, la probabilité que le test soit positif est de 0,1% (on dit que 0,1% est le taux de faux positifs).

1. Un patient fait le test, qui est positif. Quelle est la probabilité (conditionnelle) que l'individu soit réellement malade ?
2. Le même patient refait le test, qui est de nouveau positif. Quelle est la probabilité que l'individu soit malade ?

Solution. On choisit comme espace d'états Ω l'ensemble des individus de la population. On appelle M l'événement « l'individu est malade », et P_1 l'événement « le test est positif ». Alors, l'énoncé nous dit que

$$\mathbb{P}(M) = \frac{1}{1000};\ \text{d'où}\ \mathbb{P}(M^c) = \frac{999}{1000}; \quad \mathbb{P}(P_1 \mid M) = \frac{99}{100}; \quad \mathbb{P}(P_1 \mid M^c) = \frac{1}{1000}\,.$$

1. On cherche

$$\mathbb{P}(M \mid P_1) = \frac{\mathbb{P}(M \cap P_1)}{\mathbb{P}(P_1)} = \frac{\mathbb{P}(P_1 \mid M)\mathbb{P}(M)}{\mathbb{P}(P_1 \mid M)\mathbb{P}(M) + \mathbb{P}(P_1 \mid M^c)\mathbb{P}(M^c)} = \frac{990}{990+999} \approx 0{,}5.$$

Si le test est positif, la probabilité que l'individu soit réellement malade est $\approx 0{,}5$ (il a autant de chance que le test soit un faux positif). Donc attention avant de tirer des conclusions trop hâtives.

2. On note maintenant P_2 l'événement « le 2ème test est positif ». On cherche à déterminer

$$\mathbb{P}(M \mid P_1 \cap P_2) = \frac{\mathbb{P}(M \cap P_1 \cap P_2)}{\mathbb{P}(P_1 \cap P_2)} = \frac{\mathbb{P}(M)\mathbb{P}(P_1 \cap P_2 \mid M)}{\mathbb{P}(M)\mathbb{P}(P_1 \cap P_2 \mid M) + \mathbb{P}(M^c)\mathbb{P}(P_1 \cap P_2 \mid M^c)}\,.$$

On suppose que : (i) si l'individu est malade alors les tests seront positifs avec probabilité 0,99 indépendamment l'un de l'autre ; (ii) si l'individu n'est pas malade alors les tests seront positifs avec probabilité 0,001 indépendamment l'un de l'autre. Autrement dit, on a les valeurs $\mathbb{P}(P_1 \cap P_2 \mid M) = (99/100)^2$ et $\mathbb{P}(P_1 \cap P_2 \mid M^c) = (1/1000)^2$. Ainsi,

$$\mathbb{P}(M \mid P_1 \cap P_2) = \frac{\frac{1}{1000}(\frac{99}{100})^2}{\frac{1}{1000}(\frac{99}{100})^2 + \frac{999}{1000}(\frac{1}{1000})^2} = \frac{99^2}{99^2 + \frac{999}{100}} \approx 0{,}999\,.$$

Exercice 1.23 (Paradoxe de Simpson)**.** Voici un tableau de données de patients hospitalisés Covid en Angleterre entre le 21 juin et le 12 septembre 2021, de variant séquencé Delta[1].

groupe d'âge	décès de personnes hospitalisées	dont non vaccinées
< 50	196	126
≥ 50	2 227	552
Total	2 423	678

On choisit au hasard une personne de la population et on note les événements $A_1 =$ « la personne est hospitalisée et a moins de 50 ans », $A_2 =$ « la personne est hospitalisée et a plus de 50 ans », $V =$ « la personne est vaccinée (ou partiellement vaccinée) », $D =$ « la personne est décédée ».

1. La proportion de personnes vaccinées était de 50% parmi les moins de 50 ans. Calculer (et interpréter) le ratio $\mathbb{P}(D \mid A_1 \cap V^c)/\mathbb{P}(D \mid A_1 \cap V)$.
2. La proportion de personnes vaccinées était de 95% parmi les plus de 50 ans. Calculer (et interpréter) le ratio $\mathbb{P}(D \mid A_2 \cap V^c)/\mathbb{P}(D \mid A_2 \cap V)$.

1. Source : SARS-CoV-2 variants of concern and variants under investigation in England: technical briefings

3. Les moins de 50 ans représentent environ 63% de la population. Calculer (et interpréter) le ratio $\mathbb{P}(D \mid V^c)/\mathbb{P}(D \mid V)$.

Solution.

1. Les données de l'énoncé donnent $\mathbb{P}(V^c \mid D \cap A_1) = \frac{126}{196} \approx 64{,}3\%$ (et $\mathbb{P}(V \mid D \cap A_1) = 1 - P(V^c \mid D \cap A_1) = \frac{70}{196} \approx 35{,}7\%$) ainsi que $\mathbb{P}(V \mid A_1) = 1/2$. Par la formule des probabilités conditionnelles, on a

$$\mathbb{P}(D \mid A_1 \cap V^c) = \frac{\mathbb{P}(D \cap A_1 \cap V^c)}{\mathbb{P}(A_1 \cap V^c)} = \frac{\mathbb{P}(V^c \mid D \cap A_1)}{\mathbb{P}(D \cap A_1)} \times \frac{\mathbb{P}(A_1)}{\mathbb{P}(V^c \mid A_1)}$$

et de même

$$\mathbb{P}(D \mid A_1 \cap V^c) = \frac{\mathbb{P}(D \cap A_1 \cap V^c)}{\mathbb{P}(A_1 \cap V^c)} = \frac{\mathbb{P}(V \mid D \cap A_1)}{\mathbb{P}(D \cap A_1)} \times \frac{\mathbb{P}(A_1)}{\mathbb{P}(V \mid A_1)}\,.$$

On en conclut que

$$\frac{\mathbb{P}(D \mid A_1 \cap V^c)}{P(D \mid A_1 \cap V)} = \frac{\mathbb{P}(V^c \mid D \cap A_1)}{\mathbb{P}(V^c \mid D \cap A_1)} \times \frac{\mathbb{P}(V \mid A_1)}{\mathbb{P}(V^c \mid A_1)} = \frac{126/196}{70/196} = 1{,}8\,.$$

Ainsi, parmi les moins de 50 ans, les personnes hospitalisées non-vaccinées ont 1,8 fois plus de chance de décéder que celles qui sont vaccinées (ou partiellement vaccinées).

2. De la même manière que ci-dessus, on a $\mathbb{P}(V^c \mid D \cap A_2) = \frac{552}{2227} \approx 24{,}8\%$ (et $\mathbb{P}(V \mid D \cap A_2) = 1 - P(V^c \mid D \cap A_2) = \frac{1675}{2227} \approx 75{,}2\%$) ainsi que $\mathbb{P}(V \mid A_1) = 95/100$. Exactement comme ci-dessus, on a

$$\frac{\mathbb{P}(D \mid A_2 \cap V^c)}{P(D \mid A_2 \cap V)} = \frac{\mathbb{P}(V^c \mid D \cap A_2)}{\mathbb{P}(V^c \mid D \cap A_2)} \times \frac{\mathbb{P}(V \mid A_2)}{\mathbb{P}(V^c \mid A_2)} = \frac{552/2227}{1675/2227} \times \frac{95/100}{5/100} \approx 6{,}3\,.$$

Ainsi, parmi les plus de 50 ans, les personnes hospitalisées non-vaccinées ont environ 6,3 fois plus de chance de décéder que celles qui sont vaccinées (ou partiellement vaccinées).

3. Comme plus haut, on a $\mathbb{P}(V^c \mid D) = \frac{678}{2423} \approx 28\%$ (et $\mathbb{P}(V \mid D) = 1 - P(V^c \mid D) = \frac{1745}{2423} \approx 70\%$). On a encore la formule analogue à plus haut :

$$\frac{\mathbb{P}(D \mid V^c)}{P(D \mid V)} = \frac{\mathbb{P}(V^c \mid D)}{\mathbb{P}(V \mid D)} \times \frac{\mathbb{P}(V)}{\mathbb{P}(V^c)}\,,$$

et il reste à déterminer $\mathbb{P}(V)$. On a $\mathbb{P}(V) = \mathbb{P}(A_1)\mathbb{P}(V \mid A_1) + \mathbb{P}(A_2)\mathbb{P}(V \mid A_2) = \frac{63}{100} \times \frac{50}{100} + \frac{37}{100} \times \frac{95}{100} \approx 67\%$, et $\mathbb{P}(V^c) = 1 - \mathbb{P}(V) \approx 33\%$. On obtient donc

$$\frac{\mathbb{P}(D \mid V^c)}{P(D \mid V)} = \frac{678}{1745}\frac{67}{33} \approx 0{,}8\,.$$

Ainsi, dans la population générale, les personnes hospitalisées non-vaccinées ont moins de chance de décéder que celles qui sont vaccinées (ou partiellement vaccinées) !

Ce paradoxe est dû au fait que la population est très inhomogène : parmi les personnes non-vaccinées, la grande majorité sont des personnes de moins de 50 ans, dont le taux de mortalité du Covid est plus élevé (même en étant vaccinées).

Le but était ici d'illustrer avec de vraies données ce genre de paradoxes : cela renseigne sur les pièges liés au conditionnement, qui vont souvent à l'encontre de notre intuition. Une conclusion possible est que, pour pouvoir comparer des données, il est important de *comparer à l'intérieur d'une catégorie où les caractéristiques ne sont pas trop inhomogènes.*

1.3 Indépendance d'événements/de tribus

Définition 1.24 (Indépendance de tribus/d'événements)**.** Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé.

• Soit $(\mathcal{F}_i)_{i\in I}$ une famille de sous-tribus de $\mathcal{F}$, où I est un ensemble d'indices quelconque. On dit que $(\mathcal{F}_i)_{i\in I}$ forme une famille *indépendante* (ou par abus de langage que les $\mathcal{F}_i$ sont indépendants) si pour tout sous-ensemble d'indices $J \subseteq I$ *fini* on a

$$\forall\, A_j \in \mathcal{F}_j,\, j \in J\,, \quad \mathbb{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} \mathbb{P}(A_j)$$

• Soit $(A_i)_{i\in I}$ une famille d'événements de $\mathcal{F}$. On dit que $(A_i)_{i\in I}$ forme une famille *indépendante* d'événements (ou par abus de langage que les événements sont indépendants) si pour tout sous-ensemble d'indices $J \subseteq I$ *fini* on a

$$\mathbb{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} \mathbb{P}(A_j)\,.$$

Remarque 1.25. La notion d'indépendance est relative à la probabilité $\mathbb{P}$ utilisée $(\Omega, \mathcal{F}, \mathbb{P})$.

Exercice 1.26. Soient $A, B \in \mathcal{F}$ deux événements indépendants. On pose $\mathcal{A} = \sigma(A)$ et $\mathcal{B} = \sigma(B)$ pour tous $A \in \mathcal{A}$ et $B \in \mathcal{B}$ on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Montrer que les tribus engendrées $\sigma(\mathcal{A})$ et $\sigma(\mathcal{B})$ sont indépendantes.

Note : si $A \in \mathcal{F}$, alors $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$. Cela montre notamment que si A, B sont indépendants, alors A, B^c le sont aussi.

Lemme 1.27 (Indépendance de tribus engendrées)**.** *Si $\mathcal{A}, \mathcal{B}$ sont deux sous-ensembles de $\mathcal{F}$ tels que pour tous $A \in \mathcal{A}$ et $B \in \mathcal{B}$ on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, et si $\mathcal{A}, \mathcal{B}$ sont des π-systèmes (c'est-à-dire stable par intersections finies), alors les tribus engendrées $\sigma(\mathcal{A}), \sigma(\mathcal{B})$ sont indépendantes.*

Démonstration. La démonstration repose sur le lemme des classes monotones : il faut montrer que pour tout B, l'ensemble $\{A \in \sigma(\mathcal{A}), \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)\}$ est une classe monotone. □

Remarque 1.28 (Indépendance et indépendance deux à deux)**.** Attention, l'indépendance deux à deux n'est pas suffisante pour montrer l'indépendance complète. Un exemple : on lance deux dés et on considère les événements $A =$« le premier dé est impair », $B =$« le deuxième dé est impair » et $C =$« la somme des deux dés est impaire ». Alors A et B sont indépendants, A et C sont indépendants, B et C sont indépendants, mais A, B, C ne sont pas indépendants (exercice).

D'autre part, ce n'est pas parce que $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ que les trois événements sont indépendants (notamment indépendants deux à deux) : par exemple $A =$« le premier dé est impair », $B =$« le premier dé est supérieur ou égal à 4 » et $C =$« la somme des deux dés est égale à 9 ».

Exemple 1.29 (Développement eulérien de la fonction zeta)**.** Références [App15, p. 167], [GK19, p. 56] [BCD21, p. 44]. La notion d'indépendance permet de démontrer la formule suivante, due à Euler : pour tout $s > 1$,

$$\zeta(s) := \sum_{n\in\mathbb{N}^*} \frac{1}{n^s} = \prod_{p\in\mathfrak{P}} \frac{1}{1-p^{-s}}\,,$$

où $\mathfrak{P}$ est l'ensemble des nombres premiers.

Étape 1. Soit $s > 1$ fixé. On définit une probabilité μ_s sur $\mathbb{N}^*$ par $\mu_s(\{n\}) = \frac{1}{\zeta(s)} n^{-s}$. Il s'agit bien d'une probabilité car $\sum_{n\in\mathbb{N}^*} \mu_s(\{n\}) = 1$. On note que $\zeta(s) = 1/\mu_s(\{1\})$.

L'idée est alors d'écrire $\{1\} = \bigcap_{p\in\mathfrak{P}} A_p^c$, où $A_p := p\mathbb{N}^*$ est l'ensemble des multiples de p, vu que 1 est le seul entier qui n'est multiple d'aucun nombre premier. Le reste de la preuve consiste à montrer que : (a) $\mu_s(A_p) = p^{-s}$; (b) les événements $(A_p)_{p\in\mathfrak{P}}$ sont indépendants (par rapport à la probabilité μ_s).

Étape 2. On a facilement, par σ-additivité et en utilisant la définition de la fonction ζ :

$$\mu_s(A_p) = \sum_{k=1}^{\infty} \mu_s(\{kp\}) = \sum_{k=1}^{\infty} \frac{1}{\zeta(s)k^s p^s} = \frac{1}{p^s}\,.$$

Étape 3. Si $p_1, \ldots, p_\ell$ sont des nombres premiers distincts, alors on a $A_{p_1} \cap \cdots \cap A_{p_\ell} = A_{p_1 \cdots p_\ell}$: ainsi, en utilisant l'étape précédente,

$$\mu_s\Big(A_{p_1} \cap \cdots \cap A_{p_\ell}\Big) = \mu_s\big(A_{p_1 \cdots p_\ell}\big) = \frac{1}{(p_1 \cdots p_\ell)^s} = \frac{1}{p_1^s \cdots p_\ell^s} = \mu_s(A_{p_1}) \cdots \mu_s(A_{p_\ell}) .$$

Cela montre que les événements $(A_p)_{p \in \mathfrak{P}}$ sont indépendants.

Étape 4. Conclusion. En utilisant l'indépendance des $(A_p)_{p \in \mathfrak{P}}$ (et la continuité par le haut des probabilités), on a

$$\mu_s(\{1\}) = \mu_s\Big(\bigcap_{p \in \mathfrak{P}} A_p^c\Big) = \prod_{p \in \mathfrak{P}} \mu_s(A_p^c) = \prod_{p \in \mathfrak{P}} (1 - p^{-s}) ,$$

où on a aussi utilisé que $\mu_s(A_p) = p^{-s}$. La conclusion découle du fait que $\zeta(s) = \mu_s(\{1\})^{-1}$.

Exercice 1.30. En utilisant la formule du développement eulérien de la fonction ζ, montrer que $\prod_{p \in \mathfrak{P}} (1 - \frac{1}{p}) = 0$. En déduire que $\sum_{p \in \mathfrak{P}} \frac{1}{p} = +\infty$.

Exercice 1.31. Le but de cet exercice est de montrer qu'il n'existe pas de mesure de probabilité μ *naturelle* sur $\mathbb{N}^*$, c'est-à-dire telle que $\mu(A_k) = \frac{1}{k}$ pour tout $k \geq 1$, où $A_k = k\mathbb{N}^*$ est l'ensemble des multiples de k. On suppose donc qu'une telle probabilité μ existe.

1. Montrer que les $(A_p)_{p \in \mathfrak{P}}$ sont indépendants (par rapport à μ), où $\mathfrak{P}$ est l'ensemble des nombres premiers.
2. Montrer que pour tout $n \in \mathbb{N}^*$, $\mu(\{n\}) \leq \prod_{p \geq n, p \in \mathfrak{P}} (1 - p^{-1})$ et en déduire que pour tout $n \in \mathbb{N}^*$ on a $\mu(\{n\}) = 0$. (Indication : utiliser l'exercice précédent).
3. Conclure.

Lemme 1.32 (Indépendance par paquets). *Soit $(\mathcal{F}_i)_{i \in I}$ une famille tribus indépendantes. et soient $(J_k)_{k \in K}$ des sous-ensembles de I* disjoints. *Alors la famille de tribus $(\mathcal{G}_k)_{k \in K}$ est indépendante, où pour $\mathcal{G}_k := \sigma(\bigcup_{j \in J_k} \mathcal{F}_j)$ (la tribu engendrée par le paquet d'indices J_k).*

Note : ce lemme est parfois appelé *lemme des coalitions.*

Proposition 1.33 (Loi du 0-1 de Kolmogorov). *Soit $(\mathcal{F}_i)_{i \geq 1}$ une suite de tribus* indépendantes. *On définit $\mathcal{F}_{>n} := \sigma(\bigcup_{i > n} \mathcal{F}_i)$ la tribu engendrée par $\mathcal{F}_i$ pour $i > n$ et on pose $\mathcal{Q}_\infty := \bigcap_{n \in \mathbb{N}} \mathcal{F}_{>n}$, appelée* tribu asymptotique. *Alors pour tout $A \in \mathcal{Q}_\infty$ on a $\mathbb{P}(A) \in \{0, 1\}$.*

Rappel : l'union de tribus n'est pas nécessairement une tribu, mais l'intersection (quelconque) de tribus est toujours une tribu.

Démonstration. Soit $\mathcal{F}_{\leq n} := \sigma(\bigcup_{i \leq n} \mathcal{F}_i)$ et soit $\mathcal{A} := \bigcup_{n=1}^\infty \mathcal{F}_{\leq n}$ (qui n'est pas une tribu). Alors pour tout $A \in \mathcal{A}$, il existe $n \geq 0$ tel que $A \in \mathcal{F}_{\leq n}$: comme $\mathcal{F}_{\leq n}$ est indépendante de $\mathcal{F}_{>n}$ (par indépendance par paquet) on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ pour tout $B \in \mathcal{Q}_\infty \subseteq \mathcal{F}_{>n}$. Comme $\mathcal{A}$ est un π-système, le Lemme 1.27 montre que les tribus $\sigma(\mathcal{A})$ et $\mathcal{Q}_\infty$ sont indépendantes.

Soit maintenant $A \in \mathcal{Q}_\infty$. On a aussi $A \in \sigma(\mathcal{A}) = \sigma(\bigcup_{i \in \mathbb{N}} \mathcal{F}_i)$, donc par indépendance des tribus $\mathcal{Q}_\infty$ et $\sigma(\mathcal{A})$ on obtient $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$. L'équation $x = x^2$ n'a que 0 ou 1 comme solution : on conclut que $\mathbb{P}(A) \in \{0, 1\}$. ☐

Exemple 1.34. Considérons une suite infinie de tentatives *indépendantes*, de probabilité de succès quelconque et notons $A_i =$« la i-ème tentative est un succès » (qui sont des événements indépendants). Alors les événements suivants sont de probabilité 0 ou 1 :

- $A =$« Il y a une infinité de succès » ;
- $B =$« La différence entre le nombre de succès et le nombre d'échec forme une suite majorée » ;
- $C =$« La proportion de succès converge ».

La loi du 0-1 de Kolmogorov peut se comprendre de manière informelle de la façon suivante : tout événement qui *ne dépend pas d'un nombre fini de A_i* est de probabilité 0 ou 1.

Solution. Il suffit de montrer que les trois événements sont dans la tribu asymptotique $\mathcal{Q}_\infty$.

On peut écrire $A =$« pour tout $m \geq 1$ il y a au moins un succès après le rang m »$= \bigcap_{m\geq 1} \bigcup_{i\geq m} A_i$. Mais, pour tout $n \in \mathbb{N}$, on peut aussi réécrire A en disant $A =$« pour tout $m > n$ il y a au moins un succès après le rang m »$= \bigcap_{m>n} \bigcup_{i\geq m} A_i$. Ainsi, on a clairement $A \in \mathcal{F}_{>n}$ pour tout $n \in \mathbb{N}$ (autrement dit A ne dépend pas des n premiers A_i, pour aucun $n \in \mathbb{N}$), ce qui montre que $A \in \mathcal{Q}_\infty$.

On peut réécrire $B =$« La différence entre le nombre de succès et le nombre d'échec, comptée à partir de la n-ème tentative, forme une suite majorée », ce qui montre que $B \in \mathcal{F}_{>n}$ pour tout $n \in \mathbb{N}$, donc $B \in \mathcal{Q}_\infty$. Le cas de l'événement C est analogue.

1.4 Lemme de Borel–Cantelli

Pour $(A_i)_{i\geq 1}$ une suite dévénements on note

$$\limsup A_i = \bigcap_{n\geq 1} \bigcup_{i\geq n} A_i = \text{« il y a une infinité de } A_i \text{ réalisés »}\,,$$
$$\liminf A_i = \bigcup_{n\geq 1} \bigcap_{i\geq n} A_i = \text{« tous les } A_i \text{ sont réalisés à partir d'un certain rang »}\,.$$

Attention, ici $\limsup, \liminf$ désignent des ensembles ! L'analogie avec les *vraies* $\limsup / \liminf$ se fait en interprétant l'union comme un sup d'ensemble et l'intersection comme un inf d'ensembles. Noter que l'on a $(\limsup A_i)^c = \liminf A_i^c$.

Rappel : si $\mathbb{P}(A) = 1$, on dit que A est un événement *presque sûr* (on notera p.s.).

Lemme 1.35. *Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et soit $(A_i)_{i\geq 1}$ une famille d'événements.*

(i) Si $\sum_{i=1}^\infty \mathbb{P}(A_i) < +\infty$, alors $\mathbb{P}(\limsup A_i) = 0$. Autrement dit : presque sûrement, il n'y a qu'un nombre fini de A_i réalisés.

(ii) Si les $(A_i)_{i\geq 1}$ sont indépendants *et si $\sum_{i=1}^\infty \mathbb{P}(A_i) = +\infty$, alors $\mathbb{P}(\limsup A_i) = 1$. Autrement dit :* presque sûrement, il y a une infinité de A_i réalisés.

Démonstration. (i) Si $\sum_{i=1}^\infty \mathbb{P}(A_i) < +\infty$. Alors, par continuité par le haut des probabilités (la suite d'événements $(B_n := \bigcup_{i\geq n} A_i)_{n\geq 1}$ est décroissante) puis par sous-additivité

$$\mathbb{P}\Big(\bigcap_{n\geq 1} \bigcup_{i\geq n} A_i\Big) = \lim_{n\to+\infty} \mathbb{P}\Big(\bigcup_{i\geq n} A_i\Big) \leq \lim_{n\to+\infty} \sum_{i=n}^{\infty} \mathbb{P}(A_i)\,.$$

La série $\sum_{i=1}^\infty \mathbb{P}(A_i)$ étant convergente, le reste tend vers 0, ce qui donne la conclusion voulue.

(ii) Si les $(A_i)_{i\geq 1}$ sont *indépendants* et $\sum_{i=1}^\infty \mathbb{P}(A_i) = +\infty$. En passant au complémentaire il suffit de montrer que $\mathbb{P}\big(\bigcup_{n\geq 1} \bigcap_{i\geq n} A_i^c\big) = 0$. Par sous-additivité, on a

$$\mathbb{P}\Big(\bigcup_{n\geq 1} \bigcap_{i\geq n} A_i^c\Big) \leq \sum_{n=1}^{\infty} \mathbb{P}\Big(\bigcap_{i\geq n} A_i^c\Big)\,,$$

et il suffit donc de montrer que $\mathbb{P}\big(\bigcap_{i\geq n} A_i^c\big) = 0$, pour tout $n \geq 1$. Par continuité par le haut des probabilités, pour tout $n \geq 1$ on a

$$\mathbb{P}\Big(\bigcap_{i\geq n} A_i^c\Big) = \lim_{k\to+\infty} \mathbb{P}\Big(\bigcap_{i=n}^{n+k} A_i^c\Big) = \lim_{k\to+\infty} \prod_{i=n}^{n+k} (1 - \mathbb{P}(A_i))\,,$$

où on a utilisé l'indépendance des A_i pour la dernière égalité. Ensuite, en notant que $1 - \mathbb{P}(A_i) \leq \exp(-\mathbb{P}(A_i))$, on obtient

$$0 \leq \prod_{i=n}^{n+k} (1 - \mathbb{P}(A_i)) \leq \exp\Big(-\sum_{i=n}^{n+k} \mathbb{P}(A_i)\Big) \xrightarrow{k\to\infty} 0\,,$$

car la série $\sum_{i=n}^\infty \mathbb{P}(A_i)$ est divergente. Cela conclut la démonstration. □

Exemple 1.36. Considérons une suite infinie de tentatives *indépendantes*, de même probabilité de succès $p \in]0,1[$ et notons $A_i =$« la i-ème tentative est un succès » (qui sont des événements indépendants). Comme $\sum_{i\geq 1} \mathbb{P}(A_i) = \sum_{i\geq 1} p = +\infty$, d'après le lemme de Borel–Cantelli on obtient que, presque sûrement, il y a une infinité de succès.

Exemple 1.37 (Le singe de Borel)**.** Un chimpanzé tape à la machine à écrire en appuyant chaque seconde sur une touche choisie au hasard. Quelle est la probabilité qu'il parvienne à écrire le texte complet de *Hamlet*, c'est-à-dire qu'à un certain moment il écrive d'une traite le texte de cette pièce ?

Solution. Soit n la longueur, en caractères, de la pièce Hamlet. La probabilité p qu'il tape la pièce du premier coup est faible, mais strictement positive. Pour tout entier naturel k, on définit l'événement $A_k =$« les caractères $nk+1$ à $n(k+1)$ correspondent exactement au texte de la pièce ». Ces événements sont indépendants (ils dépendent de différents paquets d'événements) et de probabilité $p > 0$. D'après le lemme de Borel–Cantelli, presque sûrement, il y a une infinité de A_k réalisés, c'est-à-dire que presque sûrement, le chimpanzé écrira Hamlet une infinité de fois (donc a fortiori au moins une fois).

2 Variables aléatoires et lois

2.1 Définitions générales

De manière informelle, une variable aléatoire est un nombre (ou un objet) associé au résultat d'une expérience aléatoire. Autrement dit, c'est une fonction du résultat $\omega \in \Omega$ de l'expérience.

Définition 2.1. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. On dit que X est une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans l'espace mesurable $(E, \mathcal{E})$ si $\omega \mapsto X(\omega)$ est une fonction mesurable.

Autrement dit, pour tout $A \in \mathcal{E}$, l'ensemble $X^{-1}(A) = \{\omega \in \Omega, X(\omega) \in A\}$ est un événement de $\mathcal{F}$. Notation : on notera plus simplement $\{X \in A\} := X^{-1}(A)$ et $\mathbb{P}(X \in A) = \mathbb{P}(\{X \in A\})$; remarquer que par définition $\{X \in A\} = X^{-1}(A) \in \mathcal{F}$. On écrira aussi souvent v.a. au lieu de variable aléatoire.

Exemple 2.2. Prenons l'expérience consistant à tirer un nombre au hasard dans $[0,1]$: l'espace probabilisé est $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbb{P}$ =mesure de Lebesgue sur $[0,1]$. On peut considérer les variables aléatoires suivantes :

- $X : \omega \mapsto \omega$ ($X =$ le nombre tiré) ;
- $Y : \omega \mapsto \max\{n \in \mathbb{N}, n \leq 1/\omega\}$ ($Y =$ la partie entière de l'inverse du nombre tiré) ;
- $Z : \omega \mapsto (\omega_i)_{i\geq 1}$ où ω_i est la i-ème décimale de ω ($Z =$ développement décimal du nombre tiré).

2.1.1 Loi d'une variable aléatoire

Définition 2.3. Soit X une variable aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans l'espace mesurable $(E, \mathcal{E})$. On appelle *loi de* X la mesure de probabilité $\mathbb{P}_X$ sur $(E, \mathcal{E})$ définie par

$$\mathbb{P}_X : \begin{array}{l} \mathcal{E} \to [0,1] \\ A \mapsto \mathbb{P}_X(A) := \mathbb{P}(X \in A)\,. \end{array}$$

Autrement dit, $\mathbb{P}_X$ est la mesure image de $\mathbb{P}$ par l'application X.

Remarque 2.4. La loi de $\mathbb{P}_X$ est une probabilité sur l'espace d'arrivée $(E, \mathcal{E})$. Elle permet de calculer toutes les probabilités liées à la v.a. X. On verra dans la suite plusieurs manières de caractériser la loi de X. Par exemple, si E est dénombrable, il suffit de donner la valeur de $\mathbb{P}(X = x)$ pour tout $x \in E$.

On considérera principalement des v.a. à valeurs dans $\mathbb{R}^d$, ou dans un ensemble dénombrable. On verra dans les Sections 2.4-2.5 deux grandes classes de variables aléatoires : les variables aléatoires discrètes et les variables aléatoires à densité.

Exemple 2.5. Reprenons l'exemple précédent : l'espace probabilisé est $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbb{P}$ =mesure de Lebesgue sur $[0,1]$.

• Soit $X : \omega \mapsto \omega$. Alors clairement $\mathbb{P}_X = \mathbb{P}$, la mesure de Lebesgue sur $[0,1]$. On dit que X est de loi uniforme sur $[0,1]$, on note $X \sim \mathcal{U}([0,1])$.

• Soit $Y : \omega \mapsto \max\{n \in \mathbb{N}, n \geq 1/\omega\}$, qui est une variable aléatoire à valeurs dans $\mathbb{N}^*$. Pour déterminer la loi de Y, il suffit de déterminer $\mathbb{P}_Y(\{n\}) = \mathbb{P}(Y = n)$ pour tout $n \in \mathbb{N}^*$. On a

$$\mathbb{P}(Y=n) = \mathbb{P}\Big(1/\omega \in [n, n+1[\,\Big) = \mathbb{P}\Big(\omega \in \Big]\frac{1}{n+1}, \frac{1}{n}\Big]\Big) = \frac{1}{n} - \frac{1}{n+1} = \frac{1}{n(n+1)}\,.$$

On peut vérifier que $\sum_{n=1}^{\infty} \mathbb{P}(Y = n) = 1$ (il s'agit d'une somme téléscopique).

Exercice 2.6. Soient $X, \tilde{X}$ deux variables aléatoires à valeurs dans $(E, \mathcal{E})$ de même loi, et soit $f : (E,\mathcal{E}) \to (E', \mathcal{E}')$ une fonction mesurable. Montrer que $f(X)$ et $f(\tilde{X})$ ont la même loi.

2.1.2 Loi jointe et lois marginales

On peut considérer une variable aléatoire $X = (X_1, \dots, X_d)$ à valeurs dans $E_1 \times \cdots \times E_d$ muni de la tribu produit $\mathcal{E}_1 \otimes \cdots \otimes \mathcal{E}_d$ engendrée par les $A_1 \times \cdots \times A_d$, $A_i \in \mathcal{E}_i$.

La loi $\mathbb{P}_X = \mathbb{P}_{X_1,\dots,X_d}$ de $X = (X_1, \dots, X_d)$ est appelée la *loi jointe* des X_i, et les lois $\mathbb{P}_{X_i}$ pour $1 \leq i \leq d$ les *lois marginales* des X_i.

Remarque 2.7. Si on connaît la loi jointe $\mathbb{P}_X$ des X_i, on peut en déduire les lois marginales $\mathbb{P}_{X_i}$ (où les lois de $(X_i)_{i\in I}$ pour un sous-ensemble d'indices $I \subset \{1,\dots,d\}$) en notant que

$$\begin{aligned}\forall A \in \mathcal{E}_i\,, \qquad \mathbb{P}_{X_i}(A) = \mathbb{P}(X_i \in A) &= \mathbb{P}(X_1 \in E_1, \dots, X_i \in A, \dots, X_d \in E_d) \\ &= \mathbb{P}_X\big(E_1 \times \cdots \times A \times \cdots \times E_d\big)\,.\end{aligned}$$

Le réciproque est fausse : on ne peut pas en général déduire la loi jointe des lois marginales.

Exercice 2.8. Soit E un ensemble fini : on dit qu'une variable aléatoire X est de loi uniforme sur E si $\mathbb{P}(X = x) = \frac{1}{|E|}$ pour tout $x \in E$.

1. Soit $X = (X_1, X_2)$ de loi uniforme sur $E = E_1 \times E_2$, où E_1, E_2 sont deux ensembles finis. Montrer que X_1 est de loi uniforme sur E_1 et X_2 de loi uniforme sur E_2.
2. Soit X_1 une variable aléatoire de loi uniforme sur $\{0,1\}$ et posons $X_2 = 1 - X_1$. Montrer que X_2 est aussi de loi uniforme sur $\{0,1\}$, mais que $X = (X_1, X_2)$ n'est pas de loi uniforme sur $\{0,1\}\times\{0,1\}$.

Note : de manière générale, lorsque l'on s'intéresse uniquement à la loi de variables aléatoires, l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel elles sont définies n'est pas forcément pertinent : on peut faire les calcul directement sur les variables aléatoires, sans utiliser la définition explicite de $(\Omega, \mathcal{F}, \mathbb{P})$.

2.2 Variables à valeurs dans $\mathbb{R}^d$: la fonction de répartition

Définition 2.9. Soit $X = (X_1, \dots, X_d)$ une variable aléatoire à valeurs dans $\mathbb{R}^d$ (on parle de vecteur aléatoire). La fonction de répartition de X est la fonction $F_X : \mathbb{R}^d \to [0,1]$ définie par

$$F_X(t_1, \dots, t_d) = \mathbb{P}\Big(\forall 1 \leq i \leq d\,,\, X_i \leq t_i\Big) = \mathbb{P}_X\Big(\,]-\infty, t_1] \times \cdots \times]-\infty, t_d]\Big)\,.$$

Théorème 2.10. *Si deux variables aléatoires X et Y à valeurs dans $\mathbb{R}^d$ ont la même fonction de répartition, alors elles ont la même loi.*

Cela montre que la fonction de répartition caractérise la loi de variables aléatoires à valeurs dans $\mathbb{R}^d$. Voici l'idée de la démonstration (voir [GK19, p. 118]).

Démonstration. Les fonctions de répartitions étant identiques, les lois $\mathbb{P}_X$ et $\mathbb{P}_Y$ coïncident sur la classe d'ensemble

$$\mathcal{R} := \Big\{\,]-\infty, t_1] \times \cdots \times]-\infty, t_d]\,, (t_1, \dots, t_d) \in \mathbb{R}^d\Big\}\,.$$

Notons que $\mathcal{R}$ est stable par intersection finie (c'est un π-système) et $\sigma(\mathcal{R}) = \mathcal{B}(\mathbb{R}^d)$. Par le théorème π-λ, cela signifie que $\mathbb{P}_X$ et $\mathbb{P}_Y$ coïncident sur $\mathcal{B}(\mathbb{R}^d)$. ∎

2.2.1 Cas réel : propriétés

Proposition 2.11. *Soit X une variable aléatoire à valeurs dans $\mathbb{R}$ et soit $F_X : \mathbb{R} \to [0,1]$ sa fonction de répartition. Alors F_X vérifie les propriétés suivantes :*

$$\begin{aligned}
&\text{(i)} \quad F_X \text{ est croissante ;}\\
&\text{(ii)} \quad \lim_{t\to-\infty} F_X(t) = 0 \text{ et } \lim_{t\to+\infty} F_X(t) = 1;\\
&\text{(iii)} \quad F_X \text{ est continue à droite, c'est-à-dire } F_X(x^+) := \lim_{\varepsilon\to 0^+} F_X(x+\varepsilon) = F_X(x)\,.
\end{aligned} \tag{$*$}$$

Démonstration. (i) Il suffit d'observer que $\{X \le t\} \subseteq \{X \le t'\}$ si $t \le t'$. On en déduit que $F_X(t) = \mathbb{P}(X \le t) \le \mathbb{P}(X \le t') = F_X(t')$.

(ii) Par croissance, on a $\lim_{t\to-\infty} F_X(t) = \lim_{n\to+\infty} F_X(-n)$. Ensuite, par continuité par le haut des probabilités, on obtient

$$\lim_{n\to+\infty} F_X(-n) = \lim_{n\to\infty} \mathbb{P}(X \le -n) = \mathbb{P}\Big(\bigcap_{n\ge 1}\{X \le -n\}\Big) = \mathbb{P}(X = -\infty) = 0\,.$$

Le même raisonnement donne $\lim_{n\to+\infty} F_X(-n) = \mathbb{P}(X < +\infty) = 1$.

(iii) Pour tout $x \in \mathbb{R}$, par croissance, on a $\lim_{\varepsilon\to 0^+} F_X(x+\varepsilon) = \lim_{n\to+\infty} F_X(x+\frac{1}{n})$. Ensuite, par continuité par le haut des probabilités, on obtient

$$\lim_{n\to+\infty} F_X\Big(x+\frac{1}{n}\Big) = \lim_{n\to\infty} \mathbb{P}\Big(X \le x+\frac{1}{n}\Big) = \mathbb{P}\Big(\bigcap_{n\ge 1}\Big\{X \le x+\frac{1}{n}\Big\}\Big) = \mathbb{P}(X \le x) = F_X(x)\,.$$ □

Donnons quelques autres propriétés de la fonction de répartition, dont la démonstration est laissée en exercice.

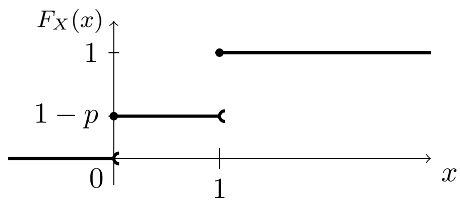
— F_X admet une limite à gauche $F_X(x^-) := \lim_{\varepsilon\to 0^+} F_X(x-\varepsilon)$ en tout $x \in \mathbb{R}$;

— $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$, donc F_X est continue en x si et seulement si $\mathbb{P}(X = x) = 0$;

— l'ensemble des points de discontinuité de F_X est au plus dénombrable.

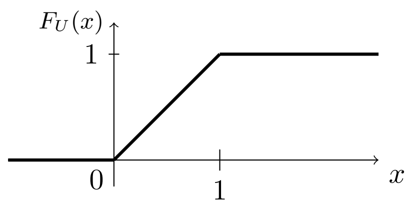
Exercice 2.12. Soit X une variable aléatoire à valeur dans $\{0,1\}$ — une telle variable aléatoire est dite de *Bernoulli* — dont la loi est donc déterminée par $p := \mathbb{P}(X = 1)$, $\mathbb{P}(X = 0) = 1 - p$. Montrer que sa fonction de répartition est donnée par la fonction suivante :

$$F_X(x) = \begin{cases} 0 & \text{si } x < 0\,;\\ 1-p & \text{si } x \in [0,1[\,;\\ 1 & \text{si } x \ge 1\,.\end{cases}$$



Soit U une variable aléatoire de loi uniforme sur $\mathcal{U}([0,1])$, c'est-à-dire telle que $\mathbb{P}_U$ est la mesure de Lebesgue sur $[0,1]$. Montrer que sa fonction de répartition est donnée par la fonction suivante :

$$F_U(x) = \begin{cases} 0 & \text{si } x < 0\,;\\ x & \text{si } x \in [0,1[\,;\\ 1 & \text{si } x \ge 1\,.\end{cases}$$



Un théorème important énonce que toute fonction F qui vérifie les propriétés (i)-(ii)-(iii) de ($*$) est la fonction de répartition d'une variable aléatoire réelle. Encore mieux : on peut construire une telle variable aléatoire comme une fonction d'une variable aléatoire de loi uniforme !

Proposition 2.13. *Soit $F : \mathbb{R} \to [0,1]$ une fonction qui vérifie les propriétés (i)-(ii)-(iii) de ($*$), c'est-à-dire F est croissante, continue à droite et de limite 0 en $-\infty$ et 1 en $+\infty$. On définit F^{-1} l'inverse généralisée de F :*

$$F^{-1}(x) = \inf\{y\,, F(y) \geq x\}\,.$$

Soit X une variable aléatoire de loi $\mathcal{U}([0,1])$ et soit $Z := F^{-1}(X)$. Alors la variable aléatoire Z a pour fonction de répartition F.

Démonstration. La première étape consiste à observer que, par définition de F^{-1} on a, pour tout $x \in]0,1[$ et tout $t \in \mathbb{R}$

$$F^{-1}(x) \leq t \quad \Longleftrightarrow \quad x \leq F(t)\,.$$

En effet, si $F(t) \geq x$, on déduit immédiatement de la définition de F^{-1} que $F^{-1}(x) \leq t$. Réciproquement, si $F(t) < x$, comme F est une fonction continue à droite, il existe $\varepsilon > 0$ tel que $F(t+\varepsilon) < x$; par monotonie, $F(z) < x$ pour tout $z \leq t+\varepsilon$ et donc $F^{-1}(x) \geq t+\varepsilon$ (toujours par définition de F^{-1}), en particulier $F^{-1}(x) > t$.

La deuxième étape consiste à observer que $\mathbb{P}(X \in]0,1[\,) = 1$ donc pour tout $t \in \mathbb{R}$

$$F_Z(t) = \mathbb{P}(Z \leq t) = \mathbb{P}(F^{-1}(X) \leq t) = \mathbb{P}(X \leq F(t)) = F(t)\,,$$

car pour tout $u \in [0,1]$ on a $\mathbb{P}(X \leq u) = \mathbb{P}(X \in [0,u]) = u$. □

2.3 Indépendance de variables aléatoires

Définition 2.14. Une famille $(X_i)_{i\in I}$ de variables aléatoires définies sur le même espace probabilisé (avec X_i à valeurs dans $(E_i, \mathcal{E}_i)$) est dite *indépendante* si la famille de tribus $\mathcal{F}_i = \sigma(X_i) = \sigma(X_i^{-1}(A), A \in \mathcal{E}_i)$ est indépendante.

De manière équivalente, la famille $(X_i)_{i\in I}$ est indépendante si pour tout sous-ensemble d'indice $J \subset I$ *fini* et tous $A_j \in \mathcal{E}_j$ on a

$$\mathbb{P}\big(\forall j \in J\,,\ X_j \in A_j\big) = \prod_{j\in J} \mathbb{P}(X_j \in A_j)\,.$$

Autrement dit, la loi jointe $\mathbb{P}_{(X_j)_{j\in J}}$ de $(X_j)_{j\in J}$ est la loi produit $\bigotimes_{j\in J} \mathbb{P}_{X_j}$ des lois marginales $\mathbb{P}_{X_j}$.

Remarque 2.15. Des v.a. $X_1, \dots, X_n$ à valeurs dans $\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_n}$ sont indépendantes si et seulement si la fonction de répartition de $X = (X_1, \dots, X_n)$ est le produit des fonctions des répartition de $X_1, \dots, X_n$.

Voici quelques propriétés de l'indépendance de variables aléatoires, dont on laisse les démonstrations en exercice.

Lemme 2.16. *Si $(X_i)_{i\in I}$ est une famille de variables aléatoires indépendantes et si $(f_i)_{i\in I}$ sont des fonctions mesurables ($f_i : (E_i, \mathcal{E}_i) \to (E_i', \mathcal{E}_i')$), alors la famille de variables aléatoires $(Y_i = f_i(X_i))_{i\in I}$ est indépendante.*

Lemme 2.17 (Regroupement par paquets)**.** *Si $(X_i)_{i\in I}$ est une famille de variables aléatoires indépendantes et si $(J_k)_{k\in K}$ sont des sous-ensembles d'indices disjoints, alors les vecteurs aléatoires $W_k = (X_j)_{j\in J_k}$ sont indépendants.*

Exemple 2.18 (Une application importante)**.** Soient $(X_i)_{i\geq 1}$ des v.a. à valeurs dans $\mathbb{R}^d$, indépendantes. On pose $S_n := \sum_{i=1}^n X_i$, avec par convention $S_0 = 0$. Alors pour toute sous-suite croissante $(n_k)_{k\geq 1}$, les v.a. $(S_{n_{k+1}} - S_{n_k})_{k\geq 1}$ sont indépendantes.

En effet, on peut écrire $S_{n_{k+1}} - S_{n_k} = f_{n_{k+1}-n_k}(X_{n_k+1}, \dots, X_{n_k})$ où $f_m(y_1, \dots, y_m) = \sum_{i=1}^m y_i$. Ainsi, les v.a. $(S_{n_{k+1}} - S_{n_k})_{k\geq 1}$ sont des fonctions de paquets disjoints de variables aléatoires indépendantes, donc sont indépendantes.

Proposition 2.19 (Loi du 0-1 de Kolmogorov, version variables aléatoires)**.** *Soit $(X_i)_{i\geq 1}$ des v.a. indépendantes. On pose $\mathcal{Q}_\infty := \bigcap_{n\geq 1} \sigma(X_i; i > n)$ la tribu asymptotique. Alors pour tout $A \in \mathcal{Q}_\infty$ on a $\mathbb{P}(A) \in \{0,1\}$.*

Exemple 2.20. Soit $(X_i)_{i\geq 1}$ des v.a. réelles indépendantes et posons $S_n := \sum_{i=1}^n X_i$, avec par convention $S_0 = 0$. Alors les événements suivants sont de probabilité 0 ou 1 :

- $A =$« la suite $(S_n)_{n\geq 0}$ converge » ;
- $B =$« $\sup_{n\geq 0} S_n = +\infty$ ».

Pour voir cela, il faut réécrire les événements pour montrer qu'ils ne dépendent pas de $X_1, \dots, X_k$, pour aucun $k \in \mathbb{N}^*$. Mais c'est assez clair si l'on écrit :

- Pour A : la suite $(S_n)_{n\geq 0}$ converge si et seulement si la suite $(S_n - S_k)_{n\geq k}$ converge ;
- Pour B : $\sup_{n\geq 0} S_n = +\infty$ si et seulement si $\sup_{n\geq k}(S_n - S_k) = +\infty$.

Exercice 2.21. Soit $(X_i)_{i\geq 1}$ des v.a. réelles indépendantes et de même loi symétrique, c'est-à-dire telles que X_i et $-X_i$ ont la même loi. On suppose que $\mathbb{P}(X_i \neq 0) > 0$.

1. Montrer qu'il existe un $\varepsilon > 0$ tel que $\mathbb{P}(X_i \geq \varepsilon) > 0$
2. Soit $k \geq 1$. On pose $A_j := \{X_{jk+1} \geq \varepsilon, \dots, X_{j(k+1)} \geq \varepsilon\}$. Montrer que p.s. il y a une infinité de A_j réalisés.
3. En déduire que, pour tout $k \geq 1$, $\mathbb{P}(\sup_n |S_n| < \frac{1}{2}\varepsilon k) = 0$, puis que $\mathbb{P}(\sup_n |S_n| < +\infty) = 0$.
4. En déduire que $\mathbb{P}(\inf S_n = -\infty) + \mathbb{P}(\sup S_n = +\infty) \geq 1$.
5. Montrer que $\mathbb{P}(\inf_n S_n = -\infty) = \mathbb{P}(\sup_n S_n = +\infty) \geq \frac{1}{2}$ puis conclure que

$$\mathbb{P}\Big(\inf_{n\geq 1} S_n = -\infty \text{ et } \sup_{n\geq 1} S_n = +\infty\Big) = 1\,.$$

Solution.

1. Si $\mathbb{P}(X_i \geq \varepsilon) = 0$ pour tout $\varepsilon > 0$, alors $\mathbb{P}(X_i > 0) = \lim_{n\to\infty} \mathbb{P}(X \geq \frac{1}{n}) = 0$. Par symétrie on aurait aussi $\mathbb{P}(X_i < 0) = 0$, d'où $\mathbb{P}(X_i \neq 0) = 0$, ce qui contredit l'énoncé.
2. Par indépendance par paquets, les événements $(A_j)_{i\geq 1}$ sont indépendants. De plus, ils sont de probabilité $\mathbb{P}(A_j) = \mathbb{P}(X_j \geq \varepsilon)^k > 0$ car les X_i sont indépendants et de même loi. Par Borel–Cantelli, on obtient donc que p.s. il y a une infinité de A_j réalisés.
3. Si un A_j est réalisé, alors $S_{j(k+1)} \geq S_{jk} + k\varepsilon$, ce qui est incompatible avec $\{S_{jk} > -\frac{1}{2}k\varepsilon\} \cap \{S_{j(k+1)} < \frac{1}{2}k\varepsilon\}$. D'après la question précédente on a donc $\mathbb{P}(\sup_n |S_n| < \frac{1}{2}\varepsilon k) = 0$. Comme $\{\sup_n |S_n| < +\infty\} = \bigcup_{k\geq 1}\{\sup_n |S_n| < \frac{1}{2}\varepsilon k\}$, par sous-additivité on a $\mathbb{P}(\sup_n |S_n| < +\infty) = 0$.
4. On a $\{\sup_n |S_n| = +\infty\} = \{\inf_n S_n = -\infty\} \cup \{\sup_n S_n = +\infty\}$. Par sous-additivité on a donc $1 \leq \mathbb{P}(\inf S_n = -\infty) + \mathbb{P}(\sup S_n = +\infty)$.
5. Par symétrie, $\mathbb{P}(\sup_n S_n = +\infty) = \mathbb{P}(\sup_n(-S_n) = +\infty) = \mathbb{P}(\inf S_n = -\infty)$. D'après la question précédente, on a $2\mathbb{P}(\sup_n S_n = +\infty) \geq 1$, c'est-à-dire $\mathbb{P}(\sup_n S_n = +\infty) \geq \frac{1}{2}$. D'après la loi du 0-1, on a donc forcément $\mathbb{P}(\sup_n S_n = +\infty) = \mathbb{P}(\inf S_n = -\infty) = 1$. Comme l'intersection de deux événements de probabilité 1 est de probabilité 1, on en déduit la conclusion voulue.

2.3.1 Construction d'une suite *infinie* de v.a. indépendantes de loi uniforme

Référence : [Ouv09, p. 52] (ou [App15, p. 713], moins détaillé). Une question importante est de savoir si l'on est capable de construire un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel est définie une suite *infinie* $(X_i)_{i\geq 1}$ de variables aléatoires *indépendantes*, toutes de loi $\mathcal{U}([0,1])$. Note : on a vu dans la Proposition 2.13 qu'à partir d'une variable aléatoire de loi uniforme sur $[0,1]$, on peut construire une variable aléatoire réelle de fonction de répartition (donc de loi) donnée.

On se place sur l'espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$, avec $\Omega = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathbb{P} =$ mesure de Lebesgue sur $[0,1]$. On a ainsi déjà une variable aléatoire $X : \omega \mapsto \omega$, de loi $\mathcal{U}([0,1])$.

Étape 1. Construction d'une suite de variables de Bernoulli de paramètre $p = 1/2$ indépendantes. L'idée centrale est que le développement dyadique de ω donne une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre $p = 1/2$. Pour $\omega \in [0,1[$, on définit de manière itérative deux suites $(D_n(\omega))_{n\geq 1}$ et $(R_n(\omega))_{n\geq 0}$ en posant $R_0(\omega) = \omega$ puis pour $n \geq 1$

$$D_n(\omega) = \lfloor 2R_{n-1}(\omega) \rfloor \in \{0,1\} \qquad R_n(\omega) = 2R_{n-1}(\omega) - D_n(\omega) \in [0,1[\,.$$

Ainsi $(D_n(\omega))_{n\geq 1}$ représente le développement dyadique de ω.

Montrons que les v.a. $(D_n)_{n\geq 1}$ ($D_n : \omega \mapsto D_n(\omega)$) sont indépendantes et de loi de Bernoulli de paramètre $p = \frac{1}{2}$. Pour cela, calculons la loi de $(D_1, \dots, D_k)$ pour $k \in \mathbb{N}^*$. Comme $(D_1, \dots, D_k) \in \{0,1\}^k$, il suffit de calculer, pour tout $(\varepsilon_1, \dots, \varepsilon_k) \in \{0,1\}^k$,

$$\mathbb{P}\Big((D_1, \dots, D_k) = (\varepsilon_1, \dots, \varepsilon_k)\Big) = \mathbb{P}\Big(\sum_{i=1}^k \frac{\varepsilon_i}{2^i} \leq \omega < \sum_{i=1}^k \frac{\varepsilon_i}{2^i} + 2^{-k}\Big) = 2^{-k},$$

où on a utilisé le fait que de connaître $(D_1, \dots, D_k)$ nous donne un intervalle dans lequel ω doit se trouver et que la mesure de Lebesgue de cet intervalle vaut 2^{-k}. On a donc la loi jointe de $(D_1, \dots, D_k)$ qui est uniforme sur $\{0,1\}^k$: on en déduit facilement que les lois marginales sont uniformes sur $\{0,1\}$, donc que les variables aléatoires sont de loi de Bernoulli de paramètre $\frac{1}{2}$, puis que $\mathbb{P}\big((D_1, \dots, D_k) = (\varepsilon_1, \dots, \varepsilon_k)\big) = \prod_{i=1}^k \mathbb{P}(X_i = \varepsilon_i)$, ce qui montre que les variables sont indépendantes.

On a donc construit une famille *infinie* $(D_n)_{n\geq 1}$ de variables aléatoires de Bernoulli($\frac{1}{2}$) indépendantes (revoir la définition de famille indépendante).

Étape 2. Construction d'une suite de variables $\mathcal{U}([0,1])$ indépendantes.
Comme $\mathbb{N}^* \times \mathbb{N}^*$ est dénombrable, il est en bijection avec $\mathbb{N}^*$, donc on peut voir la suite $(D_n)_{n\geq 1}$ comme une suite *bi-infinie* $(D_{n,k})_{n,k\geq 1}$ de variables aléatoires de Bernoulli($\frac{1}{2}$) indépendantes. On va s'en servir pour définir de nouvelles variables aléatoires $\mathcal{U}([0,1])$ indépendantes, en voyant (pour n fixé) $(D_{n,k})_{k\geq 1}$ comme le développement dyadique d'un nombre dans $[0,1[$. Pour tout $n \in \mathbb{N}^*$, on pose

$$U_n := \sum_{k=1}^\infty \frac{D_{n,k}}{2^k} = \lim_{m\to\infty} \sum_{k=1}^m \frac{X_{n,k}}{2^k},$$

où la limite existe car la série est absolument convergente ($D_{n,k} \in \{0,1\}$ pour tout $k \geq 1$).

La première remarque est que par le lemme des coalitions, les variables aléatoires $(U_n)_{n\geq 1}$ sont indépendantes, en tant que fonctions de paquets différents de la famille $(D_{n,k})_{n,k\geq 1}$ (des fonctions de $(D_{n,k})_{k\geq 1}$ pour différents n).

Il reste donc à montrer que U_n suit la loi $\mathcal{U}([0,1])$. Notons $Z_m(\omega) = \sum_{k=1}^m \frac{D_k(\omega)}{2^k}$, alors comme $(D_{n,1}, \dots, D_{n,m})$ a la même loi que $(D_1, \dots, D_k)$, c'est-à-dire la loi d'un k-uplet de variables de Bernoulli de paramètre $\frac{1}{2}$ indépendantes, on en déduit que $U_n^{(m)} := \sum_{k=1}^m \frac{D_{n,k}}{2^k}$ est de même loi que Z_m. (Rappel : si X et X' ont même loi, alors $f(X)$ et $f(X')$ ont même loi.) Ainsi, $U_n := \lim_{m\to\infty} U_n^{(m)}$ et $Z := \lim_{m\to\infty} Z_m$ ont la même loi. En effet, pour $t \in \mathbb{R}$, en notant que $\{U_n^{(m)} \leq t\}$ est une suite décroissante d'événements et $\{U_n \leq t\} = \bigcap_{m\geq 1}\{U_n^{(m)} \leq t\}$ (et de même pour Z_m, Z), on obtient

$$\mathbb{P}(U_n \leq t) = \lim_{m\to\infty} \mathbb{P}(U_n^{(m)} \leq t) = \lim_{m\to\infty} \mathbb{P}(Z_m \leq t) = \mathbb{P}(Z \leq t).$$

(Sinon, c'est une conséquence du fait que la convergence p.s. implique la convergence en loi.) Mais $Z(\omega) = \omega$ pour tout $\omega \in [0,1[$, par définition du développement dyadique $D_k(\omega)$, donc Z est de loi uniforme sur $[0,1]$. On en conclut que $U_n \sim \mathcal{U}([0,1])$.

2.4 Variables aléatoires discrètes

Définition 2.22. Soit X une variable aléatoire à valeurs dans $(E, \mathcal{E})$. La variable aléatoire X est dite *discrète* s'il existe $\tilde{E} \subseteq E$ *dénombrable* (fini ou infini) tel que $\mathbb{P}(X \in \tilde{E}) = 1$.

Remarque 2.23. On peut alors voir la loi $\mathbb{P}_X$ de X comme une probabilité sur $\tilde{E}$. Comme $\tilde{E}$ est dénombrable, $\mathbb{P}_X$ est caractérisée par ses valeurs sur les singletons, c'est-à-dire par la fonction

$$p_X : \begin{array}{l} \tilde{E} \to [0,1], \\ x \mapsto p_X(x) := \mathbb{P}(X = x), \end{array}$$

appelée *densité discrète de* X. Ainsi, pour déterminer la loi d'une variable aléatoire discrète, il faut : (i) déterminer $\tilde{E}$; (ii) déterminer la densité discrète de X, c'est-à-dire $\mathbb{P}(X = x)$ pour tout $x \in \tilde{E}$.

Si $E = \mathbb{R}$, alors on peut définir la fonction de répartition d'une variable aléatoire X. Si X est discrète, en se rappelant que $\mathbb{P}(X = x) = F_X(x) - F_X(x^-)$, on observe que l'on peut prendre $\tilde{E}$ l'ensemble des points de discontinuités de X, et la fonction F_X est constante sur chacun des morceaux de $\mathbb{R} \setminus \tilde{E}$, la hauteur des sauts de discontinuité en $x \in \tilde{E}$ étant égale à $\mathbb{P}(X = x)$.

2.4.1 Indépendance de variables aléatoires discrètes

Proposition 2.24. *Des v.a.* discrètes $X_1, \ldots, X_n$*, à valeurs respectivement dans* $E_1, \ldots, E_n$*, sont indépendantes si et seulement si*

$$\forall (x_1, \ldots, x_n) \in E_1 \times \cdots \times E_n\,, \qquad \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)\,.$$

Exercice 2.25. Montrer que si X_1, X_2 sont deux v.a. discrètes *indépendantes,* à valeurs réelles, alors $X_1 + X_2$ est discrète, de densité discrète $\mathbb{P}(X_1 + X_2 = z) = \sum_{x \in \mathbb{R}} \mathbb{P}(X_1 = x)\mathbb{P}(X_2 = z - x)$, où la somme est restreinte aux $x \in \mathbb{R}$ tels que $\mathbb{P}(X_1 = x)\mathbb{P}(X_2 = z - x) > 0$.

2.4.2 Exemples important de variables aléatoires discrètes

Bernoulli. Une v.a. X est dite de *Bernoulli* de paramètre $p \in [0, 1]$ (on note $X \sim \mathrm{Ber}(p)$) si X est à valeurs dans $\{0, 1\}$ et a pour densité discrète $\mathbb{P}(X = 0) = 1 - p$, $\mathbb{P}(X = 1) = p$.

On interprète X comme le résultat d'une tentative, où $X = 1$ désigne un succès et $X = 0$ un échec.

Binomiale. Une v.a. X est dite *Binomiale* de paramètres $n \in \mathbb{N}, p \in [0, 1]$ (on note $X \sim \mathrm{Bin}(n, p)$) si X est à valeurs dans $\{0, 1, \ldots, n\}$ et a pour densité discrète

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \qquad \text{pour } k \in \{0, 1, \ldots, n\}\,.$$

Si on pose $X =$ nombre de succès parmi n tentatives indépendantes de probabilité de succès $p \in [0, 1]$, alors $X \sim \mathrm{Bin}(n, p)$. Autrement dit, une variable de loi $\mathrm{Bin}(n, p)$ a la même loi que la somme de n variables aléatoires indépendantes de loi $\mathrm{Ber}(p)$.

Géométrique. Une v.a. X est dite *Géométrique* de paramètre $p \in]0, 1[$ (on note $X \sim \mathrm{Geo}(p)$) si X est à valeurs dans $\mathbb{N}^*$ et a pour densité discrète

$$\mathbb{P}(X = k) = (1 - p)^{k-1} p \qquad \text{pour } k \in \mathbb{N}^*\,.$$

Si $X =$ numéro du premier succès lors de tentatives indépendantes de probabilité de succès $p \in [0, 1]$, alors $X \sim \mathrm{Geo}(p)$. Notons que l'on pourrait définir par convention $X = 1$ si $p = 1$ (et $X = \infty$ si $p = 0$).

Remarque : on pourrait aussi compter le nombre X' d'échecs avant le premier succès, c'est-à-dire $X' = X - 1$ (on l'appelera aussi géométrique, donc il faut faire attention). Dans ce cas, X' est à valeurs dans $\mathbb{N}$ et $\mathbb{P}(X' = k) = (1 - p)^k p$ pour $k \in \mathbb{N}$.

Proposition 2.26. *La loi géométrique est l'unique loi de v.a. à valeurs dans* $\mathbb{N}^*$ *qui vérifie la propriété d'*absence de mémoire *suivante :*

$$\forall m, n \in \mathbb{N}, \qquad \mathbb{P}(X > m + n \mid X > m) = \mathbb{P}(X > n)\,.$$

Démonstration. Si $X \sim \mathrm{Geo}(p)$, alors $\mathbb{P}(X > n) = (1 - p)^n$ pour tout $n \in \mathbb{N}$. La propriété d'absence de mémoire en découle directement.

Si X possède la propriété d'absence de mémoire, alors pour tout $n \in \mathbb{N}^*$, $\mathbb{P}(X > n + 1 \mid X > n) = \mathbb{P}(X > 1) =: q$. Par récurrence, on en déduit que $\mathbb{P}(X > n) = q^n$ pour tout $n \in \mathbb{N}^*$, dont on déduit que $X \sim \mathrm{Geo}(1 - q)$. □

Exercice 2.27. Soient X, Y deux variables aléatoires géométriques indépendantes, de paramètres respectifs $p, q \in]0, 1[$. Montrer que $Z := \min(X, Y)$ suit une loi géométrique, de paramètre $r = 1 - (1 - p)(1 - q)$.

Poisson. Une v.a. X est dite de *Poisson* de paramètres $\lambda > 0$ (on note $X \sim \mathrm{Poi}(\lambda)$) si X est à valeurs dans $\mathbb{N}$ et a pour densité discrète

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad \text{pour } k \in \mathbb{N}\,.$$

Exercice 2.28. Soient X, Y deux variables aléatoires de Poisson indépendantes, de paramètres respectifs $\lambda, \mu > 0$. Montrer que $Z := X + Y$ suit une loi de Poisson, de paramètre $\lambda + \mu$.

Les variables aléatoires de Poisson apparaissent naturellement comme de bonnes approximations pour la variable aléatoire « nombre de succès lors d'un grand nombre de tentatives indépendantes de faible probabilité de succès ». On a par exemple le résultat suivant, appelé inégalité de Le Cam (référence [GK19, p. 208], [BCD21, p. 185]).

Proposition 2.29. *Soient* $(X_i)_{1 \le i \le n}$ *des v.a. indépendantes, avec* $X_i \sim \mathrm{Ber}(p_i)$*. Soit* $X \sim \mathrm{Poi}(\lambda)$*, où* $\lambda = \sum_{i=1}^n p_i$*. Alors on a*

$$\sum_{k \in \mathbb{N}} \left| \mathbb{P}(X = k) - \mathbb{P}\Big(\sum_{i=1}^n X_i = k \Big) \right| \le 2 \sum_{i=1}^n p_i^2\,.$$

Un corollaire de ce résultat (si $p_i = \frac{\lambda}{n}$ pout tout i) : si $Y_n \sim \mathrm{Bin}(n, \frac{\lambda}{n})$ et $X \sim \mathrm{Poi}(\lambda)$, alors

$$\forall\, k \in \mathbb{N} \qquad \lim_{n \to \infty} \mathbb{P}(Y_n = k) = \mathbb{P}(X = k)\,.$$

Exercice 2.30. Obtenir la conclusion ci-dessus par un calcul direct.

Uniforme sur un ensemble fini. Une v.a. X à valeur dans un ensemble E fini est dite *Uniforme sur* E (on note $X \sim \mathrm{Unif}(E)$) si elle a pour densité discrète

$$\mathbb{P}(X = x) = \frac{1}{|E|} \qquad \text{pour tout } x \in E\,.$$

Exercice 2.31. Soit X une v.a. à valeur dans E fini telle que $\mathbb{P}(X = x)$ ne dépend pas de $x \in E$. Montrer que X est de loi uniforme sur E.

Exercice 2.32. Montrer qu'une variable aléatoire $X = (X_1, X_2)$ est uniforme sur $E = E_1 \times E_2$ si et seulement si X_1, X_2 sont indépendantes de lois uniformes sur E_1, E_2 respectivement.

Exemple 2.33. Une v.a. X_n de loi uniforme sur $E = \mathfrak{S}_n$ est appelée permutation uniforme. On peut montrer (voir [BCD21, p. 74]) que $\mathbb{P}(X_n$ possède exactement m points fixes$) = \frac{q_{n-m}}{m!}$, où $q_k = \mathbb{P}(X_k$ n'a aucun point fixe$) = \sum_{i=0}^k \frac{(-1)^i}{i!}$. En particulier, pour tout m fixé :

$$\lim_{n \to \infty} \mathbb{P}(X_n \text{ possède exactement } m \text{ points fixes}) = \frac{e^{-1}}{m!}\,.$$

La loi de Poisson (ici de paramètre 1) apparaît donc encore une fois ! Elle correspond ici aussi à un « nombre de succès lors d'un grand nombre de tentatives de faible probabilité de succès » (ici la i-ème tentative est « i est un point fixe de X_n », de probabilité de succès $1/n$), mais cette fois les tentatives ne sont pas indépendantes...

2.5 Variables aléatoires à densité (ou absolument continues)

Définition 2.34. Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d$. On dit que X admet une densité $f : \mathbb{R}^d \to \mathbb{R}_+$ (ou plutôt que $\mathbb{P}_X$ admet une densité f par rapport à la mesure de Lebesgue) si f est mesurable et si

$$\forall A \in \mathcal{B}(\mathbb{R}^d)\,, \qquad \mathbb{P}_X(A) = \mathbb{P}(X \in A) = \int_A f(x)dx\,.$$

C'est équivalent à

$$\forall I_1, \dots, I_d\,, \qquad \mathbb{P}(X \in I_1 \times \cdots \times I_d) = \int_{I_1 \times \cdots \times I_d} f(x_1, \dots, x_d) dx_1 \cdots dx_d\,,$$

et à

$$\forall (t_1, \dots, t_d) \in \mathbb{R}^d\,, \qquad F_X(t_1, \dots, t_d) = \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_d} f(x_1, \dots, x_d) dx_1 \cdots dx_d\,.$$

Remarque 2.35. De façon générale, on peut écrire $\mathbb{P}(X \in A) = \mathbb{P}_X(A) = \int_A d\mathbb{P}_X(x)$. L'hypothèse « à densité » correspond donc à dire que l'on peut écrire « $d\mathbb{P}_X(x) = f(x)dx$ » — c'est lié au théorème de Radon–Nikodym–Lebesgue, pour ceux qui connaissent.

Propriétés des densités :

— la densité est définie *à ensemble de mesure nulle près* (en particulier, elle n'est pas unique) ;

— $\int_{\mathbb{R}^d} f(x)dx = 1$.

Remarque 2.36. Si X est une v.a. à densité, à valeurs dans $\mathbb{R}^d$, alors, pour tout ensemble $H \subset \mathbb{R}^d$ de mesure de Lebesgue d-dimensionnelle nulle (par exemple n'importe quel sous-espace affine de dimension strictement inférieure à d), alors $\mathbb{P}(X \in H) = 0$.

Exercice 2.37. Soit X une v.a. à valeurs dans $\mathbb{R}^d$ de densité $f \mathbb{1}_U$ pour un ouvert U et soit $\varphi : U \to V$ un C^1 difféomorphisme. Montrer que la v.a. $Y = \varphi(X)$ a pour densité $f_Y(y) = \frac{1}{|\det(\mathrm{Jac}_\varphi)|} f(\varphi^{-1}(y)) \mathbb{1}_V(y)$.

En particulier, si X est à valeurs dans $\mathbb{R}$ de densité f, et si $a \neq 0$, $b \in \mathbb{R}$, alors la v.a. $Y = aX + b$ a pour densité $f_Y(y) = \frac{1}{|a|} f(\frac{1}{a}(y - b))$.

2.5.1 Cas réel

Une v.a. réelle X admet pour densité f si et seulement si

$$\forall\, t \in \mathbb{R}\,, \qquad F_X(t) = \int_{-\infty}^{t} f(x)dx\,. \tag{2.1}$$

Remarque 2.38. Si X est une v.a. réelle de densité f, alors :

— F_X est continue, en particulier $\mathbb{P}(X = x) = 0$ pour tout $x \in \mathbb{R}$;

— $\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \in]a, b[\,) = \int_a^b f(x)dx$;

— F_X est dérivable presque partout, de dérivée $F'_X(x) = f(x)$ p.p.

Attention, la réciproque du dernier point est fausse ! Il existe une fonction de répartition continue, dérivable presque partout, de dérivée nulle p.p. (exemple $F_X =$ l'escalier du diable, voir [GK19, p. 125]). Par contre, on a le critère suivant pour déterminer la densité d'une variable aléatoire à partir de sa fonction de répartition.

Lemme 2.39. *Soit X une variable aléatoire réelle, de fonction de répartition F_X. Si F_X est* continue *et C^1* par morceaux, *alors X admet une densité f donnée par $f(x) = F'_X(x)$.*

Exercice 2.40. Soit X une variable aléatoire réelle, de densité $f(x) = |x| \mathbb{1}_{[-1,1]}$. On pose $Y = \max(X, 0)$. Calculer la fonction de répartition de Y. Montrer que Y n'est ni une variable aléatoire discrète, ni une variable aléatoire à densité.

2.5.2 Indépendance et densité

Proposition 2.41. *Des variables aléatoires $X_1, \dots, X_n$ à densité (à valeurs resp. dans $\mathbb{R}^{d_1}, \dots, \mathbb{R}^{d_n}$) de densités respectives $f_{X_1}, \dots, f_{X_n}$, sont indépendantes si et seulement si $X = (X_1, \dots, X_n)$ est à densité, de densité donnée par*

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

pour presque tout $(x_1, \dots, x_n) \in \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$.

Cette proposition possède deux implications :

— Si $X_1, \ldots, X_n$ sont indépendantes et à densité, alors $X = (X_1, \ldots, X_n)$ est à densité, de densité donnée par le produit des densités marginales ;

— Si $X = (X_1, \ldots, X_n)$ est à densité, de densité donnée par le produit de fonctions de $x_1, \ldots, x_n$, alors les $X_1, \ldots, X_n$ sont indépendantes et à densité (donnée par les densités marginales).

Exemple 2.42. Si X, Y sont indépendantes, à valeurs dans $\mathbb{R}^d$ et à densité, alors $\mathbb{P}(X = Y) = 0$. En effet, $\mathbb{P}(X = Y) = \mathbb{P}((X, Y) \in \Delta)$ avec $\Delta = \{(x, x), x \in \mathbb{R}^d\}$: comme (X, Y) est à densité dans $\mathbb{R}^{2d}$ et que Δ est de mesure de Lebesgue $2d$-dimensionnelle nulle (c'est un espace vectoriel de dimension d), on a $\mathbb{P}((X, Y) \in \Delta) = 0$.

Proposition 2.43. *Si X_1, X_2 sont deux v.a. à densité* indépendantes, *de densités respectives f_1, f_2, alors $X_1 + X_2$ est à densité, de densité*

$$f_{X_1+X_2}(x) = f_{X_1} * f_{X_2}(x) = \int_{\mathbb{R}} f_{X_1}(x) f_{X_2}(y - x) dx\,.$$

Démonstration. Calculons la fonction de répartition de $X_1 + X_2$: en notant que $\{X_1 + X_2 \le t\} = \{(X_1, X_2) \in D_t\}$, avec $D_t = \{(x_1, x_2) \in \mathbb{R}^2\,, x_1 + x_2 \le t\}$ on a

$$F_{X_1+X_2}(t) = \mathbb{P}\big((X_1, X_2) \in D_t\big) = \int_{D_t} f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2\,,$$

où on a utilisé l'indépendance de X_1, X_2 pour obtenir la densité $f_{X_1,X_2} = f_{X_1} f_{X_2}$. Maintenant, en appliquant Fubini–Tonelli puis avec le changement de variable $x_2 \to y - x_1$ on obtient

$$\begin{aligned} F_{X_1+X_2}(t) &= \int_{\mathbb{R}} \bigg(\int_{-\infty}^{t-x_1} f_{X_1}(x_1) f_{X_2}(x_2) dx_2 \bigg) dx_1 = \int_{\mathbb{R}} \bigg(\int_{-\infty}^{t} f_{X_1}(x_1) f_{X_2}(y - x_1) dy \bigg) dx_1 \\ &= \int_{-\infty}^{t} \bigg(\int_{\mathbb{R}} f_{X_1}(x_1) f_{X_2}(y - x_1) dx_1 \bigg) dy\,, \end{aligned}$$

où on a de nouveau utilisé Fubini–Tonelli pour la dernièr égalité. Cela conclut la démonstration de la proposition (voir (2.1)). □

2.5.3 Exemples importants de variables aléatoires à densité

Uniforme sur un compact. Soit $K \subseteq \mathbb{R}^d$ un compact de mesure de Lebesgue (d-dimensionnelle) strictement positive. Une v.a. X est dite *Uniforme sur K* (on note $X \sim \mathcal{U}(K)$) si X admet pour densité

$$f(x) = \frac{1}{\mathrm{Vol}(K)} \mathbb{1}_K(x)\,, \qquad x \in \mathbb{R}^d\,.$$

On a $X \sim \mathcal{U}(K)$ si (et seulement si) pour tout borélien B on a $\mathbb{P}(X \in B) = \frac{\mathrm{Vol}(B \cap K)}{\mathrm{Vol}(K)}$.

Un exemple important (essentiel) est le cas d'une v.a. de loi uniforme sur $[0, 1]$ ($d = 1$ et $K = [0, 1]$).

Exercice 2.44. Soit $X \sim \mathcal{U}(K)$. Montrer que pour $\lambda \in \mathbb{R}^*$ et $z \in \mathbb{R}^d$ la v.a. $\lambda X + z$ est de loi uniforme sur $K' = \{\lambda x + z, x \in K\}$.

Exponentielle. Une v.a. X réelle est dite *Exponentielle* de paramètre $\lambda > 0$ (on note $X \sim \mathcal{E}(\lambda)$) si X admet pour densité

$$f(x) = \lambda e^{-\lambda x}\, \mathbb{1}_{\mathbb{R}_+}(x)\,, \qquad x \in \mathbb{R}\,.$$

Sa fonction de répartition est $F_X(t) = (1 - e^{-\lambda t})\, \mathbb{1}_{\mathbb{R}_+}(t)$.

La variable exponentielle est l'analogue continu de la variable géométrique.

Proposition 2.45. *La loi exponentielle est l'unique loi de v.a. à valeurs dans $\mathbb{R}_+$ qui vérifie la propriété d'*absence de mémoire *suivante :*

$$\forall s, t \ge 0, \qquad \mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s)\,.$$

Démonstration. En exercice. Indication : utiliser que $\mathbb{P}(X > t) = e^{-\lambda t}$ pour $t \ge 0$. □

Exercice 2.46. Soient X, Y deux variables aléatoires exponentielles indépendantes, de paramètres respectifs $\lambda, \mu > 0$. Montrer que $Z := \min(X, Y)$ suit une loi exponentielle, de paramètre $\lambda + \mu$.

Normale. Une v.a. X réelle est dite *Normale (ou gaussienne) standard* (on note $X \sim \mathcal{N}(0,1)$) si X admet pour densité

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad x \in \mathbb{R}.$$

On dit que $X \sim \mathcal{N}(\mu, \sigma^2)$ suit une loi normale d'espérance $\mu \in \mathbb{R}$ et de variance $\sigma^2 \in \mathbb{R}_+^*$ si X admet pour densité

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad x \in \mathbb{R}.$$

Autrement dit,

$$X = \frac{Y-\mu}{\sigma} \sim \mathcal{N}(0,1) \quad \Longleftrightarrow \quad Y = \sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2).$$

Les variables gaussiennes possèdent la propriété cruciale suivante. (le théorème central limite montre en réalité que ce sont les seules variables admettant une variance à posséder une telle propriété).

Proposition 2.47. *Soient X, Y deux v.a.* indépendantes, *de lois $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$. Alors $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Démonstration. En exercice. Se ramener au cas $\mu_1 = \mu_2 = 0$. Utiliser $f_{X+Y} = f_X * f_Y$ (calcul pas facile). La preuve sera plus facile avec la fonction caractéristique. □

2.5.4 Autres exemples de variables aléatoires à densité

Gamma. Une v.a. X réelle est dite de loi *Gamma* de paramètre $a > 0, \lambda > 0$ (on note $X \sim \mathrm{Gam}(a, \lambda)$) si X admet pour densité

$$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x), \qquad x \in \mathbb{R},$$

où $\Gamma(a)$ est la fonction gamma :

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx.$$

Les variables gamma sont des généralisations de la variable exponentielle : $\mathrm{Gam}(1, \lambda) \sim \mathcal{E}(\lambda)$. Notons que $X \sim \mathrm{Gam}(a, \lambda)$ si et seulement si $\frac{1}{\lambda} X \sim \mathrm{Gam}(a, 1)$: le paramètre λ est simplement un facteur d'échelle.

Proposition 2.48. *Soient X, Y deux v.a.* indépendantes, *de lois $X \sim \mathrm{Gam}(a, \lambda)$, $Y \sim \mathrm{Gam}(b, \lambda)$. Alors $X + Y \sim \mathrm{Gam}(a + b, \lambda)$.*

Démonstration. D'après le commentaire ci-dessus, on ne traite que le cas $\lambda = 1$. Calculons $f_{X+Y}(x)$ pour $x > 0$ (la densité est évidemment nulle pour $x < 0$) :

$$\begin{aligned} f_{X+Y}(x) = f_X * f_Y(x) &= \int_0^x \frac{1}{\Gamma(a)} t^{a-1} e^{-t} \frac{1}{\Gamma(b)} (x-t)^{b-1} e^{-(x-t)} dt \\ &= \frac{1}{\Gamma(a)\Gamma(b)} \left(\int_0^1 u^{a-1} (1-u)^{b-1} du \right) x^{a+b-1} e^{-x} \end{aligned}$$

où on a fait un changement de variable $u = t/x$ pour la deuxième égalité. Cette densité devant être d'intégrale 1, la constante précédant $x^{a+b-1}e^{-x}$ doit être égale à $1/\Gamma(a+b)$, donc f_{X+Y} est la densité de la loi $\mathrm{Gam}(a + b, 1)$. □

Au passage, la démonstration ci-dessus donne que pour $a, b > 0$ on a

$$B(a, b) := \int_0^1 u^{a-1} (1-u)^{b-1} du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

La fonction $(a, b) \mapsto B(a, b)$ est appelée fonction beta.

Beta. Une v.a. X rélle est dite de loi *Beta* de paramètre $a, b > 0$, (on note $X \sim \mathrm{Beta}(a, b)$) si X admet pour densité

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}\, \mathbb{1}_{[0,1]}(x)\,, \qquad x \in \mathbb{R}\,,$$

où $B(a,b) := \int_0^1 u^{a-1}(1-u)^{b-1}du$ est la fonction beta. La loi beta est une généralisation de la loi uniforme sur $[0, 1]$. Elle apparaît naturellement dans de nombreux contextes.

Exercice 2.49. Soient X, Y deux variables aléatoires indépendantes, de lois $X \sim \mathrm{Gam}(a, \lambda)$, $Y \sim \mathrm{Gam}(b, \lambda)$. On pose $W = X + Y$ et $Z = \frac{X}{X+Y}$. Alors W et Z sont indépendantes, de loi $W \sim \mathrm{Gam}(a + b, \lambda)$, $Z \sim \mathrm{Beta}(a, b)$.

Cauchy. Une v.a. X réelle est dite de loi de *Cauchy standard* (on note $X \sim \mathrm{Cauchy}(0, 1)$) si X admet pour densité

$$f(x) = \frac{1}{\pi(1+x^2)}\,, \qquad x \in \mathbb{R}.$$

On dit que X suit une loi de Cauchy de paramètre $m \in \mathbb{R}$ (paramètre de position), $\alpha > 0$ (paramètre d'échelle), on note $X \sim \mathrm{Cauchy}(m, \alpha)$, si X admet pour densité

$$f(x) = \frac{\alpha}{\pi(\alpha^2 + (x-m)^2)}\,, \qquad x \in \mathbb{R}\,.$$

Autrement dit,

$$X = \frac{Y-m}{\alpha} \sim \mathrm{Cauchy}(0,1) \quad \Longleftrightarrow \quad Y = \alpha X + m \sim \mathrm{Cauchy}(m, \alpha)\,.$$

La loi de Cauchy possède des propriétés importantes et apparaît elle-aussi dans de nombreux contextes.

Proposition 2.50. *Soient X, Y deux v.a.* indépendantes, *de lois respectives $X \sim \mathrm{Cauchy}(m_1, \alpha_1)$, $Y \sim \mathrm{Cauchy}(m_2, \alpha_2)$. Alors $X + Y \sim \mathrm{Cauchy}(m_1 + m_2, \alpha_1 + \alpha_2)$.*

Démonstration. La preuve peut se faire en utilisant la formule $f_{X+Y} = f_X * f_Y$ (voir [BCD21, p. 262]), mais elle sera plus facile en utilisant la fonction caractéristique. □

Exercice 2.51. Soit $X \sim \mathrm{Cauchy}(0, 1)$. Montrer que l'on a aussi $1/X \sim \mathrm{Cauchy}(0, 1)$.

Exercice 2.52. Soient X, Y deux v.a. indépendantes de loi $\mathcal{N}(0, 1)$. Montrer que $Z = X/Y$ est de loi Cauchy standard.

3 Espérance, variance, inégalités probabilistes

3.1 Définitions et premières propriétés

Définition 3.1 (Espérance). Soit X une variable aléatoire *réelle* d'un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$.
• Si X est à valeurs positives, on définit l'espérance de X :

$$\mathbb{E}(X) = \int_\Omega X(\omega) d\mathbb{P}(\omega) \in [0, +\infty]\,.$$

• Si X est à valeurs réelles, on dit que X admet une espérance si $\mathbb{E}(X^+) < +\infty$ ou $\mathbb{E}(X^-) < +\infty$, où $X^+ = \max(X, 0)$ et $X^- = \max(-X, 0)$ sont les parties positives et négatives de X. On définit alors l'espérance de X comme :

$$\mathbb{E}(X) = \int_\Omega X(\omega) d\mathbb{P}(\omega) = \mathbb{E}(X^+) - \mathbb{E}(X^-) \in [-\infty, +\infty]\,.$$

Note, si $X = (X_1, \ldots, X_d)$ est une v.a. à valeurs dans $\mathbb{R}^d$, on pose $\mathbb{E}(X) = (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_d))$.

Remarque 3.2. Si $\mathbb{E}(X^+) < +\infty$ *et* $\mathbb{E}(X^-) < +\infty$, alors l'espérance de X est finie, $|\mathbb{E}(X)| < +\infty$. On note que, par linéarité de l'intégration pour des fonctions positives on a $\mathbb{E}(|X|) = \mathbb{E}(X^+) + \mathbb{E}(X^-)$, car $|X| = X^+ + X^-$. On en déduit que X est d'espérance finie (= est intégrable) si et seulement si $\mathbb{E}(|X|) < +\infty$.

Voici quelques propriétés importantes de l'espérance (qui découle de la théorie de l'intégration) :

— **Monotonie**. Si X, Y admettent une espérance et si $X \leq Y$ p.s., alors $\mathbb{E}(X) \leq \mathbb{E}(Y)$;

— **Linéarité**. Si X, Y admettent une espérance *finie*, alors pour tous $a, b \in \mathbb{R}$ on a $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$; Si X, Y sont positives, alors pour tous $a, b \in \mathbb{R}_+$, $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$;

— **Inégalité triangulaire**. Si X admet une espérance, on a $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$;

— **Positivité**. Si X est *positive* et telle que $\mathbb{E}(X) = 0$, alors $X = 0$ p.s..

Exemple 3.3. Si $X \sim \mathrm{Ber}(p)$, alors on a $\mathbb{E}(X) = p$. En effet, une variable de Bernoulli peut toujours s'écrire comme une fonction indicatrice : $X = \mathbb{1}_A$, où $A = \{\omega \in \Omega\,, X(\omega) = 1\}$; le paramètre p est clairement égal à $\mathbb{P}(X = 1) = \mathbb{P}(A)$. Alors $\mathbb{E}(X) = \int_\Omega \mathbb{1}_A d\mathbb{P}(\omega) = \mathbb{P}(A)$.

Exemple 3.4. La linéarité de l'espérance peut avoir des applications finalement assez importantes ! Elle permet notamment de calculer certaines espérances sans avoir à déterminer la loi de la variable aléatoire. Soit par exemple $\Omega = \mathfrak{S}_n$ l'ensemble des permutations de $\{1, \ldots, n\}$, muni de la probabilité uniforme $\mathbb{P}$. On pose X la v.a. $X(\omega) =$ « nombre de points fixes de ω ». On peut alors écrire $X = \sum_{i=1}^n \mathbb{1}_{A_i}$, où $A_i =$ « i est un point fixe de ω » $= \{\omega \in \mathfrak{S}_n, \omega(i) = i\}$. Par linéarité, on a alors $\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(\mathbb{1}_{A_i}) = \sum_{i=1}^n \mathbb{P}(A_i) = n \times \frac{1}{n} = 1$.

Définition-Proposition 3.5. On note $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ (où plus simplement $\mathcal{L}^1(\mathbb{P})$ ou $\mathcal{L}^1$) l'ensemble des variables aléatoires réelles d'espérance finie : $\mathcal{L}^1 = \{X : \Omega \to \mathbb{R} \text{ mesurable, } \mathbb{E}(|X|) < +\infty\}$. L'espace $\mathcal{L}^1$ est un espace vectoriel réel.

3.1.1 Quelques théorèmes d'intégration en termes probabilistes

Théorème 3.6 (Convergence monotone). *Si $(X_n)_{n\geq 1}$ est une suite croissante de v.a., c'est-à-dire $X_{n+1} \geq X_n$ p.s. $\forall\, n$, et si $\mathbb{E}(X_1) > -\infty$ alors $\mathbb{E}(\lim_{n\to\infty} X_n) = \lim_{n\to\infty} \mathbb{E}(X_n)$.*

Lemme 3.7 (Lemme de Fatou). *Si $(X_n)_{n\geq 0}$ est une suite de v.a.* positives*, alors $\mathbb{E}(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} \mathbb{E}(X_n)$.*

Exemple 3.8. Soit $(X_n)_{n\geq 1}$ une suite de v.a. *positives* telles que $X = \lim_{n\to\infty} X_n$ existe p.s. Si on a $\lim_{n\to\infty} \mathbb{E}(X_n) = 0$, alors par le lemme de Fatou on a $\mathbb{E}(\liminf_{n\to\infty} X_n) = 0$. La v.a. X étant positive, on a $X = 0$ p.s., c'est-à-dire $\lim_{n\to\infty} X_n = 0$ p.s.

Théorème 3.9 (Convergence dominée). *Soit $(X_n)_{n\geq 1}$ une suite de v.a. dominée par une même v.a. $Y \in \mathcal{L}^1$ positive, c'est-à-dire telle que $|X_n| \leq Y$ pour tout $n \geq 1$ avec $\mathbb{E}(Y) < \infty$. Alors si $X = \lim_{n\to\infty} X_n$ p.s., on a $\mathbb{E}(X) = \lim_{n\to\infty} \mathbb{E}(X_n) < \infty$.*

Une identité utile. Voici une identité qui permet de relier l'espérance d'une variable aléatoire positive à sa fonction de répartition.

Lemme 3.10. *Si X est une variable aléatoire* positive*, alors $\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X > t)dt$.*

Démonstration. Il suffit de remarquer que l'on a $X = \int_0^\infty \mathbb{1}_{\{t<X\}} dt$. En appliquant le théorème de Fubini–Tonelli, on obtient

$$\mathbb{E}(X) = \int_0^\infty \mathbb{E}(\mathbb{1}_{\{X>t\}})dt = \int_0^\infty \mathbb{P}(X > t)dt\,.$$

□

Ce résultat permet par exemple d'obtenir :

— Si X est à valeurs dans $\mathbb{N}$, $\mathbb{E}(X) = \sum_{n\in\mathbb{N}} \mathbb{P}(X > n)$;

— La v.a. X est dans $\mathcal{L}^1$ si et seulement si $t \mapsto \mathbb{P}(|X| > t)$ est intégrable.

3.2 Le théorème de transfert

D'après le théorème de transfert, pour X une v.a. réelle, on peut réécrire la formule de l'espérance comme $\mathbb{E}(X) = \int_{\mathbb{R}} x d\mathbb{P}_X(x)$. De manière générale, on peut énoncer le résultat suivant.

Théorème 3.11 (Théorème de transfert). *Soit X une variable aléatoire à valeurs dans $(E, \mathcal{E})$. Alors pour toute fonction $g : E \to \mathbb{R}$ mesurable telle que $g(X)$ admet une espérance, on a*

$$\mathbb{E}(g(X)) = \int_E g(x) d\mathbb{P}_X(x) .$$

On peut décliner cette formule de transfert pour les deux grandes classes de v.a. : les v.a. discrètes et à densité.

Théorème 3.12 (Théorème de transfert, version discrète). *Soit X une variable aléatoire à valeurs dans E dénombrable et soit $g : E \to \mathbb{R}$ une fonction mesurable. Alors $g(X)$ admet une espérance si et seulement si $\sum_{x\in E} g^+(x)\mathbb{P}(X = x) < +\infty$ ou $\sum_{x\in E} g^-(x)\mathbb{P}(X = x) < +\infty$, et on a*

$$\mathbb{E}(g(X)) = \sum_{x\in E} g(x)\mathbb{P}(X = x) .$$

Théorème 3.13 (Théorème de transfert, version absolument continue). *Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d$ admettant une densité f et soit $g : \mathbb{R}^d \to \mathbb{R}$ une fonction mesurable. Alors $g(X)$ admet une espérance si et seulement si $\int_{\mathbb{R}^d} g^+(x)f(x)dx < +\infty$ ou $\int_{\mathbb{R}^d} g^-(x)f(x)dx < +\infty$, et on a*

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}^d} g(x)f(x)dx .$$

Une application intéressante du théorème de transfert est qu'il permet de déterminer la loi d'une v.a.

Lemme 3.14. *La v.a. X à valeurs dans E est de loi μ si et seulement si pour toute fonction $g : E \to \mathbb{R}$ mesurable positive (ou mesurable bornée) on a $\mathbb{E}(g(X)) = \int_E g(x)d\mu(x)$.*

Démonstration. Si X est de loi μ, alors on a $\mathbb{E}(g(X)) = \int_E g(x)d\mu(x)$, grâce au théorème de transfert.

Pour la réciproque : si $\mathbb{E}(g(X)) = \int_E g(x)d\mu(x)$ pour toute fonction g mesurable positive, alors c'est en particulier vrai pour toute fonction $\mathbb{1}_A$, pour $A \in \mathcal{E}$. Cela donne alors $\mathbb{E}(\mathbb{1}_A(X)) = \mathbb{P}(X \in A) = \int_A d\mu(x) = \mu(A)$, c'est-à-dire $\mathbb{P}_X = \mu$. □

Remarque 3.15. Dans le Lemme 3.14, dans le cas $E = \mathbb{R}^d$, on pourrait en réalité remplacer la classe des fonctions test pour laquelle vérifier la formule de transfert par une classe plus restreinte : le lemme reste vrai si l'on remplace « pour toute fonction g mesurable positive » par

— pour toute fonction g continue et bornée ;

— pour toute fonction g C^∞ à support compact.

L'idée de la démonstration est d'approcher les fonctions indicatrices $\mathbb{1}_{\mathcal{R}}$ de pavés bornés $\mathcal{R}$ par des fonctions de la classe voulue.

Exemple 3.16. On peut se servir du Lemme 3.14 pour déterminer la loi de v.a. Par exemple, soit $X \sim \mathcal{U}([-\frac{\pi}{2}, \frac{\pi}{2}])$. Posons $Y = \tan(X)$ et déterminons la loi de Y.

Soit $g : \mathbb{R} \to \mathbb{R}_+$ une fonction mesurable positive quelconque (une *fonction test* pour la formule de transfert). Alors, en utilisant la formule de transfert pour la v.a. X, de densité $\frac{1}{\pi}\mathbb{1}_{[-\frac{\pi}{2},\frac{\pi}{2}]}(x)$, on obtient

$$\mathbb{E}(g(Y)) = \mathbb{E}(g(\tan(X))) = \int_{-\pi/2}^{\pi/2} \frac{1}{\pi} g(\tan(x))dx = \int_{\mathbb{R}} g(y)\frac{1}{\pi(1+y^2)}dy ,$$

où on a effectué le changement de variable $y = \tan(x)$ pour la dernière égalité. Cette identité étant valable pour une fonction g quelconque, on en déduit que Y est de loi $\mathbb{P}_Y$ de densité $\frac{1}{\pi(1+y^2)}$, c'est-à-dire Y est de loi de Cauchy standard.

De manière générale, cette technique fonctionne bien pour déterminer la densité (quand elle existe) d'un vecteur aléatoire Y fonction d'un vecteur aléatoire X à densité — les étapes sont toujours les mêmes : prendre une fonction test g pour calculer $\mathbb{E}(g(Y))$; appliquer la formule de transfert à X ; effectuer un changement de variable pour arriver à une intégrale du type $\int_{\mathbb{R}^d} g(y)[\cdots]dy$; identifier la densité de Y.

Exercice 3.17. Soient X, Y deux variables aléatoires indépendantes, à valeurs dans $\mathbb{R}^d$. On suppose que X admet pour densité f, mais on ne suppose rien sur Y. Montrer que $Z = X + Y$ admet une densité et la déterminer en fonction des lois de X et Y.

Solution. Soit $g : \mathbb{R}^d \to \mathbb{R}$ une fonction mesurable positive. On a

$$\mathbb{E}(g(Z)) = \mathbb{E}(g(X+Y)) = \int_{\mathbb{R}^d\times\mathbb{R}^d} g(x+y)d\mathbb{P}_X(x)d\mathbb{P}_Y(y) = \int_{\mathbb{R}^d}\bigg(\int_{\mathbb{R}^d} g(x+y)d\mathbb{P}_Y(y)\bigg)f(x)dx\,,$$

où on a utilisé la formule de transfert pour le vecteur (X, Y), dont la loi est $\mathbb{P}_{X,Y} = \mathbb{P}_X \otimes \mathbb{P}_Y$ par indépendance de X et Y. On a ensuite utilisé Fubini–Tonelli pour intégrer d'abord par rapport à y. Maintenant, l'intégrale interne est égale à $\mathbb{E}(g(x+Y))$: après avoir encore appliqué Fubini–Tonelli on peut écrire

$$\mathbb{E}(g(Z)) = \mathbb{E}\bigg(\int_{\mathbb{R}^d} g(x+Y)f(x)dx\bigg).$$

(pour ceux qui connaissent l'espérance conditionnelle, ça devrait être clair). Si on considère simplement l'intégrale à l'intérieur de l'espérance, un changement de variable $z = x + Y$ donne

$$\mathbb{E}(g(Z)) = \mathbb{E}\bigg(\int_{\mathbb{R}^d} g(z)f(z-Y)dz\bigg) = \int_{\mathbb{R}^d} g(z)\mathbb{E}\big(f(z-Y)\big)dz\,,$$

où on a de nouveau appliqué Fubini–Tonelli pour la dernière égalité. Cette identité étant valable pour toute fonction g mesurable positive, on en déduit que $Z = X + Y$ admet une densité, donnée par la fonction $z \mapsto \mathbb{E}[f(z-Y)]$.

3.3 Moments, variance et covariance

Définition 3.18. Soit X une variable aléatoire réelle, soit $p \geq 0$. Si X^p est intégrable ($X^p \in \mathcal{L}^1$), on dit que X admet un moment d'ordre p fini. La quantité $\mathbb{E}(X^p)$ est alors appelée *moment d'ordre* p de X.

On note $\mathcal{L}^p := \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ l'ensemble des v.a. à valeurs réelles qui admettent un moment d'ordre p fini. Par convention, $\mathcal{L}^0$ est l'ensemble des v.a. réelles.

Proposition 3.19. *On a les propriétés suivantes :*

— *Si $p \geq 0$, et si $X, Y \in \mathcal{L}^p$, alors $X + Y \in \mathcal{L}^p$; autrement dit, $\mathcal{L}^p$ est un espace vectoriel réel.*

— *Si $p' \leq p$ et si $X \in \mathcal{L}^p$, alors $X \in \mathcal{L}^{p'}$.*

— *Si $X, Y \in \mathcal{L}^2$, alors $XY \in \mathcal{L}^1$.*

Démonstration. Pour le premier point, il suffit d'écrire $|X+Y|^p \leq \max\{2|X|, 2|Y|\}^p \leq 2^p|X|^p + 2^p|Y|^p$, avec $|X|^p$ et $|Y|^p$ intégrables.

Pour le deuxième point, on montre facilement que $|X|^{p'} \leq 1 + |X|^p$, donc si $|X|^p$ est intégrable $|X|^{p'}$ l'est aussi.

Pour le troisième point, il suffit d'écrire $|XY| \leq \frac{1}{2}(X^2 + Y^2)$. □

Le moment d'ordre 1 d'une v.a. réelle X est son espérance. On calculera en général des moment d'ordre $p \in \mathbb{N}^*$ de X. Le moment d'ordre 2 est extrêmement utile.

Définition-Proposition 3.20. Si X admet un moment d'ordre 2 fini, alors on définit la *variance* de X :

$$\mathrm{Var}(X) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \geq 0\,.$$

On appelle $\sigma := \sqrt{\mathrm{Var}(X)}$ l'*écart-type* de X.

Contrainrement à l'espérance, la variance *n'est pas* linéaire : pour $a, b \in \mathbb{R}$

$$\mathrm{Var}(aX+b) = a^2 \mathrm{Var}(X)\,.$$

Démonstration. La deuxième égalité ci-dessus s'obtient par linéarité de l'espérance. Posons $m = \mathbb{E}(X) \in \mathbb{R}$ (comme $X \in \mathcal{L}^2$ on a aussi $X \in \mathcal{L}^1$). Alors

$$\mathrm{Var}(X) = \mathbb{E}\big((X-m)^2\big) = \mathbb{E}\big(X^2 - 2mX + m^2\big) = \mathbb{E}(X^2) - 2m\mathbb{E}(X) + m^2 = \mathbb{E}(X^2) - m^2\,.$$

Cette formule est appelée formule de König-Huygens. □

Remarque 3.21. Si $\mathrm{Var}(X) = \mathbb{E}\big((X-\mathbb{E}(X))^2\big) = 0$, alors comme $(X-\mathbb{E}(X))^2$ est positive on a $(X-\mathbb{E}(X))^2 = 0$ p.s. Autrement dit, on a $X = \mathbb{E}(X)$ p.s., c'est-à-dire X est constante p.s. Réciproquement, si X est p.s. constante, alors $\mathrm{Var}(X) = 0$.

Exercice 3.22. Soit X une v.a. réelle admettant un moment d'ordre 2 fini. Montrer que $\mathbb{E}(X)$ est le réel t qui minimise $\mathbb{E}((X-t)^2)$.

Définition-Proposition 3.23. Si $X, Y \in \mathcal{L}^1$ et $XY \in \mathcal{L}^1$ (en particulier si $X, Y \in \mathcal{L}^2$), alors on définit la *covariance* de X, Y par

$$\mathrm{Cov(X,Y)} = \mathbb{E}\Big(\big(\mathrm{X} - \mathbb{E}(\mathrm{X})\big)\big(\mathrm{Y} - \mathbb{E}(\mathrm{Y})\big)\Big) = \mathbb{E}(\mathrm{XY}) - \mathbb{E}(\mathrm{X})\mathbb{E}(\mathrm{Y})\,.$$

La deuxième égalité découle encore une fois de la linéarité de l'espérance. Notons que si $X \in \mathcal{L}^2$, $\mathrm{Var}(X) = \mathrm{Cov(X,X)}$. De plus, si X ou Y est constante p.s. alors $\mathrm{Cov(X,Y)} = 0$.

Proposition 3.24. *La covariance est une forme bilinéaire symétrique. On en déduit la formule utile suivante : pour $X_1, \ldots, X_n \in \mathcal{L}^2$, on a*

$$\mathrm{Var}\Big(\sum_{i=1}^n X_i\Big) = \sum_{i=1}^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov(X_i, X_j)}\,.$$

3.3.1 Espérance et indépendance

Proposition 3.25. *Si $X, Y \in \mathcal{L}^1$ sont deux v.a. réelles indépendantes, alors $XY \in \mathcal{L}^1$ et $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$; en particulier,* $\mathrm{Cov(X,Y)} = 0$.

Note : la même conclusion reste valable si on suppose seulement X, Y positives.

Démonstration. En utilisant la formule de transfert pour (X,Y) et en utilisant que la loi jointe de X, Y est $\mathbb{P}_X \otimes \mathbb{P}_Y$ on a :

$$\mathbb{E}(|XY|) = \int_{\mathbb{R}\times\mathbb{R}} |xy| d\mathbb{P}_X(x) d\mathbb{P}_Y(y) = \left(\int_{\mathbb{R}} |x| d\mathbb{P}_X(x)\right)\left(\int_{\mathbb{R}} |y| d\mathbb{P}_X(y)\right) = \mathbb{E}(|X|)\mathbb{E}(|Y|)\,,$$

où on a utilisé Fubini–Tonelli pour la deuxième identité. Cela montre que $XY \in \mathcal{L}^1$. En répétant la même opération en enlevant les valeurs absolues, on obtient la conclusion. □

Remarque 3.26. Attention : La réciproque est fausse ! On peut avoir $\mathrm{Cov(X,Y)} = 0$ sans avoir X et Y indépendantes. Exemple. Soient $X_1, X_2 \in \mathcal{L}^2$ deux v.a. indépendantes et soient $X = X_1 + X_2$ et $Y = X_1 - X_2$: montrer que $\mathrm{Cov(X,Y)} = 0$, et donner un exemple où X, Y ne sont pas indépendantes.

On peut en réalité caractériser l'indépendance de v.a. à l'aide de l'espérance, dans le même esprit que le Lemme 3.14.

Lemme 3.27. *Des v.a. $X_1, \ldots, X_n$ sont indépendantes si et seulement si pour toutes fonctions $f_1, \ldots, f_n$ mesurables positives on a*

$$\mathbb{E}\big(f_1(X_1) \cdots f_n(X_n)\big) = \prod_{i=1}^n \mathbb{E}\big(f_i(X_i)\big)\,.$$

Démonstration. Dans un sens, utiliser la Proposition ci-dessus. Dans l'autre sens, prendre $f_i = \mathbb{1}_{A_i}$. □

3.3.2 Espérance et variance des lois usuelles

On laisse en exercice le soin de calculer toutes les espérances et variances qui suivent :

	Définition de la loi	Espérance	Variance
$\mathrm{Bern}(p)$, $p \in [0,1]$	$\mathbb{P}(X=1)=p$ $\mathbb{P}(X=0)=1-p$	p	$p(1-p)$
$\mathrm{Bin}(n,p)$, $n \in \mathbb{N}^*$ $p \in [0,1]$	$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$, $k \in \{0,1,\dots,n\}$	np	$np(1-p)$
$\mathrm{Geo}(p)$, $p \in]0,1[$	$\mathbb{P}(X=k)=(1-p)^{k-1}p$, $k \in \mathbb{N}^*$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$\mathrm{Poi}(\lambda)$, $\lambda > 0$	$\mathbb{P}(X=k)=\frac{\lambda^k}{k!}$, $k \in \mathbb{N}$	λ	λ
$\mathrm{Unif}(\{1,\dots,n\})$, $n \in \mathbb{N}$	$\mathbb{P}(X=k)=\frac{1}{n}$, $k \in \{1,\dots,n\}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
$\mathcal{U}([a,b])$, $a<b$ réels	$f_X(x)=\frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$	$\frac{b-a}{2}$	$\frac{(b-a)^2}{12}$
$\mathcal{E}(\lambda)$, $\lambda > 0$	$f_X(x)=\lambda e^{-\lambda x}\mathbb{1}_{\mathbb{R}^+}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\mathcal{N}(\mu,\sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$	$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$	μ	σ^2
$\mathrm{Gam}(a,\lambda)$, $a,\lambda > 0$	$f_X(x)=\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x}\mathbb{1}_{\mathbb{R}^+}(x)$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$
$\mathrm{Beta}(a,b)$, $a,b>0$	$f_X(x)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}\mathbb{1}_{[0,1]}(x)$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$
$\mathrm{Cauchy}(m,\alpha)$	$f_X(x)=\frac{\alpha}{\pi(\alpha^2+(x-m)^2)}$	non définie	non définie

3.4 Inégalités probabilistes

Les inégalités suivantes, bien que parfois très simples, s'avèrent extrêmement utiles.

Théorème 3.28 (Inégalité de Markov). *Soit X une variable aléatoire* positive*, alors pour tout $a>0$ on a*

$$\mathbb{P}(X \geq a) \leq \frac{1}{a}\mathbb{E}(X)\,.$$

Démonstration. On écrit $X = X\mathbb{1}_{\{X<a\}} + X\mathbb{1}_{\{X\geq a\}} \geq a\mathbb{1}_{\{X\geq a\}}$. Par monotonie et linéarité de l'espérance, on a $\mathbb{E}(X) \geq a\mathbb{E}(\mathbb{1}_{\{X\geq a\}}) = a\mathbb{P}(X \geq a)$, ce qui donne la conclusion. □

Théorème 3.29 (Inégalité de Bienaymé–Tchebychev). *Soit X une v.a. réelle admettant un moment d'ordre 2 fini. Alors pour tout $\varepsilon > 0$ on a*

$$\mathbb{P}\big(|X-\mathbb{E}(X)| \geq \varepsilon\big) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}\,.$$

Démonstration. Notons que l'on a $|X-\mathbb{E}(X)| \geq \varepsilon$ si et seulement si $(X-\mathbb{E}(X))^2 \geq \varepsilon^2$. En appliquant l'inégalité de Markov à la v.a. (positive) $(X-\mathbb{E}(X))^2$, on obtient

$$\mathbb{P}\big(|X-\mathbb{E}(X)| \geq \varepsilon\big) = \mathbb{P}\big((X-\mathbb{E}(X))^2 \geq \varepsilon^2\big) \leq \frac{1}{\varepsilon^2}\mathbb{E}\big((X-\mathbb{E}(X))^2\big) = \frac{\mathrm{Var}(X)}{\varepsilon^2}\,.$$

□

Noter que si $\varepsilon = \alpha\sigma$, où $\sigma = \sqrt{\mathrm{Var}(X)}$ est l'écart-type de X, on obtient $\mathbb{P}\big(|X-\mathbb{E}(X)| \geq \alpha\sigma\big) \leq 1/\alpha^2$. Ainsi, la probabilité que X s'écarte de 5σ de son espérance est inférieure à $1/25 = 4\%$: cela justifie le nom d'*écart-type*.

Exemple 3.30. Supposons que pour une naisssance on ait une fille ou un garçon avec la même probabilité $1/2$, indépendamment des autres naissances. On note $n = 700\,000$ le nombre de naissances en France une année donnée et $X =$« nombre de garçons nés ». Selon notre hypothèse, on a $X \sim \mathrm{Bin}(n, 1/2)$, donc notamment $\mathbb{E}(X) = n/2$ et $\mathrm{Var}(X) = n/4$. L'inégalité de Bienaymé–Tchebychev nous donne donc que pour tout $k \geq 1$,

$$\mathbb{P}(X \geq \frac{1}{2}n + k) = \frac{1}{2}\mathbb{P}(|X - \frac{1}{2}n| \geq k) \leq \frac{n}{8k^2}$$

(on a utilisé que $X - n/2$ et $n/2 - X$ ont la même loi pour la première égalité). On observe que $355\,000$ garçon sont nés, c'est-à-dire qu'on observe $X \geq n/2 + k$ avec $k = 5\,000$: sous notre hypothèse d'équiprobabilité des naissances, il s'agit d'un événement de probabilité inférieure à $\frac{n}{8k^2} = 3{,}5 \cdot 10^{-3}$. Cela signifie donc soit qu'on a observé un événement très improbable, soit que notre hypothèse de départ (d'équiprobabilité des naissances) est très peu vraisemblable.

Proposition 3.31 (Inégalité de Jensen). *Soit X une v.a. réelle d'espérance finie et soit $\phi : \mathbb{R} \to \mathbb{R}$ une fonction convexe. Alors $\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}(X))$.*

Démonstration. La fonction ϕ étant convexe, on a pour tout $x \in \mathbb{R}$, $\phi(x) = \sup\{f(x)\,, f \in A_\phi\}$, où $A_\phi = \{f \leq \phi, f(x) = ax + b\}$ est l'ensemble des fonctions affines qui minorent ϕ.

Ainsi, pour tout $f \in A_\phi$, on a $\phi(X) \geq f(X)$, de sorte que

$$\mathbb{E}(\phi(X)) \geq \mathbb{E}(f(X)) = f(\mathbb{E}(X))\,,$$

où la dernière égalité provient de la linéarité de l'espérance, la fonction f étant affine. Cela étant vrai pour tout $f \in A_\phi$, on a

$$\mathbb{E}(\phi(X)) \geq \sup\{f(\mathbb{E}(X))\,, f \in A_\phi\} = \phi(\mathbb{E}(X))\,,$$

ce qui conclut la preuve. □

Exercice 3.32. En utilisant l'inégalité de Jensen, montrer que pour tout $x_1, \ldots, x_n \geq 0$ on a l'inégalité des moyennes géométrique et arithmétique $\big(\prod_{i=1}^n x_i\big)^{1/n} \leq \frac{1}{n}\sum_{i=1}^n x_i$.

Exercice 3.33. Montrer que si ϕ est *strictement* convexe et si X n'est pas p.s. constante, alors $\mathbb{E}(\phi(X)) > \phi(\mathbb{E}(X))$.

Proposition 3.34 (Inégalité de Cauchy–Schwarz). *Soient X, Y deux v.a. dans $\mathcal{L}^2$. Alors $\mathbb{E}(|XY|)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$.*

Cela montre notamment que $X, Y \in \mathcal{L}^2$ implique $XY \in \mathcal{L}^1$.

Proposition 3.35 (Inégalité de Hölder). *Soient $X \in \mathcal{L}^p$ et $Y \in \mathcal{L}^q$, avec $p, q \geq 1$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Alors $\mathbb{E}(|XY|) \leq \mathbb{E}(|X|^p)^{1/p}\mathbb{E}(|Y|^q)^{1/q}$.*

3.4.1 Méthode des moments : exemples

Dans bien des cas, il est possible de calculer espérance et variance (parfois des moments d'ordre supérieur) d'une variable aléatoire, sans avoir besoin de connaître sa loi en détail. On peut alors se servir des inégalités probabilistes pour en déduire des informations sur la variable aléatoire.

Exemple 3.36. Le collectionneur de vignettes (ref : [BCD21, p. 135]). Un collectionneur essaie de compléter une collection de N vignettes : il achète chaque semaine un paquet, la vignette récupérée possédant une chance $\frac{1}{N}$ d'être la vignette k, pour $k \in \{1, \ldots, N\}$. La question est de savoir combien de temps le collectionneur mettra pour compléter sa collection. Notons $A_i =$« la vignette i n'a pas été récupérée après m semaines » et $X_m := \sum_{i=1}^N \mathbb{1}_{A_i}$, égal au nombre de vignettes encore manquantes après m semaines.

On peut calculer $\mathbb{E}(X_m) = \sum_{i=1}^N \mathbb{P}(A_i) = N(1 - \frac{1}{N})^m$ par linéarité de l'espérance. L'événement « la collection n'est pas complète après m semaines » peut s'écrire sous la forme $\{X_m \geq 1\}$, donc grâce à l'inégalité de Markov on obtient $\mathbb{P}(X_m \geq 1) \leq \mathbb{E}(X_m) = N(1-\frac{1}{N})^m$. Il est alors facile de voir que si on prend $m = N \ln N + a_N N$ où a_N est une suite quelconque qui tend vers $+\infty$, on aura $\mathbb{P}(X_m \geq 1) \to 0$. Autrement dit, la probabilité que la collection ne soit pas complétée après $N \ln N + a_N N$ semaines tend vers 0.

On a aussi que l'événement « la collection est complète avant m semaines » s'écrit $\{X_m = 0\} = \{X_m \leq 0\}$. On peut alors utiliser l'inégalité de Bienaymé–Tchebychev pour montrer que $\mathbb{P}(X_m = 0) \leq \mathrm{Var}(X_m)/\mathbb{E}(X_m)^2$. Un calcul donne $\mathrm{Var}(X_m) \leq \mathbb{E}(X_m)$ et on peut conclure comme au-dessus que la probabilité que la collection soit complétée avant $N \ln N - a_N N$ semaines tend vers 0 (si $a_N \to \infty$).

Exemple 3.37. Longue suite de résultats égaux consécutifs (ref : [BCD21, p. 186]). Soit $(X_i)_{i\geq 1}$ une suite de v.a. indépendantes de loi $\mathrm{Ber}(1/2)$. On s'intéresse à la plus longue suite de résultats égaux consécutifs parmi les n permières variables. On pose

$$L_n := \max\{k\,, \exists 1 \leq i \leq n-k \text{ tel que } X_i = X_{i+1} = \cdots = X_{i+k-1}\}\,.$$

Si l'on demande à une personne combien de Pile consécutifs on pourra trouver dans 100 Pile ou Face, elle aura souvent tendance à sous-estimer ce nombre. (On a un grande probabilité d'avoir $L_{100} \geq 6$).

On peut utiliser la méthode des moments pour étudier L_n. On introduit une variable aléatoire annexe : on pose $X_k := \sum_{i=1}^{n-k+1} \mathbb{1}_{A_i}$ où $A_i = \{X_i = X_{i+1} = \cdots = X_{i+k-1}\}$ est l'événement qu'une suite de longueur k de résultats égaux consécutifs démare au i-ème lancer. Ainsi, X_k représente le nombre de suites de k résultats égauxconsécutifs (avec répétition si on a une suite de longueur $> k$), et on a ainsi $\{L_n \geq k\} = \{X_k \geq 1\}$.

Comme dans l'exemple ci-dessus, on a grâce à l'inégalité de Markov $\mathbb{P}(L_n \geq k) = \mathbb{P}(X_k \geq 1) \leq \mathbb{E}(X_k) = \sum_{i=1}^{n-k+1} \mathbb{P}(A_i) = (n-k+1)2^{-k+1}$. On en déduit que $\mathbb{P}(L_n \geq \log_2 n + C) \leq 2 \cdot 2^{-C}$. D'autre part, avec l'inégalité de Bienaymé–Tchebychev on obtient $\mathbb{P}(L_n < k) = \mathbb{P}(X_k = 0) = \mathbb{P}(X_k \leq 0) \leq \mathrm{Var}(X_k)/\mathbb{E}(X_k)^2$: un calcul donne $\mathrm{Var}(X_k) \leq 3\mathbb{E}(X_k)$, ce dont on déduit que $\mathbb{P}(L_n < \log_2 n - C) \leq 6 \cdot 2^{-C}$. Autrement dit, on a montré que $|L_n - \log_2 n| \leq C$ avec probabilité supérieure à $1 - 8 \cdot 2^{-C}$, c'est-à-dire avec probabilité proche de 1 si C est grand.

3.5 Espaces L^p

Rappel : $\mathcal{L}^p = \{X\,, \mathbb{E}(|X|^p) < +\infty\}$ est un espace vectoriel.

Définition 3.38. Soit X une v.a. réelle. Pour $p > 0$, on définit $\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$. On peut aussi définir $\|X\|_\infty := \inf\{M, \mathbb{P}(|X| \leq M) = 1\}$ (appelé supremum essentiel).

Remarque 3.39. Soient $0 < q < p$. Alors on a $\|X\|_q \leq \|X\|_p$. En effet, en appliquant l'inégalité de Jensen avec la fonction $\phi : x \mapsto x^{p/q}$, qui est concave sur $\mathbb{R}_+$, on a

$$\|X\|_q^q = \mathbb{E}\big(|X|^q\big) \leq \mathbb{E}\big(|X|^p\big)^{q/p} = \|X\|_p^q\,,$$

ce qui donne la conclusion voulue.

Notons que $\|X\|_p = 0$ implique $X = 0$ seulement au sens presque sûr. Pour remédier à cela, on définit :

$$L^p = \text{ quotient de } \mathcal{L}^p \text{ par l'espace vectoriel } V = \{X \in \mathcal{L}^p\,, \|X\|_p = 0\}\,.$$

Autrement dit, deux v.a. X, X' égales p.s. sont identifiées. On peut écrire $L^p = \mathcal{L}^p/\sim$ où $X \sim X'$ si et seulement si $X = X'$ p.s. (ou $X - X' \in V$).

Le résultat suivant (démontré dans le chapitre d'intégration) montre que, pour $p \geq 1$, $\|\cdot\|_p$ satisfait l'inégalité triangulaire : cela montre que $\|\cdot\|_p$ est une norme sur L^p. Attention, ceci n'est pas vrai pour $p \in]0,1[$, ce qui justifie le fait de considérer les espaces L^p la plupart du temps pour $p \geq 1$.

Proposition 3.40 (Inégalité de Minkowski). *Pour tout $p \geq 1$, on a $\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$.*

On a le résultat important suivant, dont la démonstration est faite dans le chapitre d'intégration.

Théorème 3.41. *Pour tout $p \geq 1$ (et $p = \infty$), l'espace vectoriel normé $(L^p, \|\cdot\|_p)$ est complet.*

Note : on montrera plus bas que l'espace L^0, muni de la distance $d_0(X,Y) = \mathbb{E}(|X-Y| \wedge 1)$, est aussi un espace complet (voir le Théorème 6.12).

Exemple 3.42. Soit $(X_i)_{i\geq 1}$ une suite de v.a. indépendantes, de loi $\mathbb{P}(X_i = +1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$. On pose $S_0 = 0$ et pour $n \geq 1$ $S_n = \sum_{i=1}^n \frac{X_i}{i}$: il s'agit de la série harmonique de signe aléatoire. On va montrer que la suite $(S_n)_{n\in\mathbb{N}}$ est de Cauchy dans L^2 : par complétude, on aura montré qu'il existe $S_\infty \in L^2$ telle que $\lim_{n\to\infty} \|S_n - S_\infty\|_2 = 0$.

Soient $m > n \geq 1$. Les X_i étant centrés, on a clairement

$$\|S_m - S_n\|_2^2 = \mathbb{E}\bigg(\Big(\sum_{i=n+1}^m \frac{X_i}{i}\Big)^2\bigg) = \sum_{i=n+1}^m \mathrm{Var}\Big(\frac{X_i}{i}\Big) = \sum_{i=n+1}^m \frac{1}{i^2}\,.$$

où on a utilisé l'indépendance des (X_i) puis le fait que $\mathrm{Var}(X_i) = 1$. Ainsi, $\sup_{m\geq n} \|S_m - S_n\|_2^2 \leq \sum_{i=n+1}^\infty \frac{1}{i^2}$ converge vers 0 quand $n \to \infty$, la série $\sum_{i\geq 1} \frac{1}{i^2}$ étant convergente. Cela montre que la suite $(S_n)_{n\in\mathbb{N}}$ est de Cauchy dans L^2 et conclut la preuve. Cela montre aussi que $\mathrm{Var}(S_\infty) = \|S_\infty\|_2^2 = \lim_{n\to\infty} \|S_n\|_2^2 = \sum_{i=1}^\infty \frac{1}{i^2} = \frac{\pi^2}{6}$.

Remarque 3.43. La même démonstration montre que si les Y_i sont indépendants avec $\mathbb{E}(Y_i) = 0$ et $\sum_{i=1}^\infty \mathrm{Var}(Y_i) < \infty$, alors $S_n = \sum_{i=1}^n Y_i$ converge pour la norme L^2. Dans ce cadre, on a en fait aussi que la limite $\lim_{n\to\infty} S_n$ existe p.s., mais c'est nettement plus dur (cela utilise l'inégalité maximale de Tchebychev, voir les références [App15, p. 649], [BCD21, p. 394].)

4 Caractérisation des lois

Dans cette section, on donne des conditions nécessaires et suffisantes pour identifier les lois de variables aléatoires. Ces conditions permettent par exemple de déterminer des lois, ou d'avoir d'autres critères pour l'indépendance de variables aléatoires.

4.1 Déjà vu : fonction de répartition & formule de transfert

4.1.1 Fonction de répartition

Le premier outil qu'on a déjà vu est la fonction de répartition, voir le Théorème 2.10. On rappelle que si $X = (X_1, \ldots, X_d)$ est une v.a. à valeurs dans $\mathbb{R}^d$ alors sa fonction de répartition est $F_X(t) = F_X(t_1, \ldots, t_d) = \mathbb{P}(X_1 \leq t_1, \ldots, X_d \leq t_d)$, définie sur $\mathbb{R}^d$. Le Théorème 2.10 annonce que deux v.a. qui ont la même fonction de répartition ont la même loi.

Une conséquence est que des v.a. $X_1, \ldots, X_n$ à valeurs dans $\mathbb{R}^{d_1}, \ldots, \mathbb{R}^{d_n}$ sont indépendantes si et seulement si $F_{X_1,\ldots X_n}(t_1, \ldots, t_n) = F_{X_1}(t_1) \cdots F_{X_n}(t_n)$ pour tout $t_1 \in \mathbb{R}^{d_1}, \ldots, t_n \in \mathbb{R}^{d_n}$.

Exemple 4.1. La fonction de répartition permet donc de calculer des lois. Voici un exemple ultra-classique. Soit $U \sim \mathcal{U}([0,1])$ et $X = -\ln U$. Déterminons la loi de X avec sa fonction de répartition.

Comme $X > 0$ p.s., on a déjà $F_X(t) = 0$ pour $t \leq 0$. Pour $t > 0$, on a

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(-\ln U \leq t) = \mathbb{P}(U \geq e^{-t}) = \mathbb{P}(U \in [e^{-t}, 1]) = 1 - e^{-t}\,.$$

On en conclut que $F_X(t) = (1 - e^{-t})\mathbb{1}_{\mathbb{R}_+}(t)$, qui est exactement la fonction de répartition de la loi exponentielle de paramètre 1 : on a donc $X \sim \mathcal{E}(1)$. (On aurait aussi pu remarquer que F_X est C^0 et C^1 par morceaux, donc on peut trouver la densité de X en dérivant F_X, voir Lemma 2.39.)

4.1.2 Formule de transfert

Le deuxième outil qu'on a déjà vu est le Lemme 3.14, qui consiste à vérifier la formule de transfert pour toute une classe de fonctions test. Rappelons le résultat ici.

Une v.a. X à valeurs dans E est de loi μ si et seulement si pour toute fonction $g \in \mathcal{T}$ on a $\mathbb{E}(g(X)) = \int_E g(x)d\mu(x)$. L'ensemble $\mathcal{T}$ est un ensemble de fonctions test : on peut prendre $\mathcal{T}$ l'ensemble des fonctions mesurables positives ou l'ensemble des fonctions mesurables bornée ; dans le cas $E = \mathbb{R}^d$, on peut prendre $\mathcal{T}$ l'ensemble des fonctions C^0 bornées ou l'ensemble des fonctions C^∞ à support compact.

Une autre conséquence de la formule de transfert est que des v.a. $X_1, \ldots, X_n$ à valeurs dans $E_1, \ldots, E_n$ sont indépendantes si et seulement si $\mathbb{E}(g_1(X_1) \cdots g_n(X_n)) = \mathbb{E}(g_1(X_1)) \cdots \mathbb{E}(g_n(X_n))$ pour toutes fonctions $g_1, \ldots, g_n \in \mathcal{T}$.

Exemple 4.2. La méthode de la fonction test pour la formule de transfert permet aussi de calculer des lois. Donnons une illustration en résolvant l'Exercice 2.49 : soient X, Y deux variables aléatoires indépendantes, de lois $X \sim \mathrm{Gam}(a, \lambda)$, $Y \sim \mathrm{Gam}(b, \lambda)$. Posons $W = X+Y$ et $Z = \frac{X}{X+Y}$ et déterminons la loi de (W, Z).

Soit $g : \mathbb{R}^2 \to \mathbb{R}_+$ une fonction mesurable positive : en appliquant la formule de transfert au vecteur (X, Y) de densité jointe $\frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} x^{a-1} y^{b-1} e^{-\lambda(x+y)}$ (par indépendance), on obtient

$$\mathbb{E}(g(W, Z)) = \mathbb{E}\Big(g\Big(X+Y, \frac{X}{X+Y}\Big)\Big) = \int_{]0,\infty[^2} g\Big(x+y, \frac{x}{x+y}\Big) \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} x^{a-1} y^{b-1} e^{-\lambda(x+y)} dxdy\,.$$

On effectue le changement de variable avec le difféomorphisme $\varphi :]0,\infty[^2 \to]0,\infty[\times]0,1[$ défini par $(w, z) = \varphi(x, y) := (x+y, \frac{x}{x+y})$, d'inverse $\varphi^{-1}(w, z) = (wz, w(1-z))$: on a $\mathrm{Jac}_{\varphi^{-1}} = \begin{pmatrix} z & w \\ 1-z & -w \end{pmatrix}$ donc $|\mathrm{Jac}_{\varphi^{-1}}| = |w| = w$, et on obtient

$$\mathbb{E}(g(W, Z)) = \int_{]0,\infty[\times]0,1[} g(w, z) \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} w^{a+b-1} z^{a-1} (1-z)^{b-1} e^{-\lambda w} dwdz\,.$$

Comme cette formule est valable pour toute fonction test, on en déduit que (W, Z) admet pour densité jointe

$$f_{W,Z}(w, z) = \frac{\lambda^{a+b}}{\Gamma(a+b)} w^{a+b-1} e^{-\lambda w} \mathbb{1}_{]0,\infty[}(w) \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} z^{a-1} (1-z)^{b-1} \mathbb{1}_{]0,1[}(z)\,,$$

dont on déduit que $W \sim \mathrm{Gam}(a+b, \lambda)$, $Z \sim \mathrm{Beta}(a, b)$ et sont indépendantes.

4.2 Fonctions génératrice et génératrice des moments

4.2.1 Fonction génératrice

La fonction génératrice est particulièrement utile pour des v.a. à valeurs dans $\mathbb{N}$: il s'agit de la série entière dont le terme général est la densité discrète de X.

Définition 4.3. Soit X une v.a. à valeurs dans $\mathbb{N}$. La *foncion génératrice* de X est définie par

$$G_X(z) := \mathbb{E}(z^X) = \sum_{k=0}^{+\infty} \mathbb{P}(X = k) z^k$$

pour tout $|z| < R_X$ où $R_X \geq 1$ est le rayon de convergence de la série entière de t.g. $(\mathbb{P}(X = k))_{k\in\mathbb{N}}$. On peut prendre $z \in \mathbb{C}$, mais on se concentrera sur le cas $z \in \mathbb{R}$. Noter que $G_X(z)$ est toujours défini (mais peut valoir $+\infty$) si $z \geq 0$.

La formule $G_X(z) = \mathbb{E}(z^X)$ permet de définir $G_X(z)$ pour une v.a. X plus générale mais on perd l'interprétation en terme de série entière. Assez clairement, par unicité de la série entière, la fonction génératrice caractérise la densité discrète d'une v.a. à valeurs dans $\mathbb{N}$, donc elle caractérise sa loi.

Proposition 4.4. *Soient X, Y deux variables aléatoires à valeurs dans $\mathbb{N}$. Si $G_X = G_Y$, alors X et Y ont la même loi.*

Note : si X, Y sont indépendantes, alors $G_{X+Y}(z) = G_X(z) G_Y(z)$ pour tout $|z| < \min(R_X, R_Y)$.

Exercice 4.5. Soient X, Y deux v.a. indépendantes, de loi $X \sim \mathrm{Poi}(\lambda)$, $Y \sim \mathrm{Poi}(\mu)$. Calculer les fonctions génératrices de X, Y et $X + Y$. En déduire que $X + Y \sim \mathrm{Poi}(\lambda + \mu)$.

Un autre avantage de la fonction génératrice est qu'elle permet de récupérer les moments de la v.a. X.

Proposition 4.6. *Soit X une v.a. à valeurs dans $\mathbb{N}$. Alors pour tout $n \geq 1$,*

$$\lim_{z\uparrow 1} G_X^{(n)}(z) = \mathbb{E}\Big(X(X-1)\cdots(X-n+1)\Big) \in [0,\infty]\,,$$

où $G_X^{(n)}$ est la dérivée n-ème de X. En particulier, si $R_X > 1$ on a $\lim_{z\uparrow 1} G_X^{(n)}(z) = G_X^{(n)}(1)$.

Démonstration. D'après le théorème de dérivation des séries entières, tout $|z| < 1$ on a

$$G_X^{(n)}(z) = \sum_{k=n}^{\infty} k(k-1)\cdots(k-n+1)z^{k-n}\mathbb{P}(X=k) = \mathbb{E}\Big(X(X-1)\cdots(X-n+1)z^{X-n}\Big)\,.$$

La conclusion de la proposition découle alors du théorème de convergence monotone. □

4.2.2 Fonction génératrice des moments

Définition 4.7. Soit X une v.a. réelle. La *fonction génératrice des moments* de X est définie par

$$M_X(t) := \mathbb{E}\big(e^{tX}\big) \in [0,\infty] \qquad t \in \mathbb{R}.$$

Il s'avère que, sous certaines conditions, la fonction génératrice M_X est la série entière dont le terme général est donné par la suite des moments $(\mathbb{E}(X^n))_{n\in\mathbb{N}}$.

Proposition 4.8. *Soit $a = \sup\{|t|, M_X(t) < +\infty\}$, de sorte que $M_X(t)$ est fini sur $]-a,a[$. Si $a > 0$, alors on a pour tout $t \in]-a,a[$ on a $M_X^{(n)}(t) = \mathbb{E}(X^n e^{tX}) < +\infty$; en particulier, $M_X^{(n)}(0) = \mathbb{E}(X^n)$ et $M_X(t) = \sum_{n=0}^{\infty} \frac{1}{n!}\mathbb{E}(X^n)t^n$ pour tout $t \in]-a,a[$.*

Remarque 4.9. Réciproquement, si $\sum_{n=0}^{\infty} \frac{1}{n!}\mathbb{E}(|X|^n)t^n < +\infty$ pour tout $t \in]-a,a[$, alors $M_X(t)$ est fini sur $]-a,a[$ et $M_X(t) = \sum_{n=0}^{\infty} \frac{1}{n!}\mathbb{E}(X^n)t^n < +\infty$.

Exemple 4.10. Si $X \sim \mathcal{N}(0,1)$, alors un calcul donne $M_X(t) = \mathbb{E}(e^{tX}) = e^{t^2/2}$ pour tout $t \in \mathbb{R}$. On peut appliquer le résultat précédent et identifier le développement en série entière de $M_X(t) = e^{t^2/2} = \sum_{k=0}^{\infty} \frac{1}{2^k k!}t^{2k}$ aux termes de la série entière $\sum_{n=0}^{\infty} \frac{1}{n!}\mathbb{E}(X^n)t^n$. On en déduit que si n est impair on a $\mathbb{E}(X^n) = 0$ et si $n = 2k$ on a $\mathbb{E}(X^{2k}) = \frac{(2k)!}{2^k k!}$.

De manière générale, la fonction génératrice des moments ne caractérise pas la loi des variables aléatoires. Mais on a le résultat suivant.

Proposition 4.11. *Soient X, Y deux v.a. réelles. Si leurs fonctions génératrices sont finies sur $]-a,a[$ pour un $a > 0$ et si $M_X = M_Y$, alors X et Y ont la même loi.*

De manière générale, les moments (entiers) d'une variable aléatoire n'en caractérise pas la loi — cependant, c'est le cas sous certaines conditions.

Exercice 4.12. Soient X et Y deux v.a. telles que $\mathbb{E}(X^k) = \mathbb{E}(Y^k)$ pour tout $k \in \mathbb{N}$ (on suppose que tous les moments existent).

1. On suppose que $|X|, |Y| \leq M$ p.s., pour un $M > 0$ donné. Montrer que X et Y ont la même loi.
2. On suppose que $\mathbb{E}(|X|^k) \leq C^k k!$ pour un $C > 0$. Montrer que X et Y ont la même loi.

Exercice 4.13. Soit X de densité $f(x) = (2\pi)^{-1}x^{-1}\exp(-(\ln x)^2/2)\mathbb{1}_{\mathbb{R}_+^*}(x)$ et Y de densité $g(x) = f(x)(1+\sin(2\pi\ln x))$.

1. Montrer que $\mathbb{E}(X^n) = e^{n^2/2}$ pour tout $n \geq 0$. (Indication : montrer que $\ln X \sim \mathcal{N}(0,1)$).
2. Montrer que $\mathbb{E}(Y^n) = \mathbb{E}(X^n)$ pour tout $n \geq 0$.

4.3 Fonction caractéristique

Définition 4.14. Soit X une v.a. à valeurs dans $\mathbb{R}^d$. La *fonction caractéristique* de X est définie par

$$\phi_X(t) := \mathbb{E}(e^{i\,t\cdot X}) = \int_{\mathbb{R}^d} e^{i\,\langle t,x\rangle} d\mathbb{P}_X(x)\,, \qquad t \in \mathbb{R}^d\,,$$

où $\langle t,x\rangle$ est le produit scalaire standard de $\mathbb{R}^d$.

Notons que $|e^{itX}| = 1$, donc $\mathbb{E}(e^{itX})$ est bien définie. La fonction caractéristique correspond à *la transformée de Fourier de la loi* $\mathbb{P}_X$; la transformée de Fourier de fonctions sera vue dans un prochain chapitre.

Remarque 4.15. Voici quelques propriétés de la fonction caractéristique, dans le cas d'une v.a. réelle :
- ϕ_X est uniformément continue (on peut facilement écrire $|\phi(t+h) - \phi(t)| \le \mathbb{E}(|e^{ihX} - 1|)$ qui tend vers 0 par convergence dominée) ;
- si X admet un moment d'ordre k fini alors ϕ est de classe C^k et $\phi_X^{(k)}(t) = \mathbb{E}((iX)^k e^{itX})$; en particulier $\mathbb{E}(X^k) = (-i)^k\,\phi_X^{(k)}(0)$;

Remarque 4.16. Dans le cas où X admet une densité f, la fonction caractéristique correspond exactement à la transformée de Fourier de f (modulo les conventions de normalisation) :

$$\phi_X(t) = \int_{\mathbb{R}^d} e^{i\,\langle t,x\rangle} f(x)dx = \hat{f}(t)\,.$$

En particulier, on a $\lim_{|t|\to\infty} \phi_X(t) = 0$ d'après le lemme de Riemann–Lebesgue ;

Exemple 4.17 (Fonction caractéristique de la gaussienne)**.** Soit $X \sim \mathcal{N}(0,1)$. Calculons sa fonction caractéristique

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx\,.$$

D'après le théorème de dérivation sous le signe intégral (bien vérifier l'hypothèse de domination), puis par une intégration par parties, on obtient

$$\phi_X'(t) = i \int_{\mathbb{R}} e^{itx} \frac{x}{\sqrt{2\pi}} e^{-x^2/2} dx = -t \int_{\mathbb{R}} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = -t\phi_X(t)\,.$$

Ainsi, ϕ_X vérifie l'équation différentielle ordinaire $\phi_X'(t) = -t\phi_X(t)$ avec la condition $\phi_X(0) = 1$. On en déduit que $\phi_X(t) = e^{-t^2/2}$.

Si $X \sim \mathcal{N}(\mu,\sigma^2)$, comme $\frac{1}{\sigma}(X-\mu) \sim \mathcal{N}(0,1)$, on en déduit que $\phi_X(t) = e^{i\mu t - \sigma^2 t^2/2}$.

Comme son nom l'indique, la fonction caractéristique caractérise la loi : voici le théorème principal de cette section, que l'on démontrera à la fin de la section.

Théorème 4.18. *Soient X, Y deux v.a. à valeurs dans $\mathbb{R}^d$. Si on a $\phi_X = \phi_Y$, alors X et Y sont de même loi.*

Ainsi, l'application $X \mapsto \phi_X$ est injective. Un corollaire naturel est la caractérisation de l'indépendance à travers les fonctions caractéristiques.

Corollaire 4.19. *Les v.a. $X_1,\dots,X_n$ à valeurs dans $\mathbb{R}^{d_1},\dots,\mathbb{R}^{d_n}$ sont indépendantes si et seulement si $\phi_{X_1,\dots,X_n}(t_1,\dots,t_n) = \phi_{X_1}(t_1)\cdots\phi_{X_n}(t_n)$ pour tous $t_1 \in \mathbb{R}^{d_1},\dots,t_n \in \mathbb{R}^{d_n}$.*

Exercice 4.20. Montrer que si $X_1 \sim \mathcal{N}(\mu_1,\sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_1,\sigma_2^2)$ sont indépendantes, alors $X_1 + X_2 \sim \mathcal{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

Dans la suite, on va se concentrer esssentiellement sur le cas de v.a. réelles : le cas de v.a. à valeurs dans $\mathbb{R}^d$ est essentiellement inchangé, avec des notations plus lourdes.

4.3.1 Démonstration du théorème 4.18

References : [GK19, p.239] ([Ouv09, p. 193] pour une autre méthode). On va montrer la formule d'inversion suivante :

$$\mathbb{P}\big(X \in]a,b[\big) + \tfrac{1}{2}\mathbb{P}(X=a) + \tfrac{1}{2}\mathbb{P}(X=b) = \lim_{T\to\infty} I_T(a,b) \\ \text{avec} \quad I_T(a,b) := \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita}-e^{-itb}}{it}\,\phi_X(t)dt\,. \tag{4.1}$$

On remplace ϕ_X par sa définition et on applique le théorème de Fubini :

$$I_T(a,b) := \frac{1}{2\pi}\int_{\mathbb{R}}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\,e^{itx}dtd\mathbb{P}_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}}\left(\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dt\right)d\mathbb{P}_X(x)\,.$$

Or $\int_{-T}^{T}\frac{e^{i\alpha t}}{it}dt = -\int_{-T}^{T}\frac{e^{-i\alpha t}}{it}dt$ donc $\int_{-T}^{T}\frac{e^{i\alpha t}}{it}dt = \int_{-T}^{T}\frac{\sin(\alpha t)}{t}dt$. Ainsi

$$\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dt = \frac{1}{2\pi}\int_{-T}^{T}\frac{\sin((x-a)t)}{t}dt - \frac{1}{2\pi}\int_{-T}^{T}\frac{\sin((x-b)t)}{t}dt\,.$$

Il reste à observer que comme $\int_{-y}^{y}\frac{\sin(t)}{t}dt$ converge vers π lorsque $y\to\infty$, on a que $\int_{-T}^{T}\frac{\sin(\alpha t)}{t}dt$ converge vers $-\pi$ si $\alpha<0$ et vers π si $\alpha>0$ (et est égale à 0 si $\alpha=0$). On en déduit donc que

$$\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{it(x-a)}-e^{it(x-b)}}{it}dt = \begin{cases} -\frac{1}{2}+\frac{1}{2}=0 & \text{si } x<a\,,\\ 0+\frac{1}{2}=\frac{1}{2} & \text{si } x=a\,,\\ \frac{1}{2}+\frac{1}{2}=1 & \text{si } a<x<b\,,\\ \frac{1}{2}-0=\frac{1}{2} & \text{si } x=b\,,\\ \frac{1}{2}-\frac{1}{2}=0 & \text{si } x>b\,.\end{cases}$$

On en conclut (par convergence dominée) que

$$\lim_{T\to\infty} I_T(a,b) = \int_{\mathbb{R}}(\tfrac{1}{2}\mathbb{1}_{\{a\}}(x) + \mathbb{1}_{]a,b[}(x) + \tfrac{1}{2}\mathbb{1}_{\{b\}}(x))dP_X(x)\,,$$

qui donne le résultat voulu.

Maintenant, cela permet de montrer que si $\phi_X=\phi_Y$, alors

$$\mathbb{P}\big(X\in]a,b[\big)+\tfrac{1}{2}\mathbb{P}(X=a)+\tfrac{1}{2}\mathbb{P}(X=b) = \mathbb{P}\big(Y\in]a,b[\big)+\tfrac{1}{2}\mathbb{P}(Y=a)+\tfrac{1}{2}\mathbb{P}(Y=b)$$

pour tous $a<b$. En particulier, pour tous a,b dans $\mathbb{R}\setminus(\mathcal{A}_X\cup\mathcal{A}_Y)$, où $\mathcal{A}_X=\{x,\mathbb{P}(X=x)>0\}$ et $\mathcal{A}_Y=\{x,\mathbb{P}(Y=x)>0\}$ sont les ensembles des atomes de X,Y, on a $\mathbb{P}\big(X\in]a,b[\big)=\mathbb{P}\big(Y\in]a,b[\big)$. Comme $\{\,]a,b[\,,a,b\in\mathbb{R}\setminus(\mathcal{A}_X\cup\mathcal{A}_Y)\}$ forment un π-système qui engendre la tribu borélienne (noter que $\mathcal{A}_X$, $\mathcal{A}_Y$ sont au plus dénombrables), on en déduit que $\mathbb{P}_X=\mathbb{P}_Y$.

Si on préfère, on peut passer par la fonction de répartition : on a

$$F_X(b)-F_X(a) = \mathbb{P}(X\in]a,b[\,)+\mathbb{P}(X=b) = \lim_{T\to\infty}I_T(a,b)+\frac{1}{2}(\mathbb{P}(X=a)-\mathbb{P}(X=b))\,,$$

de sorte que

$$F_X(b) = \lim_{a\to-\infty}\lim_{T\to\infty}I_T(a,b)-\frac{1}{2}\mathbb{P}(X=b)\,.$$

Ainsi, la fonction de répartition est déterminée par sa fonction caractéristique pour tout $b\notin\mathcal{A}_X$; comme $\mathcal{A}_X$ est au plus dénombrable et que F_X est continue à droite, on en conclut que F_X est caractérisée par ϕ_X. Comme F_X caractérise la loi de X, on en déduit que ϕ_X caractérise la loi de X.

4.3.2 Cas discret : formule d'inversion

Ref : [Ouv09, p. 228]. On va montrer la formule suivante, qui permet de déterminer la densité discrète d'une v.a. à partir de sa fonction caractéristique.

Proposition 4.21. *Pour tout $a \in \mathbb{R}$ fixé,*

$$\mathbb{P}(X = a) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-iat}\phi_X(t)dt\,.$$

Démonstration. On écrit la définition de la fonction caractéristique et on applique Fubini :

$$\frac{1}{2T}\int_{-T}^{T} e^{-iat}\phi_X(t)dt = \int_{\mathbb{R}}\left(\frac{1}{2T}\int_{-T}^{T} e^{i(x-a)t}dt\right)d\mathbb{P}_X(x)\,.$$

Maintenant, on observe que

$$\frac{1}{2T}\int_{-T}^{T} e^{i(x-a)t}dt = \begin{cases} 1 & \text{si } x = a\,, \\ \frac{\sin((x-a)T)}{(x-a)T} & \text{si } x \neq a\,, \end{cases}$$

de sorte que, par convergence dominée on a

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-iat}\phi_X(t)dt = \int_{\mathbb{R}} \mathbb{1}_{\{a\}}(x)d\mathbb{P}_X(x) = \mathbb{P}(X = a)\,,$$

ce qui conclut la démonstration. □

4.3.3 Cas à densité : formule d'inversion

Ref : [Ouv09, p. 228]. Notons que si X admet une densité f, alors on a

$$\phi_X(t) = \int_{\mathbb{R}} e^{i\,tx}f(x)dx\,,$$

qui est donc la transformée de Fourier de la fonction f (modulo les conventions différentes au niveau des notations/normalisations).

On a le résultat suivant, qui permet de déterminer la densité d'une v.a. à partir de sa fonction caractéristique (qui correspond à la formule d'inversion de Fourier vue plus loin, mais sans savoir a priori que X est à densité).

Proposition 4.22. *Soit X une variable aléatoire réelle. Si ϕ_X est intégrable sur $\mathbb{R}$, alors X admet une densité f (qui est continue et bornée), donnée par*

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\phi_X(t)dt\,.$$

Démonstration. La première chose à remarquer est que si ϕ_X est intégrable, alors $\mathbb{P}_X$ n'a aucun atome. Cela découle en effet du lemme précédent, qui montre que pour tout $a \in \mathbb{R}$,

$$|\mathbb{P}(X = a)| = \lim_{T\to\infty}\left|\frac{1}{2T}\int_{\mathbb{R}} e^{-iat}\phi_X(t)dt\right| \leq \lim_{T\to\infty}\frac{1}{2T}\int_{\mathbb{R}}|\phi_X(t)|dt = 0\,.$$

On peut alors reprendre la formule (4.1) démontrée plus haut.

$$\mathbb{P}\big(X \in]a,b[\,\big) = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\,\phi_X(t)dt\,.$$

On écrit $\frac{e^{-ita}-e^{-itb}}{it} = \int_a^b e^{-ixt}dx$ et on applique Fubini (c'est possible car les bornes a, b sont bornées et ϕ_X est intégrable) : on obtient

$$\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\,\phi_X(t)dt = \int_a^b f_T(x)dx \quad \text{avec} \quad f_T(x) := \frac{1}{2\pi}\int_{-T}^{T} e^{-ixt}\phi_X(t)dt\,.$$

Il reste à prendre la limite $T \to \infty$, et l'intervertir avec l'intégrale. Notons que l'on a $|f_T(x)| \leq \frac{1}{2\pi}\int_{\mathbb{R}} |\phi_X(t)|dt$, qui est une constante, donc intégrable sur $[a,b]$. Par convergence dominée, on a donc

$$\mathbb{P}\big(X \in]a,b[\big) = \int_a^b \lim_{T\to\infty} f_T(x)dx = \int_a^b \left(\frac{1}{2\pi}\int_{\mathbb{R}} e^{-ixt}\phi_X(t)dt\right)dx\,,$$

ce qui conclut la démonstration. □

Exemple 4.23. Donnons un exemple d'application de la formule d'inversion. Soit X une v.a. de densité $f(x) = \frac{1}{2}e^{-|x|}$ sur $\mathbb{R}$. On peut calculer sa fonction caractéristique :

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx}f(x)dx = \frac{1}{2}\int_0^\infty e^{itx}e^{-x}dx + \frac{1}{2}\int_0^\infty e^{-itu}e^{-u}du$$

où on a effectué un changement de variable $u = -x$ pour l'intégrale sur $\mathbb{R}_-$. On peut calculer ces intégrales : on obtient

$$\phi_X(t) = \frac{1}{2}\frac{1}{1-it} + \frac{1}{2}\frac{1}{1+it} = \frac{1}{1+t^2}\,.$$

On remarque que ϕ_X ressemble drôlement à la densité d'une v.a. de Cauchy standard... Soit Y une v.a. de Cauchy standard, de densité $f_Y(y) = \frac{1}{\pi(1+y^2)} = \frac{1}{\pi}\phi_X(y)$. La fonction caractéristique de Y vaut

$$\phi_Y(t) = \int_{\mathbb{R}} e^{ity}f_Y(y)dy = \int_{\mathbb{R}} e^{-ity'}f_Y(y')dy' = \int_{\mathbb{R}} e^{-itu}\tfrac{1}{\pi}\phi_X(u)du$$

où on a utilisé le changement de variable $y' = -y$ et la parité de f_Y. D'après la formule d'inversion, comme ϕ_X est intégrable, on a $\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itu}\ \phi_X(u)du = f_X(t)$: on en conclut que

$$\phi_Y(t) = 2f_X(t) = e^{-|t|}\,, \qquad t \in \mathbb{R}\,.$$

On a ainsi déterminé la fonction caractéristique d'une loi de Cauchy standard. (Une autre méthode emploie la formule des résidus, voir [GK19, p. 249].)

Exercice 4.24. Soient X, Y des v.a. indépendantes de lois $X \sim \mathrm{Cauchy}(m_1, \alpha_1)$, $Y \sim \mathrm{Cauchy}(m_2, \alpha_2)$. Montrer que $X + Y \sim \mathrm{Cauchy}(m_1 + m_2, \alpha_1 + \alpha_2)$.

5 Théorèmes limites I : convergence en loi

5.1 Convergence en loi (= convergence des lois)

Définition 5.1. Soit $(X_n)_{n\geq 1}$ une suite de v.a. à valeurs dans $\mathbb{R}^d$ et soit X une v.a. à valeurs dans $\mathbb{R}^d$. On dit que $(X_n)_{n\geq 1}$ *converge en loi* vers X (on notera $X_n \xrightarrow{\mathcal{L}} X$) si pour toute fonction $g : \mathbb{R}^d \to \mathbb{R}$ *continue et bornée* on a

$$\mathbb{E}(g(X_n)) = \int_{\mathbb{R}^d} g(x)d\mathbb{P}_{X_n}(x) \xrightarrow{n\to\infty} \mathbb{E}(g(X)) = \int_{\mathbb{R}^d} g(x)d\mathbb{P}_X(x)\,.$$

Note : on a besoin de prendre des fonctions continues car sinon la suite de v.a. (constantes) $X_n = 1/n$ ne convergerait pas vers 0 en loi !

L'hypothèse « g bornée » est aussi importante. On peut très bien avoir $X_n \xrightarrow{\mathcal{L}} 0$ mais $\mathbb{E}(X_n) \to \infty$ (par exemple : $\mathbb{P}(X_n = n^2) = 1/n$, $\mathbb{P}(X_n = 0) = 1 - 1/n$.)

Corollaire 5.2. *Si $X_n \xrightarrow{\mathcal{L}} X$ et si $h : \mathbb{R}^d \to \mathbb{R}^{d'}$ est continue, alors $h(X_n) \xrightarrow{\mathcal{L}} h(X)$.*

Démonstration. Si $g : \mathbb{R}^{d'} \to \mathbb{R}$ est continue et bornée, alors on peut appliquer la définition de la convergence en loi de X_n vers X avec la fonction $g \circ h : \mathbb{R}^d \to \mathbb{R}$ qui est continue et bornée. □

Note : on peut en fait remplacer « h continue » par « h est continue en X, p.s. » (le p.s. fait référence à la valeur de X).

Exercice 5.3. Trouver un exemple où l'on a $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{\mathcal{L}} Y$, mais pas $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, Y)$.

Solution. Prendre ξ une variable aléatoire symétrique ($\mathbb{P}(\xi = +1) = \mathbb{P}(\xi = -1) = \frac{1}{2}$ fait l'affaire), et poser, pour $n \geq 1$, $X_n = \xi$ et $Y_n = (-1)^n \xi$. Alors X_n, Y_n ont toujours la même loi que ξ, mais la loi de (X_n, Y_n) est alternativement (ξ, ξ) et $(\xi, -\xi)$, donc (X_n, Y_n) ne converge pas en loi.

Remarque 5.4. La définition de la convergence en loi ne concerne que la loi de X_n et la loi de X ! On devrait en réalité dire : « la loi de X_n converge vers la loi de X ». La notion de convergence de mesures que l'on utilise ici est celle de *convergence étroite*. Voici quelques notions de convergence de mesures $\mu_n \to \mu$ (de la plus faible à la plus forte) :

— μ_n converge *étroitement* vers μ si $\lim_{n\to\infty} \int_{\mathbb{R}^d} g(x) d\mu_n(x) = \int_{\mathbb{R}^d} g(x) d\mu_X(x)$ pour toute fonction *g continue et borné* ;

— μ_n converge *faiblement* vers μ si $\lim_{n\to\infty} \int_{\mathbb{R}^d} g(x) d\mu_n(x) = \int_{\mathbb{R}^d} g(x) d\mu_X(x)$ pour toute fonction *g continue telle que* $\lim_{\|x\|\to\infty} g(x) = 0$;

— μ_n converge *vaguement* vers μ si $\lim_{n\to\infty} \int_{\mathbb{R}^d} g(x) d\mu_n(x) = \int_{\mathbb{R}^d} g(x) d\mu_X(x)$ pour toute fonction *g continue à support compact.*

Remarque 5.5. Dans le cas de mesures de probabilité, les trois types de convergence sont équivalents. Il suffit de montrer que la convergence vague implique la convergence étroite. Supposons que μ_n converge étroitement ver μ et qu'il s'agit de mesures de probabilités : on supposera donnée une suite $(X_n)_{n\geq 1}$ de v.a. de lois μ_n et X une v.a. de loi μ. Soit g une fonction confinue bornée : on doit montrer que $\lim_{n\to\infty} \mathbb{E}(g(X_n)) = \mathbb{E}(g(X))$. Pour tout $\varepsilon \in]0, 1[$, soit ρ une fonction continue à support compact telle que $0 \leq \rho \leq 1$ et $1 \geq \mathbb{E}(\rho(X)) \geq 1 - \varepsilon$ (prendre ρ qui approche $\mathbb{1}_{[-R,R]}$ pour un R assez grand). Alors on peut majorer $|\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))|$ par

$$\begin{aligned}
&\Big|\mathbb{E}\big(g(X_n)(1 - \rho(X_n))\big)\Big| + \Big|\mathbb{E}\big(g(X_n)\rho(X_n)\big) - \mathbb{E}\big(g(X)\rho(X)\big)\Big| + \Big|\mathbb{E}\big(g(X)(1 - \rho(X))\big)\Big| \\
&\qquad \leq \|g\|_\infty \Big|\mathbb{E}(1 - \rho(X_n))\Big| + \Big|\mathbb{E}\big(g(X_n)\rho(X_n)\big) - \mathbb{E}\big(g(X)\rho(X)\big)\Big| + \|g\|_\infty \Big|\mathbb{E}(1 - \rho(X))\Big| .
\end{aligned}$$

Quand $n \to \infty$, comme ρ et $g\rho$ sont à support compact, on a par hypothèse $\mathbb{E}(\rho(X_n)) \to \mathbb{E}(\rho(X))$ et $\mathbb{E}(g(X_n)\rho(X_n)) \to \mathbb{E}(g(X)\rho(X))$ quand $n \to \infty$. Ainsi, on obtient

$$\limsup_{n\to\infty} \Big|\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))\Big| \leq 2\|g\|_\infty \Big|\mathbb{E}(1 - \rho(X))\Big| \leq 2\varepsilon \|g\|_\infty ,$$

ce qui conclut la démonstration, ε étant arbitraire.

Exercice 5.6. Montrer que si $\lim_{n\to\infty} \mathbb{E}(g(X_n)) = \mathbb{E}(g(X))$ pour toute fonction g C^∞ à support compact, alors $X_n \xrightarrow{\mathcal{L}} X$.

Exemple 5.7. Soit $(X_n)_{n\geq 1}$ une suite de v.a. de lois $\mathrm{Geo}(p_n)$, de paramètre p_n tel que $\lim_{n\to\infty} p_n = 0$. Montrons que la suite de v.a. $(Y_n := p_n X_n)_{n\geq 0}$ converge en loi vers $X \sim \mathcal{E}(1)$.

Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction continue à support compact (pour simplifier). On a alors

$$\mathbb{E}(g(Y_n)) = \mathbb{E}(g(p_n X_n)) = \sum_{k=1}^{\infty} g(p_n k)(1 - p_n)^{k-1} p_n = (1 + o(1)) \sum_{k=1}^{\infty} p_n g(p_n k) e^{-p_n k} .$$

où on a utilisé que $(1 - p_n)^k = e^{-p_n k + o(1)}$ pour $p_n k \leq A$ (où A est tel que $g(x) = 0$ pour $|x| > A$). On reconnaît une somme de Riemann : on obtient $\lim_{n\to\infty} \mathbb{E}(g(Y_n)) = \int_0^\infty g(x) e^{-x} dx = \mathbb{E}(g(X))$.

5.2 Autres critères de convergence en loi

5.2.1 Avec la fonction de répartition

Proposition 5.8. *La suite $(X_n) \xrightarrow{\mathcal{L}} X$ si et seulement si $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$ en tout point t où F_X est continue.*

Démonstration. Pour le sens direct, l'idée est d'approcher $\mathbb{1}_{]-\infty,t]}(x)$ par une fonction continue bornée. La réciproque est plus difficile... □

Il faut écarter les points de discontinuité de F_X : en effet si on considère les fonctions de répartition des v.a. constantes $X_n = 1/n$, elles convergent vers $\mathbb{1}_{]0,\infty[}$, qui n'est pas la fonction de répartition de la v.a. constante $X = 0$ (il y a un problème en $x = 0$).

Exemple 5.9. Si $(X_n)_{n\geq 1}$ et X sont des v.a. à valeurs dans $\mathbb{Z}$ alors $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$ pour tout $k \in \mathbb{Z}$.

En particulier, l'approximation de Poisson dit que $\mathrm{Bin}(n, \lambda/n) \xrightarrow{\mathcal{L}} \mathrm{Poi}(\lambda)$. On a aussi $X_n \xrightarrow{\mathcal{L}} \mathrm{Poi}(1)$, où X_n est le nombre de points fixes d'une permutation aléatoire de $\{1, \ldots, n\}$, voir l'Exemple 2.33.

Exemple 5.10. Problème du collectionneur de vignettes (références : [App15, p. 163]), voir l'Ex. 1.10. Un collectionneur essaie de compléter une collection de N vignettes : il achète chaque semaine un paquet, la vignette récupérée possédant une chance $\frac{1}{N}$ d'être la vignette k, pour $k \in \{1, \ldots, N\}$. On note T_N le nombre de semaines nécessaires pour compléter la collection. dans l'Exemple 1.10 on a vu que la formule d'inclusion-exclusion permet d'obtenir

$$\mathbb{P}(T_N \leq m) = \sum_{k=0}^{N} (-1)^{k+1} \binom{N}{k} \Big(1 - \frac{k}{N}\Big)^m .$$

On note que pour $x \in \mathbb{R}$, par convergence dominée on obtient

$$\lim_{N\to+\infty} \mathbb{P}\big(T_N \leq N \ln N + xN\big) = \sum_{k=0}^{\infty} (-1)^{k+1} \frac{1}{k!} e^{-kx} = e^{-e^{-x}} .$$

Ainsi, on a montré que si l'on pose F_n la fonction de répartition de $\frac{1}{N}(T_N - N\ln N)$, alors $\lim_{n\to\infty} F_n(x) = F(x) = e^{-e^{-x}}$. Il est facile de voir que F est la fonction de répartition d'une v.a. (il suffit de vérifier les trois propriétés des fonctions de répartition). Une telle v.a. est dite de *Gumbel*. On a donc montré que $\frac{1}{N}(T_N - N\ln N)$ converge en loi vers une v.a. de Gumbel, de fonction de répartition $F(x) = e^{-e^{-x}}$.

Exercice 5.11 (Un exercice classique). Soit $M_n := \max\{X_1, \ldots, X_n\}$, où les $(X_i)_{i\geq 1}$ sont des v.a. indépendantes de loi $\mathcal{E}(1)$. Montrer que $M_n - \ln n$ converge en loi vers une v.a. de Gumbel, de fonction de répartition $F(x) = 1 - e^{-e^{-x}}$.

Exercice 5.12. Soit $(X_n)_{n\geq 0}$ une suite de v.a. de densité f_n, et soit X une v.a. de densité f. On suppose que $f_n \to f$ p.p. : montrer que $X_n \xrightarrow{\mathcal{L}} X$.
Attention, la réciproque est fausse : considérer X_n de densité $f_n(x) = (1 - \cos(2\pi nx))\mathbb{1}_{[0,1]}(x)$.

5.2.2 Avec la fonction caractéristique

Théorème 5.13 (de Lévy). *On a $X_n \xrightarrow{\mathcal{L}} X$ si et seulement si $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$ pour tout $t \in \mathbb{R}^d$.*

Démonstration. On fait la démonstration dans de cas de v.a. à valeurs dans $\mathbb{R}$ pour simplifier. Le preuve utilise certaines propriétés de la transformée de Fourier qui seront vues plus loin dans le cours.

(i) Si $X_n \xrightarrow{\mathcal{L}} X$ alors $\lim_{n\to\infty} \phi_{X_n}(t) = \lim_{n\to\infty} \mathbb{E}(e^{itX_n}) = \mathbb{E}(e^{itX}) = \phi_X(t)$ car $x \mapsto e^{itx}$ est continue.

(ii) Réciproquement, soit $g : \mathbb{R} \to \mathbb{R}$ une fonction C^∞ à support compact (on a besoin d'avoir C^∞ pour hériter de bonnes propriétés de décroissance de la transformée de Fourier de g). D'après la formule d'inversion de Fourier, on a

$$g(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \hat{g}(t) dt , \qquad \text{où } \hat{g}(t) = \int_{\mathbb{R}} e^{itx} g(x) dx ,$$

et $\hat{g}$ est intégrable sur $\mathbb{R}$ car g est C^∞ (C^2 suffit). Ainsi, en appliquant Fubini

$$\mathbb{E}(g(X_n)) = \frac{1}{2\pi} \mathbb{E}\Big[\int_{\mathbb{R}} e^{-itX_n} \hat{g}(t) dt \Big] = \frac{1}{2\pi} \int_{\mathbb{R}} \phi_{X_n}(-t) \hat{g}(t) dt .$$

En utilisant la majoration $|\phi_{X_n}(-t)\hat{g}(t)| \leq |\hat{g}(t)|$ qui est intégrable, on en déduit par convergence dominée :

$$\lim_{n\to\infty} \mathbb{E}(g(X_n)) = \frac{1}{2\pi}\int_{\mathbb{R}} \phi_X(-t)\hat{g}(t)dt = \frac{1}{2\pi}\mathbb{E}\Big[\int_{\mathbb{R}} e^{-itX}\hat{g}(t)dt\Big] = \mathbb{E}(g(X))\,,$$

où on a de nouveau appliqué Fubini puis la formule d'inversion de Fourier pour g. □

Pour montrer une convergence en loi, on peut montrer la convergence des fonctions caractéristiques : $\phi_{X_n} \to \phi$. Mais rien ne dit a priori que ϕ est une fonction caractéristique. On admet le résultat suivant.

Proposition 5.14 (Critère de Lévy). *Si ϕ_{X_n} converge simplement vers ϕ et si ϕ est continue en* 0, *alors ϕ est la fonction caractéristique d'une v.a. X, et on a $X_n \xrightarrow{\mathcal{L}} X$.*

Exemple 5.15. Soit $(X_i)_{i\geq 1}$ une suite de v.a. indépendantes, de loi $\mathbb{P}(X_i = +1) = \mathbb{P}(X_1 = -1) = \frac{1}{2}$. On pose $S_0 = 0$ et pour $n \geq 1$ $S_n = \sum_{k=1}^n \frac{X_k}{k}$: il s'agit de la série harmonique de signe aléatoire.

Par indépendance, on a

$$\phi_{S_n}(t) = \prod_{k=1}^n \phi_{X_k/k}(t) = \prod_{k=1}^n \cos(t/k)$$

où on a utilisé un calcul exact : $\mathbb{E}(e^{itX_k/k}) = \frac{1}{2}e^{it/k} + \frac{1}{2}e^{it/k} = \cos(\frac{t}{k})$. Maintenant, on obtient facilement

$$\lim_{n\to\infty} \phi_{S_n}(t) = \phi(t) = \prod_{k=1}^\infty \cos\Big(\frac{t}{k}\Big)\,,$$

qui est bien défini car $\prod_{k=1}^\infty (1-x_k)$ est bien défini si et seulement si $\sum_k x_k < \infty$, ce qui est le cas ici car $1-\cos(\frac{t}{k}) \sim \frac{t^2}{2k^2}$ quand $k\to\infty$. Pour voir que ϕ est bien la fonction caractéristique d'une v.a., il reste à voir qu'elle est continue en 0. Mais pour $x \in [-1,1]$, on a $\cos(x) \geq 1 - \frac{x^2}{2} \geq \exp(-x^2)$: ainsi, pour $t \in [-1,1]$

$$1 \geq \phi(t) = \prod_{k=1}^\infty \cos\Big(\frac{t}{k}\Big) \geq \exp\Big(-\sum_{k=1}^\infty \frac{t^2}{k^2}\Big) = e^{-\frac{\pi^2 t}{6}}\,,$$

ce qui montre la continuité de ϕ en 0. Ainsi, ϕ est la fonction caractéristique d'une v.a. X, et $S_n \xrightarrow{\mathcal{L}} X$.

À noter que l'on sait majorer $\phi(t)$ pour montrer qu'elle est intégrable : cela signifie que la v.a. X admet une densité ! Pour la majoration de ϕ, voilà la méthode : on majore $|\cos(x)| \leq \exp(-cx^2)$ pour tout $x \in [-1,1]$ (en pratique, on peut prendre $c = 1/4$), d'où

$$|\phi(t)| = \prod_{k=1}^\infty \Big|\cos\Big(\frac{t}{k}\Big)\Big| \leq \prod_{k\geq|t|} |\cos(t/k)| \leq \exp\Big(-\sum_{k\geq|t|}\frac{t^2}{k^2}\Big) \leq \exp\big(-c'|t|\big)\,.$$

5.3 Théorème central limite et applications

Le théorème central limite (TCL) est l'un des théorèmes les plus fondamentaux des probabilités, aux applications sans limites. On notera i.i.d. = indépendantes et identiquement distribuées.

Théorème 5.16. *Soit $(X_k)_{k\geq 0}$ une suite de variables aléatoires i.i.d., admettant un moment d'ordre deux fini. On pose $\mu := \mathbb{E}(X_k)$ et $\sigma^2 = \mathrm{Var}(X_k)$. Alors*

$$Z_n := \frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma}\Big(\frac{1}{n}\sum_{k=1}^n X_k - \mu\Big) \xrightarrow[n\to\infty]{\mathcal{L}} Z \sim \mathcal{N}(0,1)\,.$$

Entre guillemets on a « $\sum_{k=1}^n X_k \approx n\mu + \sigma\sqrt{n}\,Z$ », ou encore « $\overline{X}_n := \frac{1}{n}\sum_{k=1}^n X_k \approx \mu + \frac{\sigma}{\sqrt{n}}Z$ »,

Remarque 5.17. De manière générale, on se ramènera dans la suite au cas variables aléatoires centrées et réduites, c'est-à-dire d'espérance nulle $\mu = \mathbb{E}(X_k) = 0$ et de variance unitaire $\sigma^2 = \mathrm{Var}(X_k) = 1$. Il suffit en effet de considérer les v.a. $Y_k = \frac{X_k - \mu}{\sigma}$ pour se ramener à ce cas : on parle de procédure de standarisation des v.a. X_k.

Remarque 5.18. Le théorème est trivial si $X_k \sim \mathcal{N}(0,1)$. En effet, dans ce cas on a *égalité en loi* entre $\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k$ et $Z \sim \mathcal{N}(0,1)$, sans même avoir à passer à la limite. La loi normale est un point fixe de l'opération de *normalisation* : $X_1,\dots,X_n$ i.i.d. (centrées réduites) $\mapsto \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i$ (aussi centrée réduite). Le théorème central limite nous dit que c'est le seul point fixe, et qu'il est attracteur.

Démonstration. Comme annoncé plus haut, on considère $(Y_k)_{k\ge 1}$ des v.a. i.i.d. centrées et réduites. On pose $Z_n := \frac{1}{\sqrt{n}}\sum_{k=1}^n Y_k$. La fonction caractéristique de Z_n, grâce à l'indépendance des $(Y_k)_{k\ge1}$, est donnée par

$$\phi_{Z_n}(t) = \mathbb{E}\Big(\exp\Big(\sum_{k=1}^n i\frac{t}{\sqrt{n}} Y_k\Big)\Big) = \prod_{k=1}^n \mathbb{E}\Big(e^{i\frac{t}{\sqrt{n}}Y_k}\Big) = \phi_{Y_1}\Big(\frac{t}{\sqrt{n}}\Big)^n .$$

Maintenant, on peut utiliser l'inégalité $|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i| \le \sum_{i=1}^n |z_i - w_i|$, valable pour tous complexes z_i, w_i de module $|z_i|,|w_i|\le 1$, qui se démontre facilement par récurrence sur n : on obtient

$$\Big|\phi_{Z_n}(t) - e^{-\frac12 t^2}\Big| \le n\Big|\phi_{Y_1}\Big(\frac{t}{\sqrt{n}}\Big) - e^{-\frac{t^2}{2n}}\Big| .$$

On peut alors effectuer un développement de Taylor, à t fixé, quand $n\to\infty$: on sait que $\phi'_{Y_1}(0) = (-i)\,\mathbb{E}(Y_1) = 0$ et $\phi''_{Y_1}(0) = (-i)^2\,\mathbb{E}(Y_1^2) = -1$, car Y_1 admet un moment d'ordre 2 fini, donc

$$\phi_{Y_1}\Big(\frac{t}{\sqrt{n}}\Big) = 1 + \frac{t}{\sqrt{n}}\phi'_{Y_1}(0) + (1+o(1))\frac{t^2}{2n}\phi''_{Y_1}(0) = 1-(1+o(1))\frac{t^2}{2n}, \qquad e^{-\frac{t^2}{2n}} = 1-(1+o(1))\frac{t^2}{2n}$$

(Noter que le $o(1)$ est un complexe de module $o(1)$.) On en déduit que

$$\lim_{n\to\infty} n\Big|\phi_{Y_1}\Big(\frac{t}{\sqrt{n}}\Big) - e^{-\frac{t^2}{2n}}\Big| = 0 ,$$

ce qui implique que $\lim_{n\to\infty}\phi_{Z_n}(t) = e^{-\frac12 t^2}$ pour tout $t\in\mathbb{R}$.

On aurait pu voir le résultat d'une autre manière. En prenant le logarithme (complexe) et en faisant un développement limité de ϕ_{Y_1} autour de 0 (attention il faut s'assurer que l'on garde la même détermination du logarithme complexe, c'est-à-dire que l'on prend n suffisamment grand de sorte que $\phi_{Y_1}\big(\frac{t}{\sqrt{n}}\big)$ soit de partie réelle positive), on obtient

$$\ln \phi_{Z_n}(t) = n\ln\phi_{Y_1}\big(\tfrac{t}{\sqrt{n}}\big) = n\ln\Big(1-(1+o(1))\frac{t^2}{2n}\Big) = -(1+(o(1)))\frac{t^2}{2} .$$

Cela montre que pour tout $t\in\mathbb{R}$ on a $\lim_{n\to\infty}\phi_{Z_n}(t) = e^{-\frac12 t^2} = \phi_Z(t)$, d'où la convergence en loi. □

Exercice 5.19. Le théorème central limite montre aussi la convergence (ponctuelle) des fonctions de répartition : $\lim_{n\to\infty} F_{Z_n}(x) = F_Z(x)$. On va montrer dans cet exercice que la convergence a lieu pour la norme uniforme : $\lim_{n\to\infty}\|F_{Z_n} - F_Z\|_\infty = 0$ (ref : [BCD21, p. 325]). Cela marche en réalité dès que l'on a F_Z continue—il s'agit d'un des théorèmes de Dini...

1. Soit $N\in\mathbb{N}^*$ et soient $x_1<\dots<x_{N-1}$ fixés. Montrer qu'il existe un n_0 tel que pour tout $n\ge n_0$ et $1\le k\le N-1$ on a $|F_{Z_n}(x_i) - F_Z(x_i)| \le \frac{1}{N}$.
2. Soit $1\le k\le N$ et soit $x\in]x_{k-1},x_k[$ (par convention $x_0=-\infty$ et $x_N=+\infty$). Montrer que pour $n\ge n_0$ on a $F_{Z_n}(x) - F_Z(x) \le \frac1N + F_Z(x_k) - F_Z(x_{k-1})$ et de même $F_Z(x) - F_{Z_n}(x_n) \le \frac1N + F_Z(x_k) - F_Z(x_{k-1})$. Conclure que pour $n\ge n_0$
$$\|F_{Z_n} - F_Z\|_\infty \le \frac1N + \sup_{1\le k\le N}|F_Z(x_{k+1}) - F_Z(x_k)| .$$
3. Montrer que l'on peut choisir $x_1<\dots<x_{N-1}$ tels que $\sup_{1\le k\le N}|F_Z(x_{k+1}) - F_Z(x_k)| = \frac1N$. Conclure.

Remarque 5.20. Il existe un théorème (de Berry–Esseen) qui contrôle la vitesse de convergence de la fonction de répartition. Voici un énoncé. Soit $(X_k)_{k\ge1}$ une suite de v.a. i.i.d. centrées réduites, admettant un moment d'ordre 3 fini. On pose $Z_n := \frac{1}{\sqrt{n}}\sum_{k=1}^n X_k$ et soit $Z\sim\mathcal{N}(0,1)$. Alors

$$\|F_{Z_n} - F_Z\|_\infty \le \frac{1}{2\sqrt{n}}\mathbb{E}(|X_1|^3) .$$

5.3.1 Applications en statistique

Le TCL permet de donner des intervalles de confiance pour l'estimation de paramètres. Prenons l'exemple le plus standard : on a des v.a. $(X_i)_{i\geq 1}$ i.i.d. de Bernoulli de paramètre $p \in [0,1]$.

L'énoncé du TCL nous dit que, pour tout $x > 0$

$$\lim_{n\to\infty} \mathbb{P}\bigg(\frac{\sqrt{p(1-p)}}{\sqrt{n}} \Big| \frac{1}{n}\sum_{i=1}^n X_i - p \Big| \leq x \bigg) = \lim_{n\to\infty} \mathbb{P}\bigg(\Big| \frac{1}{n}\sum_{i=1}^n X_i - p \Big| \leq \frac{x\sqrt{p(1-p)}}{\sqrt{n}} \bigg) = \mathbb{P}(|Z| \leq x)\,.$$

On peut voir la probabilité $\mathbb{P}\Big(\Big| \frac{1}{n}\sum_{i=1}^n X_i - p \Big| \leq \frac{x\sqrt{p(1-p)}}{\sqrt{n}} \Big)$ de deux façons différentes :

— soit on connaît le paramètre p, auquel cas le TCL nous donne la probabilité que $\frac{1}{n}\sum_{i=1}^n X_i$ soit proche de p (c'est-à-dire à une distance inférieure à $\frac{x\sqrt{p(1-p)}}{\sqrt{n}}$) ;

— soit on ne connaît pas le paramètre p mais on a observé la valeur de $X_1, \dots, X_n$: dans ce cas le TCL nous donne la probabilité que p soit proche de $\frac{1}{n}\sum_{i=1}^n X_i$ (c'est-à-dire à une distance inférieure à $\frac{x\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{x}{2\sqrt{n}}$, car $p(1-p) \leq \frac{1}{4}$).

Autrement dit, le TCL nous permet de dire que pour n grand,

$$\begin{aligned}
\mathbb{P}(|Z| \geq x) &\approx \mathbb{P}\bigg(\Big| \frac{1}{n}\sum_{i=1}^n X_i - p \Big| \leq \frac{x\sqrt{p(1-p)}}{\sqrt{n}} \bigg) \\
&\leq \mathbb{P}\bigg(\Big| \frac{1}{n}\sum_{i=1}^n X_i - p \Big| \leq \frac{x}{2\sqrt{n}} \bigg) = \mathbb{P}\bigg(p \in \Big[\frac{1}{n}\sum_{i=1}^n X_i - \frac{x}{2\sqrt{n}}, \frac{1}{n}\sum_{i=1}^n X_i + \frac{x}{2\sqrt{n}} \Big] \bigg)\,.
\end{aligned}$$

On dit que $\Big[\frac{1}{n}\sum_{i=1}^n X_i - \frac{x}{2\sqrt{n}}, \frac{1}{n}\sum_{i=1}^n X_i + \frac{x}{2\sqrt{n}} \Big]$ est un *intervalle de confiance* (asymptotique) pour p, de niveau de confiance $\gamma = \mathbb{P}(|Z| \geq x)$: cela signifie que le paramètre p a une probabilité supérieure à γ de tomber dans cet intervalle.

En pratique, on se fixe un niveau de confiance (par exemple $\gamma = 95\%$) ce qui nous permet de trouver le $x > 0$ tel que $\mathbb{P}(|Z| \geq x) = \gamma$ (on appelle cela un quantile) : on le trouve dans la table de la loi normale (pour $\gamma = 95\%$ on a $x \approx 1{,}96$). Cela nous fournit donc un intervalle de confiance pour p de niveau de confiance γ.

Exemple 5.21. Reprenons l'exemple des naissances. Soit $(X_i)_{i\geq 1}$ une suite de v.a. i.i.d. de loi $\mathrm{Ber}(p)$, où $X_i = 1$ signifie que la i-ème naissance est un garçon et $X_i = 0$ que la i-ème naissance est une fille. On a observé que pour $n = 700\,000$ naissances en France, on a eu $\sum_{i=1}^n X_i = 355\,000$ garçons. D'après ce qui précède, un intervalle de confiance de niveau 95% pour la probabilité p d'avoir un garçon est donc donné par

$$\Big[\frac{1}{n}\sum_{i=1}^n X_i - \frac{1{,}96}{2\sqrt{n}}, \frac{1}{n}\sum_{i=1}^n X_i + \frac{1{,}96}{2\sqrt{n}} \Big] \approx \big[0{,}5083 \pm 0,!0012 \big]\,.$$

Noter que 1/2 n'appartient pas à cet intervalle ! Il y a donc une niveau de confiance d'au moins 95% qu'il n'y ait pas une probabilité 1/2 d'avoir un garçon ou une fille.

Exemple 5.22. Reprenons l'exemple des naissances : faisons l'hypothèse (H) : « les naissances sont i.i.d. de loi Bern(1/2) ». Sous cette hypothèse, on peut appliquer le TCL, qui nous donne

$$\lim_{n\to\infty} \mathbb{P}\bigg(\Big| \frac{\sum_{i=1}^n X_i - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \Big| \geq x \bigg) = \lim_{n\to\infty} \mathbb{P}\Big(\Big| \sum_{i=1}^n X_i - \frac{n}{2} \Big| \geq \frac{x\sqrt{n}}{2} \Big) = \mathbb{P}(|Z| \geq x)\,.$$

On se fixe à l'avance un seuil α (disons égal à 1%) et on rejettera l'hypothèse (H) si on observe un nombre de garÁons ou de filles s'écartant de $n/2$ de plus de x_α, avec x_α tel que $\mathbb{P}(|Z| \geq x_\alpha|) = \alpha$. Avec les chiffres de l'exemple ci-dessus, on observe $\sum_{i=1}^n X_i \geq 355\,000$, c'est-à-dire $|\sum_{i=1}^n X_i - \frac{n}{2}| \geq \frac{x\sqrt{n}}{2}$ avec $\frac{x\sqrt{n}}{2} = 5000$, c'est-à-dire $x \approx 11{,}95$. Ainsi, sous l'hypothèse (H), on a $\mathbb{P}(A) \approx \mathbb{P}(|Z| \geq 11{,}95) \approx 10^{-64}$... On peut donc raisonnablement rejeter l'hyopthèse (H).

6 Théorèmes limites II : convergence dans L^p, en probabilité, p.s.

La convergence en loi ne concerne que les lois des v.a. et pas la convergence des v.a. elles-même. En particulier, les v.a. pourraient ne rien avoir l'une avec l'autre et en particulier ne pas être définies sur le même espace probabilisé. On va voir dans cette section des notions de convergence dites « trajectorielles » dans le sens où elles dépendent de la trajectoire de la suite $(X_n)_{n\geq 1}$ en tant que v.a. à valeurs dans les suites de $\mathbb{R}^d$ (on considérera $(X_n(\omega))_{n\geq 1}$).
Important. On ne le précisera pas à chaque fois, mais dans ce qui suit les v.a. $(X_n)_{n\geq 1}$ et la v.a. limite X sont définies sur *le même espace* $(\Omega, \mathcal{F}, \mathbb{P})$: on considère des propriétés de la suite $(X_n(\omega))_{n\geq 1}$

6.1 Convergence dans L^p

Commençons par la notion de convergence définie par la norme L^p. On ne considérera que le cas $p \geq 1$ car on rappelle que dans le cas $p < 1$, $\|X\|_p = \mathbb{E}(|X|^p)^{1/p}$ ne satisfait pas l'inégalité triangulaire et ne définit donc pas une norme.

Définition 6.1. On dit que la suite $(X_n)_{n\geq 1}$ converge *dans* L^p vers X, et on notera $X_n \xrightarrow{L^p} X$, si $\lim_{n\to\infty} \|X_n - X\|_p = 0$, ou de manière équivalente si $\lim_{n\to\infty} \mathbb{E}\big(|X_n - X|^p\big) = 0$.

Quelques observations importantes :

— Si $1 \leq q \leq p$ et $X_n \xrightarrow{L^p} X$, alors $X_n \xrightarrow{L^q} X$. En effet, on a montré que $\|X_n - X\|_q \leq \|X_n - X\|_p$.

— On rappelle que pour $p \geq 1$, $(L^p, \|\cdot\|_p)$ est complet.

— Si $X_n \xrightarrow{L^p} X$, alors $\lim_{n\to\infty} \|X_n\|_p = \|X\|_p$ (car $\|X_n - X\|_p \geq |\|X_n\|_p - \|X\|_p|$). Attention, la réciproque est fausse !

Remarque 6.2. Comme on le verra plus bas, la convergence dans L^p implique (la convergence en probabilité qui elle-même implique) la convergence en loi : si $X_n \xrightarrow{L^p} X$ alors $X_n \xrightarrow{\mathcal{L}} X$.

Exemple 6.3. Soit $(X_n)_{n\geq 1}$ une suite de v.a. telle que $\mathbb{P}(X_n = n) = 1/n^\alpha$ et $\mathbb{P}(X_n = 0) = 1 - 1/n^\alpha$, où $\alpha > 0$ est un réel donné. Notons que l'on a toujours $X_n \xrightarrow{\mathcal{L}} 0$, donc si X_n converge dans L^p, c'est nécessairement vers 0.

On a $\mathbb{E}(|X_n|^p) = n^p \mathbb{P}(X_n = n) = n^{p-\alpha}$, qui tend vers 0 si et seulement si $\alpha > p$. Ainsi, $X_n \xrightarrow{L^p} X$ pour tout $p < \alpha$ (si $\alpha > 1$) et la suite $(X_n)_{n\geq 1}$ ne converge pas dans L^p si $p \geq \alpha$.

Exercice 6.4. Soit $(X_n)_{n\geq 1}$ une suite de v.a. qui converge dans L^p vers une v.a. X. Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction mesurable telle que pour tout x, y on ait $|g(x) - g(y)| \leq (h(x) + h(y))|x - y|$, pour une fonction h telle que $h(z) = O(z^{p-1})$ quand $\|z\| \to \infty$. Montrer que $\lim_{n\to\infty} \mathbb{E}(g(X_n)) = \mathbb{E}(g(X))$.
En particulier, on a $\lim_{n\to\infty} \mathbb{E}(X_n^p) = \mathbb{E}(X^p)$.
Solution. On a

$$\begin{aligned}\Big|\mathbb{E}\big(g(X_n) - g(X)\big)\Big| &\leq \mathbb{E}\big(h(X_n)\,|X_n - X|\big) + \mathbb{E}\big(h(X)\,|X_n - X|\big) \\ &\leq \Big(\mathbb{E}\big(h(X_n)^{\frac{p}{p-1}}\big)^{\frac{p-1}{p}} + \mathbb{E}\big(h(X)^{\frac{p}{p-1}}\big)^{\frac{p-1}{p}}\Big)\mathbb{E}\big(|X_n - X|^p\big)^{\frac{1}{p}}\end{aligned}$$

où on a appliqué Hölder avec les exposants p et $q = \frac{p}{p-1}$ pour chacun des deux termes. Maintenant, on utilise que $h(x) \leq C|x|^{p-1}$ pour obtenir

$$\Big|\mathbb{E}\big(g(X_n) - g(X)\big)\Big| \leq \Big(\mathbb{E}\big(|X_n|^p\big)^{\frac{p-1}{p}} + \mathbb{E}\big(|X|^p\big)^{\frac{p-1}{p}}\Big)\mathbb{E}\big(|X_n - X|^p\big)^{\frac{1}{p}}.$$

Comme $\lim_{n\to\infty} \mathbb{E}(|X_n|^p) = \mathbb{E}(|X|^p) < +\infty$ et que $\lim_{n\to\infty} \mathbb{E}(|X_n - X|^p) = 0$, on obtient la conclusion.

Pour $g(x) = x^p$, on a d'après le théorème des accroissements finis si $x < y$ on a $g(x) - g(y) = pz^{p-1}(x - y)$ pour un $z \in [x, y]$. On en déduit que $|g(x) - g(y)| \leq p(|x^{p-1} + |y|^{p-1})|x - y|$.

Exercice 6.5. Soit $(X_n)_{n\geq 1}$ une suite de v.a. qui converge dans L^p vers une v.a. X non constante p.s., et soit $\tilde{X}$ une v.a. indépendante de X, de même loi que X. Montrer que $(X_n)_{n\geq 1}$ ne converge pas dans L^p vers $\tilde{X}$, mais que $\lim_{n\to\infty} \mathbb{E}(g(X_n)) = \mathbb{E}(g(\tilde{X}))$ pour toute fonction g .

6.2 Convergence en probabilité

Définition 6.6. On dit que la suite $(X_n)_{n\geq 1}$ converge *en probabilité* vers X, et on notera $X_n \xrightarrow{\mathbb{P}} X$, si pour tout $\varepsilon > 0$ on a $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.

Exemple 6.7. Soit $(X_n)_{n\geq 1}$ une suite de v.a. telle que $\mathbb{P}(X_n = n) = 1/n^\alpha$ et $\mathbb{P}(X_n = 0) = 1 - 1/n^\alpha$, où $\alpha > 0$ est un réel donné. Alors pour tout $\varepsilon > 0$, on a

$$\mathbb{P}(|X_n| > \varepsilon) \leq \mathbb{P}(|X_n| \neq 0) = 1/n^\alpha \xrightarrow{n\to\infty} 0\,,$$

donc X_n converge en probabilité vers 0.

Remarque 6.8. La convergence dans L^p implique la convergence en probabilité, qui implique la convergence en loi. Il existe des réciproques partielles à chacune de ces implications. On le laisse en exercice pour l'instant, on le démontre plus bas.

Exemple 6.9. Soit $(X_n)_{n\geq 1}$ une suite de v.a. qui converge en probabilité vers X. Montrer que si $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ est continue, on a $f(X_n) \xrightarrow{\mathbb{P}} f(X)$.
Indication : on pourra commencer par une fonction continue à support compact.

La proposition suivante montre que la convergence en probabilité est associée à une distance d : on peut voir la convergence en probabilité comme la convergence dans l'espace métrique (L^0, d) (on rappelle que L^0 est l'espace des v.a. à valeurs dans $\mathbb{R}^d$, où on a identifié deux v.a. égales p.s.).

Proposition 6.10. *Pour deux v.a. $X, Y \in L^0$, on définit $d(X,Y) = \mathbb{E}(|X-Y| \wedge 1)$, où $a \wedge b := \min(a,b)$. Alors $d(\cdot,\cdot)$ est une distance sur L^0 qui caractérise la convergence en probabilité :*

$$X_n \xrightarrow{\mathbb{P}} X \qquad \Longleftrightarrow \qquad \lim_{n\to\infty} d(X_n, X) = 0\,.$$

Exercice 6.11. Montrer que la même proposition est valable avec les distances $d_1(X,Y) = \mathbb{E}(\frac{|X-Y|}{1+|X-Y|})$ et $d_2(X,Y) = \inf\{\alpha\,, \mathbb{P}(|X-Y| > \alpha) \geq \alpha\}$.

Démonstration. Le fait que $d(X,Y)$ définisse une distance est trivial.
(i) Si $X_n \xrightarrow{\mathbb{P}} X$. Soit $\varepsilon > 0$. En décomposant l'espérance, on obtient

$$\begin{aligned} d(X_n, X) = \mathbb{E}\big(|X_n - X| \wedge 1\big) &= \mathbb{E}\big(|X_n - X| \wedge 1 \mathbb{1}_{\{|X_n - X| \leq \varepsilon\}}\big) + \mathbb{E}\big(|X_n - X| \wedge 1 \mathbb{1}_{\{|X_n - X| > \varepsilon\}}\big) \\ &\leq \varepsilon \mathbb{P}\big(|X_n - X| \leq \varepsilon\big) + \mathbb{P}\big(|X_n - X| > \varepsilon\big)\,. \end{aligned}$$

En utilisant la définition de la convergence en probabilité, on a $\limsup_{n\to\infty} d(X_n, X) \leq \varepsilon$: cela conclut la preuve de la première implication, ε étant arbitraire.
(ii) Supposons que $\lim_{n\to\infty} d(X_n, X) = 0$. Soit $\varepsilon \in]0,1[$. Alors, grâce à l'inégalité de Markov, on a

$$\mathbb{P}\big(|X_n - X| > \varepsilon\big) = \mathbb{P}\big(|X_n - X| \wedge 1 > \varepsilon\big) \leq \frac{1}{\varepsilon}\mathbb{E}\big(|X_n - X| \wedge 1\big) = \frac{1}{\varepsilon} d(X_n, X)\,.$$

En prenant la limite $n \to \infty$, cela montre que $X_n \xrightarrow{\mathbb{P}} X$. □

Théorème 6.12. *L'espace métrique (L^0, d) est complet.*

Démonstration. La démonstration est proche de ce qui est fait pour les espaces L^p. La principale difficulté est de construire la limite potentielle. Ref : [App15, p.391].

Soit $(X_n)_{n\geq 1}$ une suite de Cauchy pour la distance d. On peut trouver une extractrice $(n_k)_{k\geq 1}$ telle que pour tout $k \geq 1$

$$d(X_{n_{k+1}}, X_{n_k}) \leq 2^{-k}\,.$$

Par convergence monotone, cela montre que

$$\mathbb{E}\bigg(\sum_{k=1}^{\infty} |X_{n_{k+1}} - X_{n_k}| \wedge 1\bigg) = \sum_{k=1}^{\infty} d(X_{n_{k+1}}, X_{n_k}) < +\infty\,.$$

En particulier, $\sum_{k=1}^{\infty} |X_{n_{k+1}} - X_{n_k}| \wedge 1 < +\infty$ p.s., donc p.s. $|X_{n_{k+1}} - X_{n_k}| < 1$ à partir d'un certain rang, de sorte que l'on a $\sum_{k=1}^{\infty} |X_{n_{k+1}} - X_{n_k}| < +\infty$ p.s. Cela montre que p.s., la série

$$X = X_{n_1} + \sum_{k=1}^{\infty} (X_{n_{k+1}} - X_{n_k}),$$

converge absolument. Par construction, on a

$$d(X_{n_j}, X) \le \sum_{k=j}^{\infty} d(X_{n_{k+1}}, X_{n_k}) \xrightarrow{j\to\infty} 0.$$

La suite $(X_{n_j})_{j\ge1}$ converge donc vers X, et comme il s'agit d'une suite de Cauchy dans un espace métrique, la suite elle-même $(X_n)_{n\ge0}$ converge vers X. □

Exercice 6.13. Soit $(X_k)_{k\ge1}$ une suite de v.a. indépendantes de loi Cauchy$(0,1)$; en particulier $X_k \notin L^1$. On pose $S_0 = 0$ et pour $n \ge 1$ on pose $S_n = \sum_{k=1}^{n} \alpha_k X_k$, où $\alpha_k > 0$ pour tout $k > 0$.

1. Montrer que si $\alpha := \sum_{k=1}^{\infty} \alpha_k < \infty$, alors $S_n \xrightarrow{\mathbb{P}} X$, où $X \sim$ Cauchy$(0,\alpha)$.
2. Montrer que si $\sum_{k=1}^{\infty} \alpha_k = \infty$, alors X_n ne converge pas en loi, donc pas en probabilité.

Solution. Commençons par noter que l'on sait calculer la fonction caractéristique de S_n : en se rappelant qu'une loi de Cauchy de paramètre a est de fonction caractéristique $e^{-a|t|}$, on en déduit que $\phi_{S_n}(t) = e^{-\sum_{k=1}^{n} \alpha_k |t|}$, donc $S_n \sim$ Cauchy$(0, \sum_{k=1}^{n} \alpha_k)$. De la même manière, pour $m < n$ on a $S_n - S_m \sim$ Cauchy$(0, \sum_{k=m+1}^{n} \alpha_k)$. Notons aussi que si $X \sim$ Cauchy$(0,a)$ alors $\frac{1}{a} X$Cauchy$(0,1)$.

1. Commençons par le cas $\alpha := \sum_{k=1}^{\infty} \alpha_k < \infty$. Notons que $\lim_{n\to\infty} \phi_{S_n}(t) = e^{-\alpha|t|}$, qui est la fonction caractéristique d'une v.a. Cauchy$(0,\alpha)$: on a donc $S_n \xrightarrow{\mathcal{L}} X \sim$ Cauchy$(0,\alpha)$. Montrons maintenant que $(S_n)_{n\ge1}$ est une suite de Cauchy pour la distance $d(\cdot,\cdot)$, ce qui montrera que S_n converge en probabilité vers une v.a. X : comme S_n doit converger en loi vers la même v.a. X, on aura $S_n \xrightarrow{\mathbb{P}} X \sim$ Cauchy$(0,\alpha)$.

Pour $m < n$, notons que $S_n - S_m \sim$ Cauchy$(0,a)$ avec $a = a_{m,n} := \sum_{k=m+1}^{n} \alpha_k$, de sorte que $\frac{1}{a}(S_n - S_m) \sim Z$, où $Z \sim$ Cauchy$(0,1)$. Ainsi

$$d(S_n, S_m) = \mathbb{E}\big(|S_n - S_m| \wedge 1\big) = a\mathbb{E}\Big(|Z| \wedge \frac{1}{a}\Big) = a\bigg(\mathbb{E}(|Z| \mathbb{1}_{\{Z \le 1/a\}}) + \frac{1}{a}\mathbb{P}(|Z| \ge 1/a)\bigg).$$

Maintenant, notons que $\mathbb{E}(|Z|\mathbb{1}_{\{Z\le A\}}) = 2\int_0^A \frac{z}{\pi(1+z^2)} dz = \frac{1}{\pi}\ln(1+A^2)$ et $\mathbb{P}(|Z| \ge A) = 2\int_A^{\infty} \frac{1}{\pi(1+z^2)} dz \le \frac{2}{\pi}\int_A^{\infty} z^{-2} dz = \frac{2}{\pi A}$. On en conclut que

$$d(S_n, S_m) \le \frac{a}{\pi}\Big(\ln\big(1 + 1/a^2\big) + 2\Big).$$

Comme $\sup_{n\ge m} a_{n,m} = \sum_{k=m+1}^{\infty} \alpha_k \to 0$ quand $m \to \infty$, on a $\lim_{m\to\infty} \sup_{n\ge m} d(S_n, S_m) = 0$: cela montre que $(S_n)_{n\ge1}$ est une suite de Cauchy pour la distance $d(\cdot,\cdot)$ et conclut la démonstration.

2. Dans le cas $\sum_{k=1}^{\infty} \alpha_k = +\infty$, on a $\lim_{n\to\infty} \phi_{S_n}(t) = 1$ si $t = 0$ et $\lim_{n\to\infty} \phi_{S_n}(t) = 0$ si $t \ne 0$. La fonction $x \mapsto \mathbb{1}_{\{0\}}(x)$ n'est pas continue en 0, donc $\lim_{n\to\infty} \phi_{S_n}$ converge vers une fonction qui n'est pas une fonction caractéristique. Ainsi, la suite $(S_n)_{n\ge0}$ ne peut pas converger en loi, donc pas non plus en probabilité.

6.3 Convergence presque sûre

Définition 6.14. On dit que la suite $(X_n)_{n\ge1}$ converge *presque sûrement* vers X, on notera $X_n \xrightarrow{p.s.} X$, si $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.

Remarque 6.15. On verra plus bas que la convergence p.s. implique la convergence en probabilité (donc la convergence en loi), mais que la convergence p.s. n'implique pas forcément la convergence dans L^p.

Notons que l'on a facilement que si $X_n \xrightarrow{p.s.} X$ et $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ est une fonction continue, alors $f(X_n) \xrightarrow{p.s.} f(X)$.

Exercice 6.16. Montrer que l'on a la convergence $X_n \xrightarrow{p.s.} X$ si et seulement si, pour tout $\varepsilon > 0$ on a $\lim_{n\to\infty} \mathbb{P}\big(\sup_{k\geq n} |X_k - X| > \varepsilon\big) = 0$. (Comparer à la définition de la convergence en probabilité.)

Solution. Notons que l'on a $\lim_{n\to\infty} X_n = X$ si et seulement si pour tout $\varepsilon > 0$ avec $\varepsilon \in \mathbb{Q}$ il existe un n tel que $\sup_{k\geq n} |X_n - X| \leq \varepsilon$. On a donc $\mathbb{P}(\lim_{n\to\infty} X_n = X)$ si et seulement si pour tout $\varepsilon > 0$, $\varepsilon \in \mathbb{Q}$ on a $\mathbb{P}(\bigcup_{n\geq 1} A_{n,\varepsilon}) = 1$, où $A_{n,\varepsilon} = \{\sup_{k\geq n} |X_n - X| \leq \varepsilon\}$. La suite dévénements $(A_{n,\varepsilon})_{n\geq 1}$ étant croissante, on a $\mathbb{P}(\bigcup_{n\geq 1} A_n) = \lim_{n\to\infty} \mathbb{P}(A_n)$. Ainsi, $X_n \xrightarrow{p.s.} X$ si et seulement si, pour tout $\varepsilon > 0$ on a $\lim_{n\to\infty} \mathbb{P}(A_{n,\varepsilon}) = 1$, ce qui est exactement le résultat voulu (modulo un passage au complémentaire).

Exemple 6.17. Soit $(X_n)_{n\geq 1}$ une suite de v.a. i.i.d., de loi $\mathcal{U}([0,1])$. On définit la suite $(Y_n)_{n\geq 1}$ en posant $Y_n = \prod_{i=1}^n X_i$. Alors $(Y_n)_{n\geq 1}$ est clairement une suite décroissante (car $X_n \leq 1$ pour tout n), donc $(Y_n)_{n\geq 0}$ converge presque sûrement, vers une v.a. $Y := \lim_{n\to\infty} Y_n$. Maintenant, grâce au lemme de Fatou (les Y_n sont positifs), on a

$$\mathbb{E}(Y) = \mathbb{E}(\liminf_{n\to\infty} Y_n) \leq \liminf_{n\to\infty} \mathbb{E}(Y_n)\,.$$

Par indépendance, on a $\mathbb{E}(Y_n) = \mathbb{E}(X_1)^n = (\frac{1}{2})^n$, donc $\mathbb{E}(Y) \leq \lim_{n\to\infty} \mathbb{E}(Y_n) = 0$. La v.a. Y étant positive, on a $Y = 0$ p.s. En conclusion, on a montré que $Y_n \xrightarrow{p.s.} 0$.

Remarque 6.18. Pour démontrer des convergences p.s. un outil redoutable, qu'on utilisera souvent, est le lemme de Borel–Cantelli. On a par exemple le critère suivant (une condition suffisante pour la convergence p.s.) :

Si pour tout $\varepsilon > 0$ la série $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon)$ converge, alors $X_n \xrightarrow{p.s.} X$.

En effet, par Borel–Cantelli, pour tout $\varepsilon > 0$, p.s. on a $|X_n - X| > \varepsilon$ un nombre fini de fois. Autrement dit, $\forall \varepsilon > 0$, $\mathbb{P}(\exists n_0\,, \forall n \geq n_0$ on a $|X_n - X| \leq \varepsilon) = 1$. En prenant $\varepsilon \in \mathbb{Q}$ (de sorte à avoir une quantité dénombrable ; voir l'Exercice 1.14), on a $\mathbb{P}(\forall \varepsilon \in \mathbb{Q}, \exists n_0\,, \forall n \geq n_0$ on a $|X_n - X| \leq \varepsilon) = 1$, qui est exactement $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.

Exercice 6.19 (Un classique)**.** Soit $(X_n)_{n\geq 0}$ une suite de v.a. indépendantes, avec X_n de loi $\mathrm{Bern}(p_n)$. On suppose que $p_n \to 0$.

1. Montrer que $X_n \xrightarrow{\mathbb{P}} 0$.
2. Si $\sum_{n\geq 1} p_n < +\infty$, montrer que $X_n \xrightarrow{p.s.} 0$.
3. Si $\sum_{n\geq 1} p_n = +\infty$, montrer que $\limsup_{n\geq 1} X_n = 1$ p.s., donc $(X_n)_{n\geq 1}$ ne converge pas p.s.

6.4 Liens entre les différents types de convergence

On a déjà énoncé plus haut certaines implications de convergence. On les rassemble ici dans un diagramme (les démonstrations se trouvent plus bas).

$$\begin{array}{ccccc}
X_n \xrightarrow{p.s.} X & \Longrightarrow & X_n \xrightarrow{\mathbb{P}} X & \Longrightarrow & X_n \xrightarrow{\mathcal{L}} X \\
 & & \Uparrow & & \\
X_n \xrightarrow{L^q} X & \overset{q>p}{\Longrightarrow} & X_n \xrightarrow{L^p} X & &
\end{array}$$

Chacune de ces implications possède une réciproque partielle, qui est énoncée plus bas. La convergence dans L^p et la convergence p.s. n'ont pas de lien d'implication.

6.4.1 Convergence dans L^p implique convergence en probabilité, réciproque partielle

Proposition 6.20. *Si $X_n \xrightarrow{L^p} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.*

Démonstration. Soit $\varepsilon > 0$. En appliquant l'inégalité de Markov, on a

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - X|^p > \varepsilon^p) \leq \frac{1}{\varepsilon^p}\mathbb{E}\big(|X_n - X|^p\big) \xrightarrow{n\to\infty} 0\,.$$

□

Exemple 6.21. Voici un contre-exemple à la réciproque. Soit $(X_n)_{n\geq 1}$ une suite de v.a. définies par $\mathbb{P}(X_n = n) = p_n$ et $\mathbb{P}(X = 0) = 1 - p_n$. Il est très facile de voir que $X_n \xrightarrow{\mathbb{P}} 0$ dès que $p_n \to 0$. Par contre on a $\mathbb{E}(|X_n|^p) = p_n\, n^p$, ce qui montre que $(X_n)_{n\geq 1}$ ne converge pas dans L^p dès que $p_n n^p$ ne tend pas vers 0 (par exemple $p_n = 1/\ln n$).

Une réciproque partielle est la suivante.

Proposition 6.22. *Si $X_n \xrightarrow{\mathbb{P}} X$ et si $\sup_{n\geq 1} \mathbb{E}(|X_n|^r) < +\infty$ pour un $r > p$, alors $X_n \xrightarrow{L^p} X$.*

Démonstration. Soit $\varepsilon > 0$ fixé. On utilise la décomposition suivante :

$$\begin{aligned}
\mathbb{E}\big(|X_n - X|^p\big) &= \mathbb{E}\big(|X_n - X|^p \mathbb{1}_{\{|X_n - X| \leq \varepsilon\}}\big) + \mathbb{E}\big(|X_n - X|^p \mathbb{1}_{\{|X_n - X| > \varepsilon\}}\big) \\
&\leq \varepsilon^p + \mathbb{E}\big(|X_n - X|^r\big)^{p/r} \mathbb{E}\big(\mathbb{1}_{\{|X_n - X| > \varepsilon\}}\big)^{1-p/r}
\end{aligned}$$

où pour la deuxième inégalité on a appliqué l'inégalité de Hölder avec $q_1 = r/p$, $q_2 = r/(r-p)$ ($\frac{1}{q_1} + \frac{1}{q_2} = 1$). Comme $\mathbb{E}\big(|X_n - X|^r\big)^{1/r} = \|X_n - X\|_r \leq \|X_n\|_r + \|X\|_r$ est borné par une constante (uniformément en n) et que $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$, on obtient que $\limsup_{n\to\infty} \mathbb{E}\big(|X_n - X|^p\big) \leq \varepsilon^p$. Comme ε est arbitraire, cela conclut la démonstration. □

6.4.2 Convergence en probabilité implique convergence en loi, réciproque partielle

Proposition 6.23. *Si $X_n \xrightarrow{\mathbb{P}} X$, alors $X_n \xrightarrow{\mathcal{L}} X$.*

Démonstration. Soit $g : \mathbb{R}^d \to \mathbb{R}$ une fonction continue à support compact. Soit $\varepsilon > 0$ fixé. Comme g est uniformément continue, il existe $\delta > 0$ tel que $|x - x'| \leq \delta$ implique $|g(x) - g(x')| \leq \varepsilon$. On peut alors écrire

$$\begin{aligned}
\Big|\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))\Big| &\leq \mathbb{E}\big(|g(X_n) - g(X)|\big) \\
&= \mathbb{E}\big(|g(X_n) - g(X)| \mathbb{1}_{\{|X_n - X| \leq \delta\}}\big) + \mathbb{E}\big(|g(X_n) - g(X)| \mathbb{1}_{\{|X_n - X| > \delta\}}\big) \\
&\leq \varepsilon + \|g\|_\infty \mathbb{P}\big(|X_n - X| > \delta\big).
\end{aligned}$$

Comme $X_n \xrightarrow{\mathbb{P}} X$, en prenant la limite $n \to \infty$ on obtient $\limsup_{n\to\infty} \Big|\mathbb{E}(g(X_n)) - \mathbb{E}(g(X))\Big| \leq \varepsilon$, ce qui conclut la démonstration car ε est arbitraire. □

Exemple 6.24. Voici un contre-exemple à la réciproque. Soit $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes de loi de Bernoulli de paramètre $p \in]0,1[$ fixé. Alors $X_n \sim \mathrm{Bern}(p)$ pour tout $n \in \mathbb{N}^*$, donc on a clairement $X_n \xrightarrow{\mathcal{L}} \mathrm{Bern}(p)$. Par contre, $(X_n)_{n\geq 0}$ ne converge pas en probabilité car pour $n \neq m$ on a $d(X_n, X_m) = \mathbb{E}(|X_n - X_m| \wedge 1) = \mathbb{P}(X_n \neq X_m) = 2p(1-p)$ (on a $|X_n - X_m| = 1$ dès que $X_n \neq X_m$). Ainsi, $\sup_{n\geq m} d(X_n, X_m) = 2p(1-p)$ ne converge pas vers 0 et $(X_n)_{n\geq 1}$ n'est pas une suite de Cauchy pour la distance $d(\cdot,\cdot)$.

Une réciproque partielle est la suivante.

Proposition 6.25. *Si $X_n \xrightarrow{\mathcal{L}} X$ et si $\mathbb{P}(X = c) = 1$, alors $X_n \xrightarrow{\mathbb{P}} c$.*

Démonstration. Soit $\varepsilon > 0$. Prenons une fonction g_ε continue, égale à 1 en dehors de $[c - \varepsilon, c + \varepsilon]$ et égale à 0 en c. En notant que $\mathbb{1}_{\{|x-c| > \varepsilon\}} \leq g_\varepsilon(x)$, on a

$$\mathbb{P}(|X_n - c| > \varepsilon) \leq \mathbb{E}(g_\varepsilon(X_n)) \xrightarrow{n\to\infty} \mathbb{E}(g_\varepsilon(X)) = g_\varepsilon(c) = 0\,,$$

en ayant utilisé que g_ε est continue et bornée pour pouvoir appliquer la définition de la convergence en loi. □

6.4.3 Convergence p.s. implique convergence en probabilité, réciproque partielle

Proposition 6.26. *Si $X_n \xrightarrow{p.s.} X$, alors $X_n \xrightarrow{\mathbb{P}} X$.*

Démonstration. Soit $\varepsilon > 0$. On a $\mathbb{P}(|X_n - X| > \varepsilon) \leq \mathbb{P}(\sup_{k\geq n} |X_k - X| > \varepsilon)$, tout simplement car $\sup_{k\geq n} |X_k - X| \geq |X_n - X|$. Ainsi, si $X_n \xrightarrow{p.s.} X$, la majoration de $\mathbb{P}(|X_n - X| > \varepsilon)$ tend vers 0, ce qui donne la convergence en probabilité.

Autre démonstration : on pose $Y_n = \mathbf{1}_{\{|X_n - X| > \varepsilon\}}$. Alors p.s. on a $\lim_{n\to\infty} Y_n = 0$. Par convergence dominée, on a $\mathbb{E}(Y_n) = \mathbb{P}(|X_n - X| > \varepsilon) \to 0$ quand $n \to \infty$. □

Exemple 6.27. Voici un contre-exemple à la réciproque (Exercice 6.19). Soit $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes de loi $\mathrm{Bern}(p_n)$, avec $\lim_{n\to\infty} p_n = 0$ mais $\sum_{n=1}^{\infty} p_n = +\infty$. Alors on a facilement que $X_n \xrightarrow{\mathbb{P}} 0$, mais le Lemme de Borel–Cantelli (deuxième partie) assure que p.s. on a $X_n = 1$ infiniment souvent : cela montre que p.s. X_n ne converge pas vers 0.

Une réciproque partielle est la suivante.

Proposition 6.28. *Si $X_n \xrightarrow{\mathbb{P}} X$, alors il existe une sous-suite $(n_k)_{k\geq 1}$ telle que $X_{n_k} \xrightarrow{p.s.} X$.*

Démonstration. Pour tout $\varepsilon > 0$ on a $\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$. Ainsi, on peut choisir une sous-suite $(X_{n_k})_{n_k\geq 1}$ telle que $\mathbb{P}(|X_{n_k} - X| > 1/k) \leq 1/k^2$ (on fixe $\varepsilon = 1/k$ et on choisit n_k assez grand). Alors, on a $\sum_{k\geq 1} \mathbb{P}(|X_{n_k} - X| > 1/k) < +\infty$: par Borel–Cantelli, p.s. on n'a $|X_{n_k} - X| > 1/k$ qu'un nombre fini de fois. Autrement dit, p.s. $|X_{n_k} - X| \leq 1/k$ à partir d'un certain rang k_0, ce qui montre que $X_{n_k} \xrightarrow{p.s.} X$. □

6.4.4 Convergence dans L^p et convergence presque sûre : contre-exemples

Exemple 6.29. Voici un exemple où $X_n \xrightarrow{L^p} X$ mais $(X_n)_{n\geq 1}$ ne converge pas p.s.
On prend $(X_n)_{n\geq 1}$ une suite de v.a. indépendantes, de loi $\mathrm{Bern}(p_n)$, avec $p_n \to 0$ mais $\sum_{n=1}^{\infty} p_n = +\infty$. On voit facilement que $\mathbb{E}(|X_n|^p) = p_n \to 0$, mais le Lemme de Borel–Cantelli (deuxième partie) montre que p.s. $X_n = 1$ infiniment souvent (donc en particulier ne converge pas p.s. vers 0).

Exemple 6.30. Voici un exemple où $X_n \xrightarrow{p.s.} X$ mais $(X_n)_{n\geq 1}$ ne converge pas dans L^p.
On prend $(X_n)_{n\geq 1}$ une suite de v.a. de loi $\mathbb{P}(X_n = e^n) = p_n$ et $\mathbb{P}(X_n = 0) = 1 - p_n$, avec $\sum_{n=1}^{\infty} p_n < +\infty$. Par le Lemme de Borel–Cantelli on obtient que p.s. $X_n = 0$ à partir d'un certain rang, donc $X_n \xrightarrow{p.s.} 0$. D'autre part, $\mathbb{E}(|X_n|^p) = p_n e^{pn}$, donc $(X_n)_{n\geq 1}$ ne converge pas dans L^p dès que $p_n e^{pn}$ ne tend pas vers 0.

6.5 La loi (faible et forte) des grands nombres

Énonçons maintenant un autre théorème fondamental en probabilités : la loi faible et la loi forte des grands nombres. Le premier se déduit du second, mais sa démonstration est plus facile.

Théorème 6.31 (Loi faible des grands nombres)**.** *Soit $(X_k)_{k\geq 1}$ une suite de v.a. i.i.d., admettant une espérance finie. On pose $\overline{X}_n := \frac{1}{n}\sum_{k=1}^{n} X_k$, appelée* moyenne empirique. *Alors $(\overline{X}_n)_{n\geq 1}$ converge en probabilité vers $\mu = \mathbb{E}(X_1)$.*

Démonstration. En considérant $X_k - \mu$, on se ramène facilement au cas de v.a. centrées. Supposons donc que $\mu = 0$; par linéarité on a notamment $\mathbb{E}(\overline{X}_n) = 0$.

La démonstration est très facile dans le cas où $\sigma^2 = \mathrm{Var}(X_k^2) = \mathbb{E}(X_k^2) < +\infty$. Dans ce cas, par indépendance des X_k on a $\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}\sum_{k=1}^{n} \mathrm{Var}(X_k) = \frac{1}{n}\sigma^2$. Alors par Bienaymé–Tchebychev, on obtient, pour tout $\varepsilon > 0$

$$\mathbb{P}(|\overline{X}_n| > \varepsilon) \leq \frac{1}{\varepsilon^2}\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n\varepsilon^2},$$

qui tend vers 0 quand $n \to \infty$.

Le cas général s'obtient par approximation. Soit $\delta > 0$ arbitraire et soit $A = A(\delta)$ suffisamment grand de sorte que

$$\left|\mathbb{E}(X_1 \mathbb{1}_{\{|X_1| \leq A\}})\right| \leq \delta, \qquad \mathbb{E}(|X_1| \mathbb{1}_{\{|X_1| > A\}}) \leq \delta.$$

On pose $W_k := X_k \mathbb{1}_{\{|X_k| \le A\}}$ et $Z_k := X_k \mathbb{1}_{\{|X_k|>A\}}$ de sorte que $X_k = W_k + Z_k$. Alors, pour tout $\varepsilon > 0$, on a

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^n X_k\Big| > \varepsilon\Big) \le \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^n W_k\Big| > \varepsilon/2\Big) + \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^n Z_k\Big| > \varepsilon/2\Big)$$

car $|w+z| > \varepsilon$ implique que $|w| > \varepsilon/2$ ou $|z| > \varepsilon/2$. On traite les deux termes séparément.

Pour le premier terme, en passant au carré et en appliquant l'inégalité de Markov,

$$\begin{aligned}\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^n W_k\Big| > \varepsilon/2\Big) &\le \frac{4}{\varepsilon^2}\frac{1}{n^2}\mathbb{E}\Big(\Big(\sum_{k=1}^n W_k\Big)^2\Big)\\ &\le \frac{4}{n^2\varepsilon^2}\Big(\mathrm{Var}\Big(\sum_{k=1}^n W_k\Big) + \mathbb{E}\Big(\sum_{k=1}^n W_k\Big)^2\Big) \le \frac{4\mathrm{Var}(W_1)}{n\varepsilon^2} + \frac{4\mathbb{E}(W_1)^2}{\varepsilon^2}\,.\end{aligned}$$

Notons que, par définition de W_1, on a $\mathrm{Var}(W_1) \le \mathbb{E}(W_1^2) \le A^2$ et $\mathbb{E}(W_1) \le \delta$.

Pour le deuxième terme, on applique directement l'inégalité de Markov, pour obtenir

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^n Z_k\Big| > \varepsilon/2\Big) \le \frac{2}{n\varepsilon}\mathbb{E}\Big(\Big|\frac{1}{n}\sum_{k=1}^n Z_k\Big|\Big) \le \frac{2}{\varepsilon}\mathbb{E}(|Z_1|)\,.$$

Par définition de Z_1, on a $\mathbb{E}(|Z_1|) \le \delta$. Ainsi, on en conclut que

$$\limsup_{n\to\infty} \mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^n W_k\Big| > \varepsilon/2\Big) \le \frac{4\delta^2}{\varepsilon^2} + \frac{2\delta}{\varepsilon}\,,$$

ce qui conclut la démonstration car δ est arbitraire. □

Théorème 6.32 (Loi forte des grands nombres)**.** *Soit $(X_k)_{k\ge1}$ une suite de v.a. i.i.d., admettant une espérance finie. On pose $\overline{X}_n := \frac{1}{n}\sum_{k=1}^n X_k$, appelée* moyenne empirique*. Alors $(\overline{X}_n)_{n\ge1}$ converge p.s. et dans L^1 vers $\mu = \mathbb{E}(X_1)$.*

Il existe de nombreuses démonstrations différentes. La démonstration complète n'est pas facile. L'hypothèse d'indépendance peut être assouplie : il suffit d'avoir des v.a. de même loi et deux à deux indépendantes (théorème d'Etemadi, [GK19, p. 277]).

Avec des moments d'ordre 4. Ref. [App15, p. 434]. Donnons ici une démonstration jolie lorsque les X_k admettent un moment d'ordre 4 fini ; comme précédemment, on peut supposer que $\mu = 0$. Soit $\varepsilon > 0$. En prenant la puissance 4 et en appliquant l'inégalité de Markov, on a

$$\mathbb{P}\Big(\frac{1}{n}\Big|\sum_{k=1}^n X_k\Big| > \varepsilon\Big) \le \frac{1}{\varepsilon^4}\frac{1}{n^4}\mathbb{E}\Big(\Big(\sum_{k=1}^n X_k\Big)^4\Big) = \frac{1}{n^4\varepsilon^4}\sum_{i,j,k,l=1}^n \mathbb{E}(X_iX_jX_kX_l)$$

où pour la dernière identité on a développé la puissance 4 de la somme et utilisé la linéarité de l'espérance. Notons maintenant que par indépendance, $\mathbb{E}(X_iX_jX_kX_l) = 0$ dès que l'un des indices est différent des autres. Si $i = j \ne k = l$ (ou $i = k \ne j = l$ ou $i = l \ne k = l$) on a $\mathbb{E}(X_iX_jX_kX_l) = \mathbb{E}(X_1^2)^2$; notons qu'il y a $3n(n-1)$ telles configurations d'indices. Si $i = j = k = l$, on a $\mathbb{E}(X_iX_jX_kX_l) = \mathbb{E}(X_1^4)$; il y a n telles configurations d'indices. Au final, on a

$$\mathbb{P}\Big(\frac{1}{n}\Big|\sum_{k=1}^n X_k\Big| > \varepsilon\Big) \le \frac{1}{n^4\varepsilon^4}\big(3n(n-1)\mathbb{E}(X_1^2)^2 + n\mathbb{E}(X_1^4)\big)\,.$$

Cette borne supérieure étant sommable en n (pour tout $\varepsilon > 0$ fixé), on obtient que $\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{p.s.} 0$, voir la Remarque 6.18. □

Avec des moments d'ordre 2. Ref : [GK19, p.275]. Donnons ici une démonstration astucieuse lorsque les X_k admettent un moment d'ordre 2 fini (on note $\sigma^2 = \mathrm{Var}(X_1^2)$) ; comme précédemment, on peut supposer que $\mu = 0$. Bienaymé–Tchebichev nous donne l'inégalité :

$$\mathbb{P}\Big(\frac{1}{n}\Big|\sum_{k=1}^n X_k\Big| > \varepsilon\Big) \le \frac{\sigma^2}{n\varepsilon}.$$

Si on applique cette inégalité uniquement à n^2, on obtient une borne supérieure sommable, donc par Borel–Cantelli $\frac{1}{n^2}\sum_{k=1}^{n^2} X_k \xrightarrow{p.s.} 0$. Autrement dit, on a extrait une sous-suite qui converge p.s.

Il reste à « combler les trous » entre n^2 et $(n+1)^2$. Pour $m \geq 1$, on note $n = n_m$ l'unique entier tel que $n^2 \leq m < (n+1)^2$. Notons que $m/n_m^2 \to 1$ quand $m \to \infty$, donc il nous suffit de montrer que $\frac{1}{n_m^2}\sum_{k=1}^{m} X_k \xrightarrow{p.s.} 0$. On écrit

$$\frac{1}{n_m^2}\sum_{k=1}^{m} X_k = \frac{1}{n_m^2}\sum_{k=1}^{n_m^2} X_k + \frac{1}{n_m^2}\sum_{k=n_m^2+1}^{m} X_k \,.$$

D'après ce qui précède, le premier terme converge p.s. vers 0 quand $m \to \infty$, et il reste à traiter le deuxième. Pour $\varepsilon > 0$, on a grâce à Bienaymé–Tchebychev et par indépendance des X_k

$$\mathbb{P}\Big(\Big|\frac{1}{n_m^2}\sum_{k=n_m^2+1}^{m} X_k\Big| > \varepsilon\Big) \leq \frac{1}{\varepsilon^2}\mathrm{Var}\Big(\frac{1}{n_m^2}\sum_{k=n_m^2+1}^{m} X_k\Big) = \frac{1}{\varepsilon^2}\frac{1}{n_m^4}\sum_{k=n_m^2+1}^{m}\mathrm{Var}(X_k) = \frac{(m-n_m^2)\sigma^2}{\varepsilon^2 n_m^4} \leq \frac{2\sigma^2}{n_m^3}\,,$$

car $m - n_m^2 \leq (n_m+1)^2 - n_m^2 - 1 \leq 2n_m$. Comme $n_m \sim \sqrt{m}$ quand $m \to \infty$ la borne supérieure est sommable (en m) : on en déduit par Borel–Cantelli que $\frac{1}{n_m^2}\sum_{k=n_m^2+1}^{m} X_k \xrightarrow{p.s.} 0$, ce qui conclut la démonstration. □

6.5.1 Une application de la loi des grands nombres : l'urne de Pólya

Aucune référence. Le contexte de l'urne de Pólya est le suivant. Une urne contient initialement a boules rouges et b boules vertes. Lors du n-ème tour, on tire une boule dans l'urne : on la replace dans l'urne en ajoutant aussi une boule de la même couleur. Autrement dit, si après le $(n-1)$-ème tour l'urne contient r boules rouges et v boules vertes, alors au n-ème tour : avec probabilité $r/(r+v)$ on tirera une boule rouge et on ajoutera une boule rouge dans l'urne (qui contiendra $r+1$ boules rouges et v boules vertes après le n-ème tour) ; avec probabilité $v/(r+v)$ on tirera une boule verte et on ajoutera une boule verte dans l'urne (qui contiendra r boules rouges et $v+1$ boules vertes après le n-ème tour). Ainsi, la composition de l'urne évolue au fil des tours et influence la suite de l'évolution (on parle de processus renforcé). On note R_n le nombre de boules rouges dans l'urne après n tirages ; notons qu'après n tirages il y aura toujours $a+b+n$ boules au total dans l'urne. On s'intéresse à l'évolution de $W_n := \frac{R_n}{a+b+n}$, la proportion de boules rouges dans l'urne.

On peut en fait étudier ce processus très bien, après une première série de manipulations. Soit $(X_i)_{i\geq 1}$ la suite de v.a. de Bernoulli correspondant à la couleur de la boule tirée : on pose $X_i = 1$ si la boule tirée lors du i-ème tour est rouge et $X_i = 0$ sinon ; ainsi $R_n = a + \sum_{i=1}^{n} X_i$. Les $(X_i)_{i\geq 1}$ ne sont pas indépendants mais l'observation essentielle est que l'on peut déterminer la loi de $(X_1,\ldots,X_n)$, par récurrence. Soit $(\varepsilon_1,\ldots,\varepsilon_n) \in \{0,1\}^n$: on a

$$\mathbb{P}\big(X_n = \varepsilon_n \mid (X_1,\ldots,X_{n-1}) = (\varepsilon_1,\ldots,\varepsilon_{n-1})\big) = \begin{cases} \dfrac{a+\sum_{i=1}^{n-1}\varepsilon_i}{a+b+n-1} & \text{si } \varepsilon_n = 1\,,\\[2ex] \dfrac{b+n-1-\sum_{i=1}^{n-1}\varepsilon_i}{a+b+n-1} & \text{si } \varepsilon_n = 0\,.\end{cases}$$

Ainsi, on montre assez facilement par récurrence que, en posant $k = k_{\varepsilon,n} = \sum_{i=1}^{n}\varepsilon_i$, on a

$$\mathbb{P}\big((X_1,\ldots,X_n) = (\varepsilon_1,\ldots,\varepsilon_n)\big) = \frac{\prod_{i=0}^{k-1}(a+i)\prod_{i=0}^{n-k-1}(b+i)}{\prod_{i=0}^{n-1}(a+b+i)}\,.$$

Un fait remarquable est que cette probabilité ne dépend que de $k = k_{\varepsilon,n}$ et pas de la suite précise des ε_i ! Si on continue la manipulation, on peut écrire

$$\mathbb{P}\big((X_1,\ldots,X_n) = (\varepsilon_1,\ldots,\varepsilon_n)\big) = \frac{(a+k-1)!}{(a-1)!}\frac{(b+n-k-1)!}{(b-1)!}\frac{(a+b-1)!}{(a+n-1)!} = \frac{B(a+k,b+n-k)}{B(a,b)}\,,$$

où $B(u,v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)} = \int_0^1 x^{u-1}(1-x)^{v-1}dx$ est la fonction beta. Au final, on a donc obtenu

$$\begin{aligned}\mathbb{P}\big((X_1,\ldots,X_n) = (\varepsilon_1,\ldots,\varepsilon_n)\big) &= \frac{1}{B(a,b)}\int_0^1 x^k(1-x)^{n-k}x^{a-1}(1-x)^{b-1}dx \\ &= \int_0^1 \mathbb{P}_x\big((\tilde{X}_1,\ldots,\tilde{X}_n) = (\varepsilon_1,\ldots,\varepsilon_n)\big)f_{a,b}(x)dx\,,\end{aligned} \tag{6.1}$$

où $f_{a,b}$ est la densité d'une loi beta de paramètre a, b, et les $(\tilde{X}_i)$ sont des v.a. *indépendantes* de loi $\mathrm{Bern}(x)$ (sous $\mathbb{P}_x$).

Conclusion 1. Avant de rentrer dans l'interprétation de cette expression, notons que l'on peut en déduire la convergence en loi de la proportion W_n de boules rouges dans l'urne. Soit $t \in [0,1]$. Alors

$$\mathbb{P}(W_n \leq t) = \mathbb{P}(R_n \leq (a+b+n)t) = \mathbb{P}\Big(\sum_{i=1}^{n} X_i \leq (a+b+n)t - a\Big)$$

et par Fubini

$$\mathbb{P}(W_n \leq t) = \int_0^1 \mathbb{P}_x\Big(\sum_{i=1}^{n} X_i \leq (a+b+n)t - a\Big) f_{a,b}(x)dx\,.$$

Maintenant, d'après la loi (faible) des grands nombres, comme les $\tilde{X}_i$ sont i.i.d. $\mathrm{Bern}(x)$ sous $\mathbb{P}_x$, on a

$$\lim_{n\to\infty} \mathbb{P}_x\Big(\sum_{i=1}^{n} X_i \leq (a+b+n)t - a\Big) = \lim_{n\to\infty} \mathbb{P}_x\Big(\frac{1}{n}\sum_{i=1}^{n} X_i \leq t + \tfrac{bt+a(t-1)}{n}\Big) = \begin{cases} 0 & \text{si } t < x\,, \\ 1 & \text{si } t > x\,. \end{cases}$$

(Le cas $t = x$ n'est pas utile.) Ainsi, par convergence dominée (en majorant la probabilité par 1), on obtient

$$\lim_{n\to\infty} \mathbb{P}(W_n \leq t) = \int_0^1 \mathbb{1}_{\{x<t\}} f_{a,b}(x)dx = \int_0^t f_{a,b}(x)dx\,.$$

Cela montre que W_n converge en loi vers une v.a. Z de loi $\mathrm{Beta}(a,b)$.

Conclusion 2 (difficile). Considérons le vecteur $(Z, \tilde{X}_1, \dots, \tilde{X}_n)$ tel que Z est une v.a. aléatoire de loi $\mathrm{Beta}(a,b)$ et $(\tilde{X}_1, \dots, \tilde{X}_n)$ sont des v.a. indépépendantes de loi $\mathrm{Ber}(Z)$ (une fois la valeur de Z fixée) : une manière de l'écrire est de prendre $(U_i)_{i\geq 1}$ des v.a. i.i.d. de loi $\mathcal{U}([0,1])$ et de considérer $(Z, \mathbb{1}_{\{U_1 \leq Z\}}, \dots, \mathbb{1}_{\{U_n \leq Z\}})$. Alors le membre de droite de (6.1) est exactement la loi marginale de $(\tilde{X}_1, \dots, \tilde{X}_n)$ et donc (6.1) montre que $(X_1, \dots, X_n)$ possède la même loi que $(\tilde{X}_1, \dots, \tilde{X}_n)$. Autrement dit, si on tire une v.a. $Z \sim \mathrm{Beta}(a,b)$ et qu'ensuite on tire des v.a. $(\tilde{X}_i)_{i\geq 1}$ i.i.d. de loi $\mathrm{Bern}(Z)$, alors la suite $(\tilde{X}_i)_{i\geq 1}$ a la même loi que la suite $(X_i)_{i\geq 1}$ qui encode la suite des boules tirées dans le schéma des urnes de Pólya[2]. Notamment, les propriétés presque sûres des deux suites $(X_i)_{i\geq 1}$ et $(\tilde{X}_i)_{i\geq 1}$ sont les mêmes.

Une conséquence directe de la loi forte des grands nombres est que $\frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i \xrightarrow{p.s.} Z$ (la limite dépend de la valeur Z du paramètre commun des Bernoulli indépendantes). On en déduit que $W = \lim_{n\to\infty} W_n = \lim_{n\to\infty} \frac{1}{a+b+n}(a + \sum_{i=1}^{n} X_i)$ existe p.s., et que $W \sim \mathrm{Beta}(a,b)$. En conclusion, la proportion W_n de boules rouges converge p.s. vers une v.a. W de loi $\mathrm{Beta}(a,b)$.

2. Il s'agit en réalité d'un théorème profond de De Finetti : si des v.a. $(X_i)_{i\geq 1}$ sont *échangeables*, dans le sens où pour tout n les vecteurs $(X_1, \dots, X_n)$ et $(X_{\sigma(1)}, \dots, X_{\sigma(n)})$ ont la même loi pour toute permutation σ, alors la suite $(X_i)_{i\geq 1}$ peut s'écrire comme un *mélange de v.a. i.i.d.*

Chapitre VI

Analyse numérique

Dans ce chapitre, nous faisons référence à la section 13 du programme de tronc commun. Elle constitue une incursion modeste dans le monde de l'analyse numérique, et elle concerne tous les étudiants, et non seulement ceux choisissant l'option B (Calcul scientifique), qui iront plus loin dans cette direction. Dans les titres des leçons **226** du programme d'analyse et **149** de la liste d'algèbre, il est explicitement fait référence à la notion de "résolution approchée", il est donc attendu du jury de connaître quelques éléments sur ces questions.

Il est également naturel d'évoquer ces notions dans de nombreuses leçons, comme par exemple **209-219-220-221-222-223-236-253-262** (je ne prétends pas être exhaustif).

Pour les références, citons les classiques [Dem16, Sch04] et les polycopiés [Boya, Boyb, Gon, Fro]. On pourra aussi trouver dans [All12, Chapitres 9-10] les algorithmes d'optimisation cités ici.

1 Interpolation polynomiale

La première direction que nous allons explorer, est d'essayer d'*approcher* des fonctions par des "objets" plus simples. Considérons par exemple qu'on veut approcher une fonction par des polynômes. Une façon "naïve" de demander que ce polynôme soit proche de la fonction considérée, est de demander à ce que ce polynôme soit égal à la fonction initiale en un certain nombre de points.

Théorème 1.1. *Soit $a < b$, $f : [a,b] \to \mathbb{R}$, $n \in \mathbb{N}$, et $n+1$ points distincts de $[a,b]$ notés $(x_0, \dots, x_n) \in [a,b]$. Alors il existe un unique polynome $P \in \mathbb{R}_n[x]$ tel que*

$$\forall i \in [\![0,n]\!], \quad P(x_i) = f(x_i).$$

On appelle P le polynôme d'interpolation de Lagrange de f en les points $(x_i)_{i\in[\![0,n]\!]}$.

Remarque 1.2. La référence à la fonction f n'est pas très utile pour le moment : avec $y_i = f(x_i)$ pour $i \in [\![0,n]\!]$, on cherche P un polynome de degré au plus n tel que $P(x_i) = y_i$ pour tout $i \in [\![0,n]\!]$.

Remarque 1.3. Afin de ne pas se tromper entre le nombre de points et le degré attendu du polynôme, remarquons qu'en cherchant un polynome de degré inférieur à n, on a $n+1$ inconnues, à savoir les coefficients du polynôme. Il est donc naturel (même si ça ne constitue pas une preuve, voir les détails ci-dessous) d'avoir $n+1$ équations, i.e. $n+1$ point (x_i).

Démonstration. **Preuve 1 :** On cherche $P = \sum_{k=0}^n a_k x^k$ où $(a_k)_{k\in[\![0,n]\!]}$ sont les inconnues. Le système d'équations se réécrit :

$$\left\{\begin{array}{ccccccccc} a_0 & + & a_1 x_0 & + & \dots & + & a_n x_0^n & = & f(x_0) \\ a_0 & + & a_1 x_1 & + & \dots & + & a_n x_1^n & = & f(x_1) \\ & & & & \vdots & & & & \\ a_0 & + & a_1 x_n & + & \dots & + & a_n x_n^n & = & f(x_n) \end{array}\right.$$

qui est un système linéaire à n inconnues. Ce système se réécrit

$$AX = B, \quad \text{où} \quad A = \begin{pmatrix} 1 & x_0 & \dots & x_0^n \\ 1 & x_1 & \dots & x_1^n \\ & & \vdots & \\ 1 & x_n & \dots & x_n^n \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix}$$

La matrice A est dite de Vandermonde, et elle est classiquement inversible si les x_i sont deux à deux disjoints [1].
Preuve 2 : On pose pour $i_0 \in [\![0, n]\!]$

$$L_{i_0}(x) = \prod_{j \neq i_0} \frac{x - x_j}{x_{i_0} - x_j} \in \mathbb{R}_n[x].$$

Alors le polynôme $P(x) = \sum_{i=0}^n f(x_i) L_i(x)$ est bien un polynôme de degré inférieur à n, et il convient car $L_i(x_j) = \delta_{i,j}$ pour tout $(i, j) \in [\![0, n]\!]$. Ceci montre l'existence.

Pour l'unicité, on constate que si P_1 et P_2 conviennent, alors $P_1 - P_2$ est un polynôme de degré inférieur à n et ayant $n + 1$ racines, il est donc nul, c'est-à-dire que $P_1 = P_2$.
Preuve 3 : On considère l'application [2]

$$\Phi_n : P \in \mathbb{R}_n[X] \longmapsto \begin{pmatrix} P(x_0) \\ P(x_1) \\ \vdots \\ P(x_n) \end{pmatrix} \in \mathbb{R}^{n+1}.$$

On doit montrer qu'elle est bijective ; mais elle est clairement linéaire entre deux espaces vectoriels de même dimension, il suffit donc de montrer qu'elle est injective. Or si $\Phi_n(P) = 0$, alors P est un polynôme de degré inférieur à n ayant $n + 1$ racines, il est donc nécessairement nul. □

Remarque 1.4. La seconde preuve ci-dessus donne en prime la formule

$$P = \sum_{i=0}^{n} f(x_i) L_i,$$

qui est l'unique polynôme de degré inférieur à n qui passe par les points $(x_i, f(x_i))_{i \in [\![0,n]\!]}$.

Définition-Proposition 1.5. Etant donnés $(x_0, \dots, x_n)$ des réels deux à deux distincts, la famille $(L_i)_{i \in [\![0,n]\!]}$ définie par

$$\forall i \in [\![0, n]\!], \quad L_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}$$

est une base de $\mathbb{R}_n[x]$, appelée base de Lagrange associée aux points $(x_0, \dots, x_n)$.

Démonstration. Comme il s'agit d'une famille à $(n + 1)$ éléments de $\mathbb{R}_n[x]$ qui est de dimension $n + 1$, il suffit de montrer que la famille est libre, ou qu'elle est génératrice.
Preuve par le fait que la famille est libre : Soit donc $(\alpha_i)_{i \in [\![0,n]\!]}$ tel que $\sum_{i=0}^n \alpha_i L_i = 0$. Soit $i_0 \in [\![0, n]\!]$. La valeur en x_{i_0} donne $\alpha_{i_0} = 0$, d'où le résultat.
Preuve par le fait que la famille est génératrice : Il suffit d'appliquer les idées de la preuve 2. qu'on a donné ci-dessus du Théorème 1.1 : en effet, en appliquant la remarque 1.4 à la fonction $f = P$, on a

$$P = \sum_{i=0}^{n} P(x_i) L_i$$

donc la famille est bien génératrice. □

1. On peut par exemple montrer que le déterminant de cette matrice est $\prod_{i<j} (x_i - x_j)$
2. La matrice de l'application Φ_n dans la base $(1, X, \cdots, X^n)$ de $\mathbb{R}_n[X]$ n'est autre que la matrice Vandermonde A de la preuve 1.

Remarque 1.6. En fait $(L_i)_{i\in[\![0,n]\!]}$ peut être interprêté comme la base de $\mathbb{R}_n[x]$ dont la base duale est la famille des masses de Dirac en x_i, c'est-à-dire

$$\delta_{x_i} : P \in \mathbb{R}_n[x] \longmapsto P(x_i).$$

En effet on a [3]

$$\forall (i,j) \in [\![0,n]\!], \;\; \delta_{x_i}(L_j) = \delta_{i,j}$$

Remarque 1.7. La base de Lagrange n'est pas toujours adaptée au calcul numérique : par exemple, si on a calculé $P = \sum f(x_i)L_i$, mais qu'on décide finalement d'ajouter un point à la liste $(x_i)_{i\in[\![0,n]\!]}$ (on cherche alors un polynôme de degré $n+1$), tous les termes de la base vont changer, ainsi que les coefficients du nouveau polynôme. Pour remédier à ce problème, on peut préférer la "forme de Newton du polynôme de Lagrange", qui consiste à calculer le polynôme d'interpolation P dans la base de Newton des points (x_i), à savoir

$$N_i(x) = \prod_{0\leq j<i} (x - x_j).$$

On cherche maintenant à étudier à quel point l'interpolation polynomiale donnée par la méthode de Lagrange fournit une approximation de la fonction f initiale :

Théorème 1.8. *Soit $n \in \mathbb{N}$, et $n+1$ points distincts de $[a,b]$ qu'on note $A = (x_0,\ldots,x_n)$. On suppose $f : [a,b] \to \mathbb{R}$ de classe C^{n+1} et on note P_A le polynôme d'interpolation de Lagrange de f en les points $(x_0,\ldots,x_n)$. Alors pour tout $x \in [a,b]$, il existe $\xi_x \in]a,b[$ tel que*

$$f(x) - P_A(x) = \frac{\prod_{j=0}^{n}(x-x_j)}{(n+1)!} f^{(n+1)}(\xi_x).$$

Démonstration. **Etape 1 : Théorème de Rolle généralisé :** Soit $a = x_1 < x_2 < \ldots < x_k = b$ une subdivision de $[a,b]$ en $k \geq 2$ points distincts, et $(\alpha_i)_{1\leq i\leq k} \in (\mathbb{N}^*)^k$. On pose $n = \left(\sum_{i=1}^{k} \alpha_i\right) - 1$ et on suppose $f : [a,b] \to \mathbb{R}$ de classe C^n. Si f s'annule à l'ordre α_i en x_i pour tout $1 \leq i \leq k$, alors il existe $\xi \in]a,b[$ tel que

$$f^{(n)}(\xi) = 0.$$

On va montrer ce résultat par récurrence sur $n \in \mathbb{N}^*$, plus précisément on pose pour $n \in \mathbb{N}^*$:

$H(n)$: pour tout $k \geq 2$ et $(\alpha_i)_{1\leq i\leq k} \in (\mathbb{N}^*)^k$ tel que $n = \left(\sum_{i=1}^{k} \alpha_i\right) - 1$, pour toute subdivision $a = x_1 < x_2 < \ldots < x_k = b$ et toute fonction $f : [a,b] \to \mathbb{R}$ de classe C^n, si f s'annule à l'ordre α_i en x_i pour tout $1 \leq i \leq k$, alors il existe $\xi \in]a,b[$ tel que

$$f^{(n)}(\xi) = 0.$$

$H(1)$: On a nécessairement $k = 2$, $x_1 = a$, $x_2 = b$ et $\alpha_1 = \alpha_2 = 1$. Le théorème de Rolle donne donc le résultat.

Soit $n \in \mathbb{N}^*$ tel que $H(n)$. Soit $k \geq 2$ et $(\alpha_i)_{1\leq i\leq k} \in (\mathbb{N}^*)^k$ tel que $n+1 = \left(\sum_{i=1}^{k} \alpha_i\right) - 1$, soit $a = x_1 < x_2 < \ldots < x_k = b$ et $f : [a,b] \to \mathbb{R}$ de classe C^{n+1}. Pour tout $i \in [\![1,k-1]\!]$, on applique le théorème de Rolle à f sur les intervalles $[x_i, x_{i+1}]$, et donc il existe $y_i \in]x_i, x_{i+1}[$ tel que $f'(y_i) = 0$. Ainsi f' s'annule [4] en les x_i à l'ordre $(k-1)$ et en les y_i à l'ordre 1, ce qui donne un ordre total de

$$\sum_{i=1}^{k}(\alpha_i - 1) + (k-1) = n + 1 - k + (k-1) = n$$

donc on peut appliquer l'hypothèse de récurrence à f sur cette nouvelle subdivision, et ainsi il existe $\xi \in]a,b[$ tel que $(f')^{(n)}(\xi) = 0$, d'où le résultat attendu sur f, ce qui montre $H(n+1)$ et achève la récurrence.

3. Ne pas confondre la notation masse de Dirac (qui prend un seul indice) et le symbole de Kronecker (qui prend deux indices).

4. Si $\alpha_{i_0} = 1$ on dit que f' s'annule à l'ordre 0 en x_{i_0}, ce qui revient à ne rien dire, on a une information seulement si $\alpha_{i_0} \geq 2$.

Etape 2 : Soit n, $A = (x_0, \ldots, x_n)$, f et P_A comme dans le théorème 1.8. Soit également $x \in [a, b]$. Si x est l'un des x_i, toutes les valeurs de $\xi \in]a, b[$ conviennent, on prend par exemple $\xi = (a + b)/2$. On suppose donc que x n'est pas l'un des x_i, et on pose

$$\forall t \in [a, b], \quad \Psi(t) = P_A(t) - f(t) + M \prod_{j=0}^{n} (t - x_j)$$

où M est un réel. On choisit M de sorte que $\Psi(x) = 0$, c'est-à-dire

$$M = \frac{f(x) - P_A(x)}{\prod_{j=0}^{n}(x - x_j)}.$$

La fonction Ψ s'annule en $n+2$ points disctincts, les x_i et x. L'étape 1 appliquée à Ψ donne l'existence de $\xi \in]a, b[$ tel que $\Psi^{(n+1)}(\xi) = 0$. Or $\deg(P_A) \leq n$, donc

$$\Psi^{(n+1)}(\xi) = -f^{(n+1)}(\xi) + M(n+1)!$$

d'où le résultat. □

Remarque 1.9. On a trouvé utile dans l'étape 1 ci-dessous de mettre un énoncé "fin" d'une possible généralisation du théorème de Rolle. Mais dans le cas auquel on l'applique ici, à savoir que la fonction $\Psi : [a, b] \to \mathbb{R}$ a $n+2$ points distincts d'annulation, il est facile de se convaincre rapidement du résultat par le raisonnement suivant :

— par le théorème de Rolle appliqué aux $n+1$ intervalles formés par les $n+2$ zéros de Ψ, on obtient que Ψ' s'annule en n points distincts.

— On applique ensuite le théorème de Rolle à Ψ' sur les $n-1$ intervalles formés par les zéros de Ψ' obtenus précédemment, et on itère le processus, pour finir avec le fait que $\Psi^{(n+1)}$ s'annule en au moins un point de $[a, b]$.

Mais attention à ce type de pseudo-récurrence dont il ne faut pas user dans le cadre de l'agrégation. Dans un écrit d'agrégation, il faut proscrire ce type de rédaction, au moins au début d'un sujet. Dans un oral, on peut se permettre un tel argument rapide pour répondre à une question du jury, mais dans le cadre d'un développement il vaut mieux éviter de tels raccourcis peu rigoureux.

Remarque 1.10. Voir aussi l'exercice 3.64 du Chapitre I pour un autre cas du théorème de Rolle généralisé et pour une application à la preuve du théorème de Taylor-Lagrange.

On a donc l'estimation d'erreur suivante :

Corollaire 1.11. *Avec les notations du Théorème 1.8, on a*

$$\|f - P_A\|_\infty \leq \frac{\|f^{(n+1)}\|_{\infty,[a,b]}}{(n+1)!} \|w_A\|_\infty \qquad \text{où } w_A(x) := \prod_{j=0}^{n} (x - x_j)$$

Considérons que pour tout $n \in \mathbb{N}$ on choisit $n+1$ points d'interpolation dans $[a, b]$ et on note P_n le polynome d'interpolation de Lagrange en ces points. On se demande alors si P_n converge vers f quand n tend vers $+\infty$. Donnons quelques exemples dans lesquels on peut répondre à cette question :

Exemple 1.12. Si on suppose qu'il existe $M > 0$ tel que pour tout $k \in \mathbb{N}$, $\|f^{(k)}\|_{\infty,[a,b]} \leq M$, alors

$$P_n \xrightarrow[n \to \infty]{} f \quad \text{uniformément sur } [a, b].$$

En effet on a

$$\|f - P_n\|_\infty \leq \frac{\|f^{(n+1)}\|_{\infty,[a,b]}}{(n+1)!} (b-a)^{n+1} \leq \frac{M(b-a)^{n+1}}{(n+1)!}.$$

Exemple 1.13. Si on suppose que f est développable en série entière autour de $c = \frac{a+b}{2}$ et de rayon de convergence $R > \frac{3\ell}{2}$ où $\ell = b - a$ est la longueur de $[a, b]$, alors on a

$$P_n \xrightarrow[n\to\infty]{} f \quad \text{uniformément sur } [a,b].$$

En effet, on peut écrire

$$\forall x \in]c - R, c + R[, \quad f(x) = \sum_{n\in\mathbb{N}} a_n(x-c)^n$$

où $(a_n) \in \mathbb{R}^{\mathbb{N}}$. Cela permet d'étendre (par la même série entière) f en une fonction holomorphe sur $D(c, R) \subset \mathbb{C}$. Les inégalités de Cauchy (voir la proposition 3.35 au Chapitre II) donnent

$$\forall x \in [a,b], \forall n \in \mathbb{N}, \quad |f^{(n)}(x)| \leq \frac{n!}{r^n} \sup_{C(x,r)} |f|$$

où $C(x, r)$ est le cercle de centre x et de rayon r tel que ce cercle est inclu dans $D(c, R)$. On en déduit :

$$\|f - P_n\|_\infty \leq \frac{(b-a)^{n+1}}{(n+1)!}\|f^{(n+1)}\|_{\infty,[a,b]} \leq \frac{(b-a)^{n+1}}{(n+1)!}\frac{(n+1)!}{r^{n+1}} \sup_{\overline{D}(c,\ell/2+r)} |f| = \left(\frac{b-a}{r}\right)^{n+1} \|f\|_{\infty,\overline{D}(c,\ell/2+r)}$$

à condition que tous les cercles $C(x, r)$ avec $x \in [a, b]$ soient dans $D(c, R)$ (on constate en effet que tous ces cercles sont dans le disque $\overline{D}(c, \ell/2 + r)$). Ceci est valide si $\ell/2 + r < R$. Or $R > 3\ell/2$ donc on peut choisir r de sorte que $\ell < r < R - \ell/2$, et pour un tel choix de r, l'erreur estimée ci-dessus tend bien vers 0 quand n tend vers $+\infty$.

Remarque 1.14. Dans les exemples précédents, les convergences ont été établies sous des hypothèses fortes, mais elles sont valables quels que soient les points d'interpolation : en effet on a seulement utilisé l'estimation "brutale"

$$\forall x \in [a,b], \quad \left|\prod_{j=0}^{n}(x - x_i)\right| \leq (b-a)^{n+1}.$$

1. Considérons par exemple le cas de points équidistants $x_i = a + i\frac{b-a}{n}$ pour $i \in [\![0, n]\!]$. Alors on peut montrer que si $A_n = \{x_i, i \in [\![0, n]\!]\}$, alors

$$\|w_{A_n}\|_\infty = O(\left(\frac{b-a}{e}\right)^n)$$

ce qui améliore l'estimation précédente. Ceci permet d'améliorer la condition sur le rayon de convergence de l'exemple 1.13 en

$$R > (\frac{1}{e} + \frac{1}{2})(b-a).$$

2. On peut aller plus loin et se demander comment choisir les points $A = \{x_i\}$ de façon à minimiser $\|w_A\|_\infty$. La réponse est obtenue par les points dits de Tchebychev : si $A_0 = \{x_i, 0 \leq i \leq n\}$ avec $x_i = \frac{a+b}{2} + \frac{b-a}{2}\cos\frac{(2i+1)\pi}{2n+2}$ pour $0 \leq i \leq n$, alors pour tout choix de $n+1$ points d'interpolation A, on a

$$\|w_A\|_{\infty,[a,b]} \geq \|w_{A_0}\|_{\infty,[a,b]} = 2\left(\frac{b-a}{4}\right)^{n+1}.$$

Ceci permet d'améliorer la condition sur le rayon de convergence de l'exemple 1.13 en

$$R > (\frac{1}{4} + \frac{1}{2})(b-a).$$

En fait, dans ce cas, une analyse beaucoup plus détaillée (voir [Dem16, Section 4.3]) permet de montrer

Théorème 1.15. *On suppose $f : [a, b] \to \mathbb{R}$ lipschitzienne. Alors la suite $(P_n)_{n\in\mathbb{N}}$ des polynômes d'interpolation associée aux points de Tchebychev converge uniformément vers f sur $[a, b]$.*

Pour conclure cette section, remarquons que malgré ces réponses positives de convergence, il n'y a pas convergence de l'interpolation de Lagrange en général,

1. ni pour des points d'interpolation quelconques,
2. ni pour les points équidistants (on parle de phénomène de Runge) même pour des fonctions régulières, voir l'exemple 1.16 ci-après,
3. ni pour les points de Tchebychev quand la fonction est seulement continue, voir la remarque 1.17.

Exemple 1.16. Considérons l'exemple très classique

$$f_\alpha(x) = \frac{1}{x^2 + \alpha^2}$$

pour $\alpha > 0$, qu'on cherche à interpoler sur $[-1, 1]$. Cette fonction est développable en série entière en 0, de rayon de convergence $\alpha > 0$.

1. Par l'exemple 1.13, si $\alpha > 3$, il y a bien convergence de l'interpolation de Lagrange sur $[-1, 1]$ pour toute subdivision.
2. Par la remarque 1.14, il y a convergence pour les points uniformes si $\alpha > 1 + \frac{2}{e} \sim 1,74$. On peut affiner l'analyse et montrer (voir [Dem16, Chap II - Section 2.3]) qu'il y a convergence (uniforme) de l'interpolation dès que $\alpha > 0,527$, mais que pour $\alpha < 0,526$, l'interpolation diverge au bord de l'intervalle $]-1, 1[$. On parle de phénomène de Runge.

Remarque 1.17. [*, voir [Dem16, Chap II - Section 4] et [ACL99, Chap XXII]] Pour voir qu'il existe des fonctions telle que la méthode d'interpolation diverge, on peut utiliser une approche de type "analyse fonctionnelle". Plus précisément, étant donnés des points x_i distincts de $[a, b]$, on peut étudier

$$\begin{array}{cccc} \Lambda_n : & C^0([a,b]) & \longrightarrow & \mathbb{R}_n[X] \\ & f & \longmapsto & P_n \end{array}$$

où P_n est l'interpolation de Lagrange de f en les x_i. On peut montrer que Λ_n est une application linéaire continue (on munit $C^0([a, b])$ de la norme $\|\cdot\|_\infty$) et que

$$\|\Lambda_n\| = \sup_{x \in [a,b]} \sum_{i=0}^{n} |L_i(x)|$$

Néanmoins, même dans le cas des points de Tchebychev, on peut voir que $\|\Lambda_n\| \xrightarrow[n\to\infty]{} +\infty$, donc par le théorème de Banach-Steinhaus, il existe $f \in C^0([a, b])$ telle que les polynômes d'interpolation de f en les points de Tchebychev ne converge pas vers f.

En conclusion, on a montré que l'idée de l'interpolation de Lagrange, qui consiste à trouver un polynôme de haut degré et qui est égal à la fonction f en de nombreux points ne fonctionne pas bien en général. On préfèrera en général l'idée suivante : on discrétise l'intervalle $[a, b]$ par des points équidistants $x_i = a + i(b - a)/n$, et on fait une interpolation de petit degré sur $[x_i, x_{i+1}]$, et ensuite on recolle les polynômes obtenus sur chacun des intervalles $[x_i, x_{i+1}]$ on parle d'interpolation polynomiale par morceaux. Par contre, plutôt que de demander uniquement à être égal à f en les points x_i, on peut aussi mettre des conditions sur les dérivées afin d'améliorer la régularité de la fonction obtenue.

1. si on ne met pas de conditions sur les dérivées, il suffit de prendre des polynômes de degré 1 sur chaque intervalle $[x_i, x_{i+1}]$ et on parle d'interpolation affine par morceaux.
2. On peut montrer qu'on peut obtenir une interpolation de classe C^2 en cherchant des polynômes de degré 3, et en fixant les dérivées en a et b. Il s'agit des splines cubiques.

2 Résolution approchée d'équations non linéaires

On se propose ici de trouver des approximations des solutions d'équations de la forme $f(x) = 0$ où $f : U \to \mathbb{R}^n$ où U est un sous-ensemble de $\mathbb{R}^n$. Quand $n \geq 2$ on pourra parler de système d'équations.

Quand f est affine, on parle en général d'équation linéaire, et quand on ne suppose pas cette propriété sur f, on dit que l'équation est (a priori) non-linéaire. Remarquons que le cas des équations linéaires possède des méthodes spécifiques (que nous n'aborderons pas ici, elles ont été abordées en cours d'algèbre), néanmoins certaines méthodes dans ce qui suit peuvent également servir dans les cas linéaires.

Les méthodes que nous allons aborder dans ce paragraphe seront dites itératives, c'est-à-dire qu'elles font intervenir une récurrence, et on espère qu'à chaque étape de la récurrence l'objet calculé sera de plus en plus proche d'une solution de l'équation. Un grand avantage de ces méthodes, est que le temps de calcul est une fonction croissante de la précision souhaitée : si on veut juste une idée approximative de la solution, on peut se contenter de quelques étapes de l'algorithme. Alors qu'une méthode non itérative ne donnera une solution que quand l'algorithme aura "terminé" : une méthode non itérative typique est la méthode du pivot de Gauss, qui ne fournit aucun élément de réponse si on l'arrête en cours de route. Pour les méthodes itératives, il n'est pas rare de combiner des algorithmes : un premier algorithme peu couteux et qui n'est pas très sensible à la condition initiale, mais qui converge lentement, et qui donnera une première estimation de la solution qui pourra ensuite être utilisée par un second algorithme beaucoup plus précis, mais qui n'aurait pas convergé si on était parti trop loin de la solution.

Remarquons déjà que ce problème est étroitement lié à l'approximation numérique des solutions du problème d'optimisation

$$\min\left\{g(x),\ \ x\in U\right\} \tag{2.1}$$

où $g : U \to \mathbb{R}$ (on entend par là trouver la ou les valeurs de $x_0 \in U$ telles que $g(x_0) = \inf_U g$). En effet,

— Si U est un ouvert, g est différentiable, et si x_0 est une solution de (2.1), on verra à la section 5 du Chapitre VII que x_0 est solution de l'équation $\nabla g(x_0) = 0$ (on dit que x_0 est un point critique de g). Cette condition n'est pas suffisante en général (voir le point suivant pour une condition où elle devient suffisante), mais elle permet tout de même dans de nombreux cas d'identifier des "candidats" à l'optimalité : dans le cas où il y a un nombre fini (raisonnable) de solutions à cette équation, il n'est pas difficile choisir le bon candidat, par exemple en comparant la valeur de g en ces valeurs.

— Si U est un ouvert convexe, g est convexe et différentiable, alors

$$g(x_0) = \inf_U g \quad \Leftrightarrow \quad \nabla g(x_0) = 0.$$

— Si $U = \{x,\ h(x) = 0\}$ avec $h : \mathbb{R}^n \to \mathbb{R}^p$, nous verrons que sous certaines conditions (voir le Théorème 5.20 au Chapitre **??**), si x_0 est un minimum de g sur U, alors il existe $\lambda \in \mathbb{R}^p$ tel que le couple (x_0, λ) est solution du système :

$$\begin{cases} \nabla g(x_0) = \sum_{i=1}^p \lambda_i \nabla h_i(x_0) \\ h(x_0) = 0 \end{cases}$$

Encore une fois il ne s'agit que d'une condition nécessaire.

2.1 Cas de la dimension 1

On se donne $f : [a, b] \to \mathbb{R}$ continue et telle que $f(a)f(b) < 0$. Par le théorème des valeurs intermédiaires, on sait qu'il existe $c \in]a, b[$ tel que $f(c) = 0$.

Méthode de Dichotomie : elle consiste à diviser à chaque itération l'intervalle en deux sous-intervalles et de choisir un des intervalles dont on est sûr qu'il contient encore une solution de l'équation.

Description de l'algorithme :

— Initialisation : On pose $a_0 = a$ et $b_0 = b$.

— Récurrence : Pour $n \in \mathbb{N}$, on suppose déjà calculés (a_n, b_n), et on calcule $c_n = \frac{a_n+b_n}{2}$, $f(a_n), f(b_n), f(c_n)$.

1. Si $f(c_n) = 0$, l'algorithme s'arrête et on a trouvé une solution.

2. Si $f(a_n)f(c_n) > 0$, on pose $a_{n+1} = c_n$ et $b_{n+1} = b_n$.
3. Si $f(a_n)f(c_n) < 0$, on pose $a_{n+1} = a_n$ et $b_{n+1} = c_n$.

Il est facile de montrer le résultat suivant :

Théorème 2.1. *On suppose $f : [a, b] \to \mathbb{R}$ continue et telle que $f(a)f(b) < 0$, et on note $(a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}$ la suite obtenue par l'algorithme précédent, dans le cas où l'algorithme ne s'arrête jamais. Alors (a_n) est croissante, (b_n) est décroissante, et les deux suites converge vers un élément c tel que $f(c) = 0$. De plus*

$$\forall n \in \mathbb{N}, \quad |a_n - c| \leq \frac{b-a}{2^n}, \quad |b_n - c| \leq \frac{b-a}{2^n}.$$

Ainsi, à chaque itération, l'erreur est divisé par 2, ou dit autrement on gagne un chiffre de précision dans l'écriture en base deux. On dit que l'algorithme converge linéairement : plus précisément on a les définitions :

Définition 2.2 (Ordre de convergence d'une suite ou d'un algorithme)**.** Soit (x_n) une suite et $x \in \mathbb{R}$.

1. On dit que (x_n) converge (au moins) linéairement (ou à l'ordre 1) vers x s'il existe $C > 0$ et $\alpha \in]0, 1[$ telles que
$$\forall n \in \mathbb{N}, \ \ |x_n - x| \leq C\alpha^n.$$
La borne inférieure des α tels que cette inégalité est valable pour une certaine constante C s'appelle le taux de convergence linéaire de la suite, et si cette borne inférieure est 0 on dit que la convergence est super-linéaire.
2. On dit que (x_n) converge (au moins) à l'ordre $\beta > 1$ vers x s'il existe $C > 0$ et $\alpha \in]0, 1[$ telles que
$$\forall n \in \mathbb{N}, \ \ |x_n - x| \leq C\alpha^{(\beta^n)}.$$
Si cette propriété est valable pour $\beta = 2$, on dit que la convergence est (au moins) quadratique. Plus généralement la borne inférieure des β tels que cette inégalité est valable pour un certain couple (C, α) s'appelle l'ordre de convergence de la suite.

Exercice 2.3. 1. En pratique, pour estimer l'erreur d'un algorithme, on obtient souvent une relation de récurrence sur $|x_n - x|$:
   (a) Montrer que s'il existe $\alpha \in]0, 1[$
   $$\forall n \in \mathbb{N}, \ \ |x_{n+1} - x| \leq \alpha |x_n - x|$$
   alors (x_n) converge (au moins) linéairement vers x.
   (b) De même, s'il existe $\beta > 1$ et $C > 0$ tel que
   $$\forall n \in \mathbb{N}, \ \ |x_{n+1} - x| \leq C|x_n - x|^\beta$$
   et si [5] $C|x_0 - x|^{\beta-1} < 1$, alors la (x_n) convergence vers x à l'ordre β.

2. Enfin, il arrivera que l'estimation précédente ne soit pas simple à obtenir car on ne connait pas x : on montre ici qu'on peut s'en sortir à partir d'une estimation récurrente de $|x_n - x_{n-1}|$ qui sont les incréments de l'algorithme :
   (a) Montrer que s'il existe $\alpha \in]0, 1[$
   $$\forall n \in \mathbb{N}^*, \ \ |x_{n+1} - x_n| \leq \alpha |x_n - x_{n-1}|$$
   alors il existe x tel que (x_n) converge linéairement vers x.
   (b) De même, s'il existe $\beta > 1$ et $C > 0$ tel que
   $$\forall n \in \mathbb{N}, \ \ |x_{n+1} - x_n| \leq C|x_n - x_{n-1}|^\beta$$
   et si $C^{\frac{1}{\beta-1}}|x_1 - x_0| < 1$, alors il existe x tel que (x_n) converge vers x à l'ordre β.

5. On peut évidemment demander cette condition à un certain rang n_0 plutôt qu'au rang 0. Notons que si on sait déjà que (x_n) converge vers x, alors justement la condition est vérifiée à un certain rang n_0.

Méthodes de point fixe La première remarque basique est que l'équation $f(x) = 0$ peut-être vue comme une équation de point fixe, par exemple par

$$x + f(x) = x, \quad x + \alpha f(x) = x \quad \text{ou plus généralement} \quad x + h(x)f(x) = x$$

où $\alpha \in \mathbb{R}^*$ et h est une fonction qui ne s'annule pas. On revient sur ce choix à la fin du paragraphe.

On se donne donc I un intervalle et $\varphi : I \to \mathbb{R}$ une fonction. Un algorithme de point fixe consiste simplement en l'itération de φ, c'est-à-dire la donnée de $x_0 \in I$ et

$$\forall n \in \mathbb{N}, \quad x_{n+1} = \varphi(x_n). \tag{2.2}$$

Attention, si on ne suppose pas a priori que I est stable par φ, cette récurrence n'est pas forcément définie pour tout $n \in \mathbb{N}$.

On résume dans l'énoncé suivant les propriétés de cette méthode :

Théorème 2.4. *Soit I un intervalle, $\varphi : I \to \mathbb{R}$, et (x_n) définie par $x_0 \in I$ et (2.2)*

1. *Si (x_n) est bien définie pour tout $n \in \mathbb{N}$ et converge vers $\ell \in I$, alors $\varphi(\ell) = \ell$.*
2. *Si I est un segment $[a, b]$ et $\varphi([a, b]) \subset [a, b]$, alors il existe $\ell \in [a, b]$ tel que $\varphi(\ell) = \ell$. De plus, la suite (x_n) définie par (2.2) est bien définie, par contre, on ne peut pas conclure en général qu'elle converge.*
3. *Si I est fermé, $\varphi(I) \subset I$ et si φ est contractante[6], alors il existe un unique ℓ tel que $\varphi(\ell) = \ell$ et la suite x_n converge linéairement vers ℓ.*
4. *Si φ est de classe C^1 et possède un point fixe ℓ à l'intérieur de I tel que $|\varphi'(\ell)| < 1$, alors il existe $\alpha > 0$ tel que l'intervalle $]\ell - \alpha, \ell + \alpha[$ est stable par φ.*
   - *(a) De plus, la suite (x_n) définie par (2.2) avec $x_0 \in]\ell - \alpha, \ell + \alpha[$ converge linéairement vers ℓ.*
   - *(b) Si de plus φ est C^2 et $\varphi'(\ell) = 0$ alors la convergence de (x_n) vers ℓ est quadratique.*

Remarque 2.5. Un point fixe ℓ tel que $|\varphi'(\ell)| > 1$ est dit répulsif, si x_0 est proche mais différent de ℓ, les itérées x_n vont rapidement sortir de tout voisinage de ℓ, voir la proposition 6.4 au Chapitre I.

Remarque 2.6. Lorsque ℓ est un point fixe attractif, la convergence est linéaire et on peut voir (regarder la preuve de la proposition 6.4 au Chapitre I) que le taux de convergence est égal à $|\varphi'(\ell)|$. La méthode est donc plus rapide que la méthode de dichotomie si $|\varphi'(\ell)| < \frac{1}{2}$.

Démonstration.
1. Il suffit de passer à la limite dans $x_{n+1} = \varphi(x_n)$ et d'utiliser la continuité de φ en ℓ.
2. Il s'agit d'appliquer le théorème des valeurs intermédiaires à $x \mapsto \varphi(x) - x$ qui est positive (ou nulle) en a et négative (ou nulle) en b.
   Pour voir qu'on ne peut pas conclure à la convergence de (x_n) il suffit d'observer le cas de $\varphi : x \in [-10, 10] \to -x \in \mathbb{R}$ et $x_0 = 5$ qui donne $x_n = (-1)^n 5$ bien que φ ait effectivement un point fixe en 0.
3. Il s'agit du cas particulier du théorème de point fixe pour les fonctions réelles : en effet I étant fermé dans $\mathbb{R}$, il s'agit d'un espace métrique complet (on le munit bien sûr de la métrique induite par la valeur absolue).
4. (a) Voir la proposition 6.4 au Chapitre I.
   (b) Par le point précédent, si x_0 est suffisamment proche de ℓ alors la récurrence est bien définie. De plus, par la formule de Taylor-Lagrange, il existe $c_n \in [\ell - \alpha, \ell + \alpha]$ tel que

$$x_{n+1} - \ell = \varphi(x_n) - \varphi(\ell) = \varphi'(\ell)(x_n - \ell) + \frac{1}{2}(x_n - \ell)^2 \varphi''(c_n)$$

   d'où

$$\forall n \in \mathbb{N}, \quad |x_{n+1} - \ell| \leq \frac{1}{2} \|\varphi''\|_\infty |x_n - \ell|^2$$

   et donc avec l'exercice 2.3 on conclut que la convergence est quadratique.

□

6. Rappelons qu'une fonction est dite contractante si elle est k-lipschitzienne pour une constante $k < 1$.

Concluons en revenant sur la transformation de l'équation $f(x) = 0$ en $\varphi(x) = x$: si on regarde le cas de $\varphi(x) = x + \alpha f(x)$, on a $\varphi'(\ell) = 1 + \alpha f'(\ell)$, et afin d'accélérer la convergence au maximum, le choix optimal semble être

$$\alpha = \frac{-1}{f'(\ell)}$$

si bien sûr $f'(\ell) \neq 0$. Le problème est qu'on ne connait pas ℓ donc on ne peut pas choisir α de cette façon, mais on peut conserver cette idée en essayant de choisir α assez proche de cette valeur.

Cette remarque fournit une heuristique intéressante de la prochaine méthode, à savoir la méthode de Newton.

Méthode de Newton Cette fois, pour résoudre l'équation $f(x) = 0$, on va appliquer l'approche d'itération de la méthode de point fixe à la fonction

$$\varphi(x) = x - \frac{f(x)}{f'(x)},$$

Précisément, étant donné $f : I \to \mathbb{R}$ dérivable et $x_0 \in I$, on pose

$$\forall n \in \mathbb{N}, \;\; x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{2.3}$$

tant que $f'(x_n) \neq 0$.

Remarque 2.7. On peut aussi donner une autre heuristique de la méthode de Newton, qui fournit aussi une interprétation graphique : en effet, l'idée est de chercher à approcher f par son développement limité à l'ordre 1 (qui est une fonction affine) et de résoudre l'équation obtenue explicitement. On constate en effet que x_{n+1} est la solution de l'équation affine

$$f(x_n) + f'(x_n)(x - x_n) = 0.$$

Ainsi l'interprétation graphique est que x_{n+1} est l'intersection avec l'axe des abscisses de la tangente à f en x_n.

Le théorème suivant donne des éléments sur la convergence de cette méthode : le premier point montre que la méthode converge si on est assez proche de la solution, le second montre que dans le cas d'une fonction convexe, la méthode fonctionne dès qu'on part "à droite" de la solution.

Théorème 2.8. *On suppose f de classe C^2, et ℓ tel que $f(\ell) = 0$ et $f'(\ell) \neq 0$.*

1. *Alors il existe $\alpha > 0$ tel que $[\ell - \alpha, \ell + \alpha]$ est stable par φ et tel que pour tout $x_0 \in [\ell - \alpha, \ell + \alpha]$, la suite (x_n) définie par (2.3) est bien définie et converge vers ℓ de façon quadratique.*
2. *Si on suppose de plus qu'il existe $b > \ell$ tel que f est convexe sur $[\ell, b]$. Alors pour tout $x_0 \in [\ell, b]$, la suite (x_n) est décroissante et converge vers ℓ.*

On pourra voir une variante de ces résultats dans [Rou03, Exercice 49].

Exemple 2.9. On souhaite calculer numériquement $\sqrt{2}$, c'est-à-dire résoudre l'équation $f(x) := x^2 - 2 = 0$. L'algorithme de Newton s'écrit

$$\forall n \in \mathbb{N}, \;\; x_{n+1} = x_n - \frac{x_n^2 - 2}{2x_n} = \frac{1}{2}\left(x_n + \frac{2}{x_n}\right), \qquad \text{ou autrement dit} \quad \varphi(x) = \frac{1}{2}\left(x + \frac{2}{x}\right).$$

Par le second point du théorème 2.8, toute condition initiale $x_0 > \sqrt{2}$ donne une suite (x_n) qui converge vers $\sqrt{2}$. Remarquons que si $x_0 \in [0, \sqrt{2}[$, alors en fait $x_1 > \sqrt{2}$, donc la convergence a encore lieu.

Exemple 2.10. Pour factoriser un polynôme scindé simple dans $\mathbb{R}[X]$, on peut procéder ainsi : on note $r_1 < \ldots < r_d$ les racines réelles de P et on s'arrange pour que P soit unitaire.

— Pour $x \geq r_d$, on a $P'(x) > 0$ et $P''(x) > 0$ [7], et la méthode de Newton converge pour toute initialisation suffisamment grande.

— Pour trouver r_{d-1} (et successivement les autres racines), on peut chercher Q tel que $P = (X - r_d)Q$, et appliquer la méthode précédente à Q qui satisfait les mêmes hypothèses que P. Ce calcul de Q n'est néanmoins pas simple, et on peut le contourner en constatant que

$$\frac{P'}{P} = \frac{1}{X - r_d} + \frac{Q'}{Q},$$

donc la méthode de Newton appliquée à Q peut en fait s'écrire :

$$x_{n+1} = x_n - \frac{1}{\frac{P(x_n)}{P'(x_n)} - \frac{1}{x_n - r_d}}.$$

Remarque 2.11 (Méthode de la sécante)**.** L'inconvénient majeur de la méthode de Newton, est la nécessité de calculer la dérivée de f [8]. La méthode de la sécante consiste (en ayant en tête l'interprétation graphique de la méthode de Newton) à remplacer la tangente par la droite passant par les points $(x_{n-1}, f(x_{n-1})$ et $(x_n, f(x_n))$. D'un point de vue plus analytique, on remplace $f'(x_n)$ par le taux de variation $\frac{f(x_n)-f(x_{n-1})}{x_n - x_{n-1}}$. On obtient l'algorithme :

— on se donne $(x_0, x_1) \in I$,

— pour tout $n \geq 1$, on pose

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} f(x_n). \tag{2.4}$$

On peut montrer le résultat surprenant (voir par exemple [Gon]) :

Théorème 2.12. *On suppose $f : \mathbb{R} \to \mathbb{R}$ de classe C^3, et ℓ tel que $f(\ell) = 0$ et $f'(\ell) \neq 0$. Si x_0, x_1 sont assez proches de ℓ, alors la suite (x_n) donnée par (2.4) est bien définie, et converge vers ℓ à l'ordre au moins $\frac{1+\sqrt{5}}{2}$.*

2.2 Cas de la dimension $n \geq 2$

Méthode de point fixe Le théorème de point fixe qu'on a vu au Chapitre III et qu'on a vu dans le cas particulier des fonctions réelles à la section 2.1 est valable dans le cadre général des espaces de Banach, il s'appliquera donc si $\varphi : U \to \mathbb{R}^d$ est définie sur U fermé, et telle que $\varphi(U) \subset U$ et φ est contractante sur U.

Pour avoir la propriété de contractance, on verra au Chapitre VII qu'une condition suffisante est d'avoir $\sup_U \|D\varphi\| < 1$ (lorsque U est convexe) où $Dg(x)$ peut être vue comme la matrice jacobienne de φ en x, à savoir

$$J_\varphi(x) = \left(\frac{\partial \varphi_i}{\partial x_j}\right)_{1 \leq i,j \leq d}$$

et $\|\cdot\|$ est une norme subordonnée sur $\mathcal{M}_d(\mathbb{R})$.

Rappelons le résultat

$$\forall A \in \mathcal{M}_d(\mathbb{C}), \quad \inf\left\{N(A), N \text{ norme subordonnée}\right\} = \rho(A)$$

où $\rho(A)$ est le rayon spectral de A, c'est-à-dire $\rho(A) = \sup\{|\lambda|,\ \lambda \text{ valeur propre de } A\}$.

Ainsi le point 4 du théorème 2.4 devient :

Théorème 2.13. *Soit $\varphi : \mathbb{R}^d \to \mathbb{R}^d$ de classe C^1, et ℓ un point fixe de φ tel que $\rho(Dg(\ell)) < 1$. Alors il existe K un compact tel que $\ell \in \mathring{K}$, qui est stable par φ et il existe une norme N sur $\mathbb{R}^d$ tel que $\varphi_{|K}$ est contractante.*

1. *Ainsi si $x_0 \in K$ alors la suite (x_n) définie par $\forall n \in \mathbb{N},\ x_{n+1} = \varphi(x_n)$ est bien définie et converge vers ℓ linéairement.*
2. *Si de plus φ est de classe C^2 et $Dg(\ell) = 0$ alors la convergence est en fait quadratique.*

7. Les $d-1$ racines de P' sont dans les intervalles $]r_i, r_{i+1}[$, et les $d-2$ racines de P'' sont dans les intervalles formés par les racines de P'

8. Et n'oublions pas que si on veut résoudre un problème d'optimisation de g, on va appliquer la méthode à f' et donc devoir calculer f''.

Méthode Newton-Raphson Quand $f : U \subset \mathbb{R}^d \to \mathbb{R}^d$ et qu'on veut résoudre $f(x) = 0$, la récurrence (2.3) se généralise naturellement en

$$x_0 \in U, \quad \forall n \in \mathbb{N}, \;\; x_{n+1} = x_n - (J_f(x_n))^{-1}.f(x_n) \tag{2.5}$$

où $J_f(x_n)$ est la matrice Jacobienne de f en x_n, élément de de $\mathcal{M}_d(\mathbb{R})$, qu'on a supposée inversible ici, et $J_f(x_n)^{-1}.f(x_n)$ est le produit d'une matrice carrée de taille d et d'un vecteur de $\mathbb{R}^d$. Cette méthode s'appelle la méthode de Newton-Raphson.

Le point 1 du Théorème 2.8 (qui traite de la convergence locale) reste valide et on a :

Théorème 2.14. *Soit $f : U \subset \mathbb{R}^d \to \mathbb{R}^d$ de classe C^2, et $\ell \in U$ tel que $f(\ell) = 0$ et $J_f(\ell)$ est inversible. Alors il existe $r > 0$ tel que pour tout $x_0 \in B(\ell, r)$, la suite (x_n) définie par (2.5) est bien définie, reste dans $B(\ell, r)$, et converge quadratiquement vers ℓ.*

La preuve de ce résultat est très similaire à la preuve de la version uni-dimensionnelle.

Exemple 2.15. Dans le cas où on veut résoudre un problème d'optimisation de $g : U \to \mathbb{R}$, on peut appliquer l'algorithme précédent à $f = \nabla g$ et alors l'algorithme se réécrit

$$x_{n+1} = x_n - H_g(x_n)^{-1}.\nabla g(x_n).$$

2.3 Méthodes d'optimisation

On prend maintenant le point de vue de l'optimisation. On verra dans des exemples que ce point de vue peut en fait être appliqué à la résolution de certaines équations non-linéaires.

On cherche donc à approcher $x^* \in U$ qui est tel que

$$g(x^*) = \min\{g(x), \; x \in U\}.$$

Le principe général des méthodes itératives ici est de passer de x_n à x_{n+1} de façon à avoir $g(x_{n+1}) < g(x_n)$ et même si possible de rendre $g(x_n) - g(x_{n+1})$ le plus grand possible.

Proposons de chercher x_{n+1} de la forme $x_n + th$ où $h \in \mathbb{R}^d$ est pensée comme une direction (la norme de h importe peu puisqu'on multiplie par t qui a vocation à être petit). Si on suppose g différentiable, on peut naturellement écrire[9] :

$$g(x_n + th) = g(x_n) + t\nabla g(x_n).h + o(t)$$

ce qui amène à dire que si $\nabla g(x_n).h < 0$ alors h est une direction de descente de g en x_n. On a plus précisément :

Lemme 2.16. *Soit $x \in U$ tel que $\nabla g(x) \neq 0$. Alors*

$$h = -\frac{\nabla g(x)}{\|\nabla g(x)\|}$$

est une direction de descente, et c'est même la meilleure au sens où

$$-\frac{\nabla g(x)}{\|\nabla g(x)\|} = \inf\{\nabla g(x_0).h, \; \|h\| = 1\}.$$

Démonstration. Par l'inégalité de Cauchy-Schwarz, on a $\nabla g(x_0).h \geq -\|\nabla g(x_0)\|$ pour tout h tel que $\|h\| = 1$, avec égalité pour $h = -\frac{\nabla g(x_0)}{\|\nabla g(x_0)\|}$. □

9. On munit $\mathbb{R}^d$ de sa structure euclidienne, d'où la possibilité d'assimiler $dg(x)$ à $\nabla g(x)$.

Méthode de gradient à pas constant On se donne $\tau > 0$ fixé, et on propose l'algorithme :

$$x_{n+1} = x_n - \tau \nabla g(x_n), \tag{2.6}$$

où $x_0 \in U$ est donné.

Dans l'énoncé suivant on donne des conditions suffisantes d'existence, dans un cas local et dans un cas global (voir aussi [Fro]) :

Théorème 2.17. *Soit $g : \mathbb{R}^d \to \mathbb{R}$ de classe C^2, et x^* un minimiseur de g.*

1. *On suppose qu'il existe $\alpha > 0$ tel que $H_g(x^*) \geq \alpha I_d$. Alors pour x_0 suffisamment proche de x^* et τ assez petit, la suite (x_n) définie par (2.6) converge linéairement vers x^*.*
2. *On suppose qu'il existe $\alpha > 0$ et $L \in \mathbb{R}_+$ tels que g est α-convexe, c'est-à-dire*

$$\forall (x,y) \in \mathbb{R}^d, \forall t \in [0,1], \quad g((1-t)x + ty) \leq (1-t)g(x) + tg(y) - \alpha \frac{t(1-t)}{2} \|x-y\|^2,$$

et ∇g est L-lipschitzienne sur $\mathbb{R}^d$. Alors si $\tau < 2\alpha/L^2$, pour tout $x_0 \in \mathbb{R}^d$ la suite (x_n) définie par (2.6) converge linéairement vers x^*.

Voir par exemple [Gon] pour le premier point, et [All12, Théorème 10.5.4] pour le second.

Remarque 2.18. 1. La condition $H_g(x^*) \geq \alpha I_d$ signifie par définition

$$\forall y \in \mathbb{R}^d, \quad {}^t y H_g(x^*) y \geq \alpha \|y\|^2$$

et est équivalente à avoir $\lambda_{min}(H_f(x^*)) \geq \alpha$ où $\lambda_{min}(S)$ est la plus petite valeur propre d'une matrice $S \in \mathbb{S}_d^+(\mathbb{R})$.

2. La condition g est α-convexe est équivalente (dans le cas où g est C^2) à supposer

$$\forall x \in \mathbb{R}^d, \quad H_g(x) \geq \alpha I_d,$$

voir l'exercice 5.10 au Chapitre VII.

Remarque 2.19. Il existe une méthode dite à pas optimal, qui consiste à remplacer (2.6) par

$$x_{n+1} = x_n - \tau_n \nabla g(x_n)$$

où $\tau_n > 0$ est choisi de sorte que

$$g(x_{n+1}) = \inf_{t \in \mathbb{R}} g(x_n - t\nabla g(x_n))$$

c'est-à-dire qu'on optimise g sur la droite affine passant par x_n et dirigée par $\nabla g(x_n)$. Du coup, à chaque étape de cet algorithme, on doit au fond résoudre un problème d'optimisation unidimensionnel, pour lequel on a les méthodes du paragraphe précédent. On trouvera par exemple un théorème de convergence dans [All12, Théorème 10.5.2].

3 Résolution numérique d'EDO

Commençons par préciser le cadre : on se donne I un intervalle de $\mathbb{R}$, U un ouvert de $\mathbb{R}^d$, $f : I \times U \to \mathbb{R}^d$ et enfin $(t_0, y_0) \in I \times U$. On cherche à résoudre le problème de Cauchy

$$\left\{ \begin{array}{rcl} y' & = & f(t,y) \\ y(t_0) & = & y_0 \end{array} \right.$$

où la première ligne signifie $y'(t) = f(t, y(t))$ pour tout t dans un intervalle contenant t_0. L'inconnue est donc la donnée du couple (J, y) où J est un intervalle contenant t_0 et y est une fonction définie et dérivable sur J.

Nous analyserons ce type d'équation au Chapitre X ; sous certaines hypothèses sur f (par exemple si f est de classe C^1), le problème précédent admet effectivement une unique solution dont l'ensemble de définition est le plus grand possible (on parle de solution "maximale").

Dans ce paragraphe, on se demande plutôt comment résoudre numériquement ce type de problème : en effet, il est relativement rare de pouvoir les résoudre explicitement, il est donc naturel de chercher à obtenir une approximation raisonnable de la solution.

3.1 Définition du schéma d'Euler explicite

Afin de fixer les idées, on va supposer pour simplifier que f est bien définie sur $[0,T]\times\mathbb{R}^d$, et $t_0=0$. On se donne $T>0$ et on cherche à résoudre

$$\left\{\begin{array}{rcl} y' & = & f(t,y) \text{ sur } [0,T] \\ y(0) & = & y_0 \end{array}\right. \tag{3.1}$$

Définition 3.1. Soit $N\in\mathbb{N}^*$ et $h=T/N$ qu'on appelle le pas de la méthode. On rappelle que $f:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$ et $y_0\in\mathbb{R}^d$ sont donnés. On pose alors $\forall n\in[\![0,N]\!],\quad t_n=nh$ qui fournit une subdivision à pas constant de $[0,T]$. Le schéma d'Euler explicite est le vecteur $(y_n)_{n\in[\![0,N]\!]}$ obtenu par la récurrence

$$\forall n\in[\![0,N-1]\!],\quad y_{n+1}=y_n+hf(t_n,y_n). \tag{3.2}$$

Remarque 3.2. — Attention, gardons en tête que pour chaque $N\in\mathbb{N}^*$, on a donc un vecteur $(y_n)_{n\in[\![0,N]\!]}$ dont toutes les composantes dépendent de N, même si on n'en a pas noté la dépendance pour ne pas alléger les notations.

— On a vocation à faire N grand, ce qui revient à avoir h petit. Dans ce cas, on espère que y_n pour $n\in[\![0,N]\!]$ sera une bonne approximation de $y(t_n)$ où y est la[10] solution de (3.1).

— On comprend la définition précédente on constatant que si $y:[0,T]$ est solution de (3.1), alors

$$\forall n\in[\![0,N-1]\!],\quad y(t_{n+1})=y(t_n)+\int_{t_n}^{t_{n+1}}f(s,y(s))ds$$

donc on a approché l'intégrale du terme de droite par la méthode des rectangles à gauche (avec un seul rectangle car $t_{n+1}-t_n=h$ a vocation à être petit).

Exemple 3.3. Regardons le cas où f ne dépend pas de y, c'est-à-dire $f:\mathbb{R}_+\to\mathbb{R}$, et on veut résoudre

$$\left\{\begin{array}{rcl} y'(t) & = & f(t),\ \forall t\in[0,T] \\ y(0) & = & y_0 \end{array}\right.$$

et dont la solution est bien sûr (disons si f est continue) :

$$y(t)=y_0+\int_0^t f(s)ds.$$

Le schéma d'Euler s'écrit :

$$\forall n\in[\![0,N-1]\!],\quad y_{n+1}=y_n+hf(t_n)$$

et sa solution est

$$\forall n\in[\![0,N]\!],\quad y_n=y_0+\sum_{k=0}^{n}hf(t_k),$$

qui est effectivement la méthode des rectangles à gauche pour l'intégrale de f sur $[0,t_n]$ sur la subdivision $(t_k)_{k\in[\![0,n]\!]}$.

Exemple 3.4. On regarde maintenant $f(t,y)=ay$ où $a\in\mathbb{R}$, c'est-à-dire qu'on veut résoudre

$$\left\{\begin{array}{rcl} y' & = & ay,\ \forall t\in[0,T] \\ y(0) & = & y_0 \end{array}\right.$$

dont la solution est $t\mapsto y(t)=y_0e^{at}$. Le schéma d'Euler s'écrit :

$$\forall n\in[\![0,N-1]\!],\quad y_{n+1}=(1+ha)y_n$$

et sa solution est

$$\forall n\in[\![0,N]\!],\quad y_n=y_0(1+ah)^n.$$

Regardons la valeur pour $n=N$[11], on a $y_N=y_0(1+aT/N)^N$ qui effectivement converge vers y_0e^{aT} quand N tend vers $+\infty$.

10. Pour le moment on ne sait pas si cette solution existe et/ou est unique car on n'a fait aucune hypothèse sur f, mais on imagine pour simplifier qu'on est dans une telle situation

11. Sinon à n fixé il n'y a pas vraiment de sens à faire tendre h vers 0 : en effet y_n est censé approcher $y(t_n)$ qui dépend de N (rappelons que $t_n=nT/N$), donc si on fixait n et on faisait $N\to\infty$, on a $t_n\to 0$ et effectivement $y_0(1+ah)^n$ converge vers 1 mais ça ne nous apprend rien.

On termine par définir ce que signifie la convergence du schéma d'Euler.

Définition 3.5. On suppose que (3.1) a une unique solution y. On appelle erreur du schéma d'Euler la suite

$$\sup_{n\in[\![0,N]\!]} \|y(t_n) - y_n\|$$

et on dit que le schéma est convergent si

$$\lim_{N\to\infty} \sup_{n\in[\![0,N]\!]} \|y(t_n) - y_n\| = 0.$$

et on dit qu'il est convergent d'ordre $p \in \mathbb{N}^*$ si

$$\sup_{n\in[\![0,N]\!]} \|y(t_n) - y_n\| \leq Ch^p,$$

pour une certaine constante C.

3.2 Analyse du schéma d'Euler explicite

On va étudier dans l'énoncé suivant une estimation de l'erreur du schéma d'Euler explicite sous certaines hypothèses sur f. On prend la peine aussi de regarder la dépendance en la condition initiale : en effet, il se peut qu'on fasse une erreur sur cette condition initiale, il est donc important de prendre en compte cette erreur dans les estimations :

Théorème 3.6. *On suppose $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ de classe C^1 et globalement lipschitzienne en la seconde variable, c'est-à-dire qu'il existe $L \in \mathbb{R}_+$ tel que*

$$\forall t \in [0,T],\ \forall (x_1, x_2) \in (\mathbb{R}^d)^2,\ \ \|f(t,x_2) - f(t,x_1)\| \leq L\|x_2 - x_1\|.$$

Sous ces conditions, pour $\hat{y}_0 \in \mathbb{R}^d$ le problème

$$\begin{cases} y' &= f(t,y) \text{ sur } [0,T] \\ y(0) &= \hat{y}_0 \end{cases}$$

admet une unique solution sur $[0,T]$ qu'on note $\hat{y}$, voir le Théorème 1.17 au Chapitre X. On définit pour $N \in \mathbb{N}^*$ le vecteur $(y_n)_{n\in[\![0,N]\!]}$ obtenu par le schéma d'Euler explicite partant de $y_0 \in \mathbb{R}^d$, voir la définition 3.1. Alors il existe $A, B \in \mathbb{R}_+$ tels que[12]

$$\sup_{n\in[\![0,N]\!]} |\hat{y}(t_n) - y_n| \leq A\|\hat{y}_0 - y_0\| + Bh.$$

En particulier, sous ces conditions, le schéma d'Euler est convergent d'ordre 1.

Démonstration. **Étape 1 : relation de récurrence sur l'erreur au temps t_n** : On pose $\forall n \in [\![0,N]\!],\ \ z_n = \hat{y}(t_n) - y_n$. Alors pour $n \leq N-1$ on a

$$\begin{aligned} z_{n+1} &= \hat{y}(t_{n+1}) - y_{n+1} = \hat{y}(t_{n+1}) - y_n - hf(t_n,y_n) && (3.3)\\ &= z_n + \hat{y}(t_{n+1}) - \hat{y}(t_n) - hf(t_n,y_n) && (3.4)\end{aligned}$$

On remarque que comme f est de classe C^1, $\hat{y}$ est de classe C^2 vu que $\hat{y}'$ est composée de fonctions C^1. Sa dérivée seconde est donc une fonction bornée sur $[0,T]$. La formule de Taylor montre donc que[13]

$$\hat{y}(t_{n+1}) = \hat{y}(t_n) + h\hat{y}'(t_n) + O(h^2)$$

et donc en remplaçant $\hat{y}'(t_n)$ par $f(t_n,\hat{y}(t_n))$ on a finalement

$$z_{n+1} = z_n + h\Big[f(t_n,\hat{y}(t_n)) - f(t_n,y_n)\Big] + O(h^2)$$

12. On rappelle que $h = T/N$.
13. Ici $O(h^2)$ fait référence à une constante qui ne dépend pas de n ou de N.

avec, par hypothèse sur f :

$$\left\|f(t_n, \hat{y}(t_n)) - f(t_n, y_n)\right\| \leq L\|z_n\|.$$

Ainsi il existe $M \in \mathbb{R}_+$ tel que

$$\forall n \in [\![0, N-1]\!], \quad \|z_{n+1}\| \leq (1+Lh)\|z_n\| + Mh^2$$

Étape 2 : Résolution de cette récurrence : On pose $\forall n \in [\![0, N]\!], \quad \delta_n = \dfrac{\|z_n\|}{(1+Lh)^n}$. Alors

$$\forall n \in [\![0, N-1]\!], \quad \delta_{n+1} \leq \delta_n + \frac{Mh^2}{(1+Lh)^{n+1}}$$

et donc on peut faire la somme téléscopique des $(\delta_{k+1} - \delta_k)_{k \in [\![0, n-1]\!]}$ et on obtient

$$\forall n \in [\![0, N]\!], \quad \delta_n \leq \delta_0 + \sum_{k=0}^{n-1} \frac{Mh^2}{(1+Lh)^{k+1}} = \|z_0\| + \frac{Mh^2}{(1+Lh)} \frac{1 - \frac{1}{(1+Lh)^n}}{1 - \frac{1}{(1+Lh)}} \leq \|z_0\| + \frac{M}{L} h.$$

Mais on a de plus l'estimation

$$(1+Lh)^n \leq (1+Lh)^N = \exp\left(N \ln\left(1 + L\frac{T}{N}\right)\right) \leq e^{TL}$$

où on a utilisé l'inégalité $\ln(1+x) \leq x$ (qui peut se montrer par exemple par concavité de ln). En conclusion :

$$\forall n \in [\![0, N]\!], \quad \|z_n\| \leq e^{TL}\|z_0\| + \frac{Me^{TL}}{L} h.$$

□

Remarque 3.7. Dans la preuve ci-dessus, on a montré que si $(a_n)_{n \in [\![0,N]\!]}, (b_n)_{n \in [\![0,N]\!]}$ sont deux suites telles que

$$\forall n \in [\![0, N-1]\!], \ a_{n+1} \leq \lambda a_n + b_n$$

pour un $\lambda > 0$, alors

$$\forall n \in [\![0, N]\!], \ a_n \leq \lambda^n a_0 + \sum_{k=0}^{n-1} \lambda^{n-k-1} b_k$$

qui peut s'appeler le lemme de Gronwall discret, en référence au Lemme 1.17 du Chapitre X.

3.3 Les méthodes à un pas

On peut généraliser la récurrence (3.6) en :

$$y_{n+1} = y_n + h\Phi(t_n, y_n, h) \tag{3.5}$$

où Φ est une fonction continue.

On a voir désormais les propriétés attendues d'un tel schéma, et qui amènent à une convergence du schéma :

Définition 3.8 (Consistance)**.** Si y est une solution de (3.1), on appelle erreur de consistance du schéma (3.5) la suite

$$R_n = y(t_{n+1}) - y(t_n) - h\Phi(t_n, y(t_n), h).$$

On dira que le schéma est consistant si pour tout y solution de (3.1) on a

$$\sum_{n=0}^{N} \|R_n\| \xrightarrow[h \to 0]{} 0$$

et on dit que le schéma est consistant d'ordre $p \in \mathbb{N}^*$ si on a même

$$\sum_{n=0}^{N} \|R_n\| \leq Ch^p$$

pour une constante C.

Dans la définition ci-dessus, on a mis $y(t_n)$ à la place de y_n dans la récurrence (3.5) et R_n est l'erreur effectuée à l'étape n, c'est pourquoi la convergence est définie via la somme sur n.

Maintenant, on va définir la stabilité du schéma, qui correspond intuitivement au fait que le cumul des erreurs (d'approximation, de consistance, d'arrondis, etc...) au cours du calcul est contrôlé par la taille des erreurs commises :

Définition 3.9 (Stabilité)**.** On dit que le schéma (3.6) est stable s'il existe $M > 0$ une constante telle que pour toutes suites $(y_n)_{n\le N}$, $(\hat{y}_n)_{n\le N}, (\varepsilon_n)_{n\le N}$ telle que

$$\forall n \in [\![0, N-1]\!],\ \ y_{n+1} = y_n + h\Phi(t_n, y_n, h)$$

$$\forall n \in [\![0, N-1]\!],\ \ \hat{y}_{n+1} = \hat{y}_n + h\Phi(t_n, \hat{y}_n, h) + \varepsilon_n$$

on a

$$\sup_{n\in[\![0,N]\!]} |\hat{y}_n - y_n| \le M \left(\|\hat{y}_0 - y_0\| + \sum_{k=0}^{N-1} \|\varepsilon_k\| \right).$$

On résumé dans le résultat suivant des conditions pour que les propriétés précédentes soient satisfaites :

Théorème 3.10. *1. Si $\Phi(t, x, 0) = f(t, x)$ alors le schéma (3.6) est consistant.*

2. Si Φ est Lipschitzienne en la seconde variable, c'est-à-dire s'il existe L tel que :

$$\forall (x_1, x_2) \in (\mathbb{R}^d)^2,\ \forall t \in [0, T], \forall h > 0,\ \ \|\Phi(t, x_2, h) - \Phi(t, x_1, h)\| \le L\|x - y\|$$

alors la méthode est stable.

3. Si le schéma est stable et consistant, alors il est convergent. S'il est stable et consistant d'ordre $p \in \mathbb{N}^*$, alors il converge à l'ordre p.

Démonstration. 1. On a

$$\begin{aligned} R_n = y(t_{n+1}) - y(t_n) - h\Phi(t_n, y(t_n), h) &= \int_{t_n}^{t_{n+1}} f(s, y(s))ds - h\Phi(t_n, y(t_n), h) \\ &= \int_{t_n}^{t_{n+1}} \Big[\Phi(s, y(s), 0) - \Phi(t_n, y(t_n), h)\Big] ds. \end{aligned}$$

D'après le théorème de Heine, la fonction $(t, h) \mapsto \Phi(t, y(t), h)$ est uniformément continue sur le compact $[0, T] \times [0, h_0]$. Donc pour $\varepsilon > 0$, il existe $\delta > 0$ tel que

$$\forall (s, t) \text{ tel que } |t - s| < \delta, \forall h < \delta,\ \ \Big\|\Phi(t, y(t), h) - \Phi(s, y(s), 0)\Big\| \le \varepsilon$$

et alors si $t_{n+1} - t_n = h < \delta$, on peut écrire

$$\|R_n\| \le \int_{t_n}^{t_{n+1}} \varepsilon ds = \varepsilon(t_{n+1} - t_n)$$

d'où

$$\sum_{n=0}^{N} \|R_n\| \le T\varepsilon.$$

2. On pose $z_n = \hat{y}_n - y_n$. Alors pour $n \le N - 1$,

$$z_{n+1} = z_n + h\Big[\Phi(t_n, \hat{y}_n, h) - \Phi(t_n, y_n, h)\Big] + \varepsilon_n$$

d'où

$$\|z_{n+1}\| \le (1 + Lh)\|z_n\| + \|\varepsilon_n\|$$

ce qui donne par récurrence :

$$\forall n \le N,\ \ \|z_n\| \le (1 + Lh)^n \|z_0\| + \sum_{k=0}^{n-1} (1 + hL)^{n-k-1} \|\varepsilon_k\|.$$

Comme à la preuve du Théorème 3.6 on utilise l'inégalité $(1+hL)^n \leq e^{TL}$ et on obtient

$$\|z_n\| \leq e^{TL}\left(\|z_0\| + \sum_{k=0}^{n-1} \|\varepsilon_k\|\right),$$

d'où finalement

$$\max_{0\leq n\leq N} \|z_n\| \leq e^{TL}\left(\|z_0\| + \sum_{k=0}^{N-1} \|\varepsilon_k\|\right).$$

3. On applique la stabilité à la suite $\hat{y}_n = y(t_n)$ qui satisfait

$$\hat{y}_{n+1} = \hat{y}_n + h\Phi(t_n, \hat{y}_n, h) + \varepsilon_n, \quad \text{avec} \quad \varepsilon_n = y(t_{n+1}) - y(t_n) - h\Phi(t_n, y(t_n), h) = R_n.$$

□

Exemple 3.11. On peut citer comme exemple qui rentre naturellement dans cette catégorie de "méthode à un pas" la méthode Runge, qui est basée sur la méthode du point milieu pour approcher une intégrale. Cette méthode s'écrit :

$$y_{n+1} = y_n + hf(t_n + \frac{h}{2}, y_n + \frac{h}{2}f(t_n, y_n))$$

qui fait intervenir le point $t_n + \frac{h}{2}$ milieu de $[t_n, t_{n+1}]$, et on peut montrer que cette méthode est d'ordre 2.

Cette méthode s'appelle aussi la méthode RK2 pour Runge-Kutta d'ordre 2, qui se généralise avec notamment la méthode RK4 qui est d'ordre 4, en faisant intervenir plus de points intermédiaires de l'intervalle $[t_n, t_{n+1}]$ et qui se base sur une méthode dite de quadrature pour approcher l'intégrale sur $[t_n, t_{n+1}]$.

Remarque 3.12. D'autres méthodes numériques sont obtenues avec d'autres méthode d'approximation des intégrales : la méthode des rectangles à droite aboutit par exemple au schéma d'Euler implicite :

$$y_{n+1} = y_n + hf(t_{n+1}, y_{n+1}).$$

Cette relation n'est pas une relation de récurrence traditionnelle, on dit qu'elle est implicite car elle nécessite la résolution de l'équation

$$y = y_n + hf(t_{n+1}, y)$$

d'inconnue y pour définir y_{n+1}. Donc pour chaque n on doit résoudre une équation non linéaire en utilisant les outils du paragraphe 2, et il faut faire les hypothèses pour permettre qu'une telle équation ait une unique solution, au moins quand h est suffisamment petit [14]

3.4 Le Théorème de Peano

On a montré au Théorème 3.6, sous certaines hypothèses sur f, que la méthode d'Euler convergeait à l'ordre 1 vers l'unique solution du problème de Cauchy (3.1). On va démontrer ici que sous la seule hypothèse que f est continue, le schéma d'Euler converge à extraction près vers une solution locale [15]du problème de Cauchy (3.1). Il est important de noter que le théorème de Cauchy-Lipschitz fait des hypothèses plus fortes sur f que la simple continuité, et donc que ce résultat d'existence est très intéressant. Notons qu'on ne peut pas conclure à l'unicité en général, comme on le verra à l'exemple 1.14 du Chapitre X.

Théorème 3.13 (Théorème de Cauchy-Arzela-Peano)**.** *Soit U un ouvert de $\mathbb{K}^d$, I un intervalle ouvert de $\mathbb{R}$, $f : I \times U \to \mathbb{K}^d$ une fonction continue. Alors pour tout $(t_0, y_0) \in I \times U$, il existe un intervalle ouvert J contenant t_0 pour lequel le problème (1.8) admet une solution y définie sur J.*

14. En fait la fonction $\varphi : y \mapsto y + hf(t_{n+1}, y)$ est hL-Lipschitzienne si f est L-Lipschitzienne en la seconde variable, donc si $hL < 1$ le théorème de point fixe peut s'appliquer pour montrer l'existence et l'unicité d'un point fixe de cette fonction.

15. C'est-à-dire définie sur un voisinage de $t_0 = 0$.

Démonstration. [16] Comme $I \times U$ est ouvert, il existe $r_0 > 0$ tel que $\overline{B}_{r_0} := [t_0 - r_0, t_0 + r_0] \times \overline{B}(y_0, r_0) \subset I \times U$. On pose $M = \sup_{\overline{B}_{r_0}} \|f\|$ qui est fini par continuité de f.

On se donne $T \in]0, r_0]$ qui est tel que $T \leq \frac{r_0}{M}$.

1. On fixe $N \in \mathbb{N}^*$ et on va définir le schéma d'Euler sur l'intervalle $[t_0, t_0 + T]$ qui est légère adaptation de la définition 3.1 : [17]

$$\forall n \in [\![0, N]\!], \quad t_n := t_0 + n\frac{T}{N}$$

qui fournit une subdivision de $[t_0, t_0 + T]$ en N intervalles. Ensuite, on définit

$$\forall n \in [\![0, N-1]\!], \;\; y_{n+1} = y_n + \frac{T}{N} f(t_n, y_n).$$

Alors par récurrence immédiate on a

$$\forall n \in [\![0, N]\!], \|y_n - y_0\| \leq \frac{nT}{N} M$$

et comme $TM \leq r_0$ cela implique $y_n \in \overline{B}(y_0, r_0)$ pour tout $n \in [\![0, N]\!]$.

2. On définit $y_{(N)} : [t_0, t_0 + T] \to U$ la fonction continue et affine par morceaux [18] qui passe par les points (t_n, y_n) pour $n \in [\![0, N]\!]$. Par convexité de $\overline{B}_{r_0}$ et par le point précédent, on en déduit que $y_{(N)}$ est bien à valeurs dans $\overline{B}(y_0, r_0)$.
La suite de fonctions $(y_{(N)})_{N \in \mathbb{N}^*}$ ainsi définie est uniformément bornée, et elle satisfait

$$\forall N \in \mathbb{N}^*, \; \forall n \in [\![0, N-1]\!], \forall t \in [t_n, t_{n+1}], \quad \|y'_{(N)}(t)\| = \|f(t_n, y_n)\| \leq M$$

donc la famille $(y_{(N)})_{N \in \mathbb{N}^*}$ est équi-lipschitzienne.

3. Par le théorème d'Ascoli 3.7 au Chapitre III, il existe une suite extraite (on omet d'écrire l'extraction) de $(y_{(N)})_{N \in \mathbb{N}^*}$ et y_∞ telle que $(y_{(N)})_{N \in \mathbb{N}^*}$ converge vers y_∞ uniformément.

4. On veut montrer que y_∞ satisfait l'équation différentielle $y' = f(t, y)$. On va écrire une équation intégrale satisfaite par $y_{(N)}$ et passer à la limite. On fixe $N \in \mathbb{N}^*$ et on constate

$$\forall n \in [\![0, N-1]\!], \; \forall t \in [t_n, t_{n+1}], \;\; y_{(N)}(t) = y_n + (t - t_n) f(t_n, y_n) = y_n + \int_{t_n}^{t} f(t_n, y_n) ds \tag{3.6}$$

donc on est amené à poser

$$\forall n \in [\![0, N-1]\!], \; \forall t \in [t_n, t_{n+1}[, \;\; \bar{y}_{(N)}(t) = y_n$$

la fonction constante par morceaux passant par les points $(t_n, y_n)_{n \in [\![0, N]\!]}$ et

$$\forall n \in [\![0, N-1]\!], \; \forall t \in [t_n, t_{n+1}[, \; \forall y \in U \quad \bar{f}_{(N)}(t, y) = f(t_n, y)$$

et alors (3.6) devient après sommation

$$\forall t \in [t_0, t_0 + T], \;\; y_{(N)}(t) = y_0 + \int_{t_0}^{t} \bar{f}_{(N)}(s, \bar{y}_{(N)}(s)) dt. \tag{3.7}$$

5. Soit $\varepsilon > 0$. Par le théorème de Heine, la fonction f est uniformément continue sur $\overline{B}_{r_0}$, donc il existe $\eta > 0$ tel que

$$\forall (t_1, y_1), (t_2, y_2) \in \overline{B}_{r_0}, \quad \|f(t_2, y_2)) - f(t_1, y_1)\| \leq \varepsilon.$$

16. On trouve une variante de la preuve exposée ci-dessous dans [Dem16, pages 121-128] mais la présentation y est très longue ! On suit plutôt [Boyb, page 8] et [LM99, Exercice 1.3.5].
17. Attention, les nombres (t_n, y_n) dépendent en fait de N ; on n'a pas souhaité noter (t_n^N, y_n^N) pour ne pas alourdir les notations.
18. Cela signifie que pour $t \in [t_n, t_{n+1}]$, $\;y_{(N)}(t) = y_n + (t - t_n) f(t_n, y_n)$.

Pour $N \in \mathbb{N}^*$ on a

$$\|\bar{y}_{(N)} - y_\infty\|_\infty \leq \|\bar{y}_{(N)} - y_{(N)}\|_\infty + \|y_{(N)} - y_\infty\|_\infty \leq \frac{TM}{N} + \|y_{(N)} - y_\infty\|_\infty$$

donc par convergence uniforme de $y_{(N)}$ vers y_∞, il existe $N_0 \in \mathbb{N}^*$ tel que pour tout $N \geq N_0$,

$$\|\bar{y}_{(N)} - y_\infty\|_\infty \leq \delta,$$

condition à laquelle on ajoute $\frac{T}{N} \leq \delta$.
Ainsi pour $t \in [t_0, t_0 + T]$, et $N \geq N_0$, on note $n \in [\![0, N-1]\!]$ tel que $t \in [t_n, t_{n+1}]$. Alors

$$\|f(t, y_\infty(t)) - \bar{f}_{(N)}(t, \bar{y}_{(N)}(t))\| = \|f(t, y_\infty(t)) - f(t_n, \bar{y}_{(N)}(t))\| \leq \varepsilon$$

où la majoration par ε vient du fait que $|t - t_n| \leq \frac{T}{N} \leq \delta$ et $\|\bar{y}_{(N)}(t)) - y_\infty(t)\| \leq \delta$.
On a donc montré que

$$\forall \varepsilon > 0, \exists N_0 \in \mathbb{N}^*, \forall N \geq N_0, \ \ \forall t \in [t_0, t_0 + T], \ \ \|f(t, y_\infty(t)) - \bar{f}_{(N)}(t, \bar{y}_{(N)}(t))\| \leq \varepsilon$$

ce qui montre une convergence uniforme de $\bar{f}_{(N)}(\cdot, \bar{y}_{(N)}(\cdot))$ vers $f(\cdot, y_\infty(\cdot))$, et au final on peut passer à la limite dans (3.7) et on obtient

$$\forall t \in [t_0, t_0 + T], \ \ y_\infty(t) = y_0 + \int_{t_0}^{t} f(s, y_\infty(s)) ds,$$

et montre que y_∞ est solution du problème de Cauchy sur $[t_0, t_0 + T]$.

6. Afin de traiter les temps négatifs, on peut appliquer ce qui précède à $(t, z) \mapsto -f(t_0 - t, z)$ qui est encore continue et donc il existe une solution z à $z' = -f(t_0 - \cdot, z)$ sur $[0, T]$. On pose alors $y(t) = z(t_0 - t)$ pour $t \in [t_0 - T, t_0]$ et y satisfait

$$\forall t \in [t_0 - T, t_0], \ y'(t) = -z'(t_0 - t) = f(t, z(t_0 - t)) = f(t, y(t)).$$

On peut recoller les solutions trouvées sur $[t_0 - T, t_0]$ et $[t_0, t_0 + T]$, ce qui donnera une fonction C^1 au travers de t_0 : en effet la condition de Cauchy prouve la continuité de y en t_0, et l'équation différentielle montre que la dérivée à droite et à gauche en t_0 sont les mêmes.

$\square$

Remarque 3.14. On peut noter qu'il existe d'autres démonstrations du théorème de Peano :

1. En régularisant f par convolution et en utilisant le théorème de Cauchy-Lipschitz : voir [QZ13, page 359-362].
2. En utilisant le théorème de point fixe de Schauder, qui dit que si C est un convexe fermé non vide d'un e.v.n. E et si $f : C \to C$ est une application continue telle que $\overline{f(C)}$ est compact dans E, alors f admet un point fixe [19]. Voir par exemple [CL97].

19. Ce théorème de Schauder est lui-même une conséquence du théorème de point fixe de Brouwer dans la version suivante : "toute fonction continue $f : C \to C$ où C est un convexe compact non vide de $\mathbb{K}^d$ admet un point fixe".

Chapitre VII

Calcul différentiel

On a vu au Chapitre I la notion de dérivation pour une fonction $f : \mathbb{R} \to \mathbb{R}$, et au Chapitre II la notion de $\mathbb{C}$-dérivation pour une fonction $f : \mathbb{C} \to \mathbb{C}$. Dans les deux cas, on pourrait remplacer l'espace d'arrivée par un $\mathbb{K}$-espace vectoriel normé (où $\mathbb{K}$ est respectivement $\mathbb{R}$ ou $\mathbb{C}$) : les définitions peuvent être adaptées sans difficulté à ce cadre [1] ; certaines propriétés de base seront facilement conservées (mais pas toutes). Dans les deux cas également, on a utilisé la notion de division dans l'espace de départ, et donc que ces derniers étaient munis d'une structure de corps ; on sait néanmoins que des espaces vectoriels de dimension $n \geq 3$ ne peuvent être munis d'une telle structure [2].

Dans ce chapitre, nous allons généraliser au cas où l'espace de départ est un espace vectoriel l'idée de la dérivation réelle via son interprétation sous forme de développement limité à l'ordre 1 (voir remarque 3.36 au Chapitre I), ou si on veut le voir géométriquement, le fait que la fonction est approchée par une droite. On parlera de différentiabilité. Notons déjà que l'on a expliqué au paragraphe 3.2 du Chapitre II que si on voit $\mathbb{C}$ comme un $\mathbb{R}$-espace vectoriel de dimension 2, la notion de différentiabilité abordée ici n'est pas identique à la notion de $\mathbb{C}$-dérivabilité, mais on a aussi établi les équations de Cauchy-Riemann qui permettent de voir qu'en ajoutant des conditions à la différentiabilité, on trouve une version équivalente à la $\mathbb{C}$-dérivabilité.

Au programme de l'oral, il y a deux leçons qui rentrent intégralement dans le contexte de ce chapitre : **214-215**. Mais les notions de ce chapitre sont utilisées dans des sujets comme les équations différentielles, les EDPs, le calcul d'intégrales de plusieurs variables, la notion de convexité, et les problèmes d'extremums, donc sont utiles et à valoriser pour les leçons **219-220-221-222-229-236-253**. Il semble également pertinent de les utiliser dans les leçons plus récentes **265** et **267**. Il n'est pas incongru non plus qu'elles interviennent dans les leçons **226-233**.

En ce qui concerne les références, il n'est pas difficile de trouver des livres de base sur les notions de ce chapitre ; voir par exemple [Wag12a]. On conseille fortement [Rou03] qui fait des rappels de cours, et donne des exercices très intéressants s'initier et pour agrémenter vos leçon d'agrégation. La lecture de [BMP05, Chapitre 1] est également instructive. D'autres classiques [3] sont [Ave83, Car67].

1 Fonctions différentiables sur un ouvert de $\mathbb{R}^n$

Dans tout le chapitre $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ désignent deux $\mathbb{R}$-espaces vectoriels normés [4], et Ω est un ouvert de E. Comme au Chapitre II, on se restreint à des fonctions définies sur des ouverts. [5]

1. Voir par exemple [QZ13, page 476] sur le théorème de Riesz-Thorin pour une utilisation de fonctions holomorphes à valeurs dans L^p.

2. Ceci s'énonce et se démontre rigoureusement bien sûr (on ne peut pas juste dire "qu'on n'en connaît pas" !). Voir par exemple le théorème de Frobenius.

3. Que je connais peu.

4. *Si les fonctions sont définies sur et/ou à valeur dans des $\mathbb{C}$-espaces vectoriels, on les considèrera comme des $\mathbb{R}$-espaces vectoriels (il suffit pour cela de restreindre la multiplication par un "scalaire" (définie a priori pour tout complexe) à $\mathbb{R}$).

5. Dans le cas de $\mathbb{R}$, on a principalement travaillé sur des intervalles (ouverts ou non), mais cela est possible car il y a la notion de gauche et droite dans $\mathbb{R}$.

1.1 Définition

Définition-Proposition 1.1. Soit $f : \Omega \to F$ une application, et $x_0 \in \Omega$. On dit que f est différentiable en x_0 s'il existe $L \in \mathcal{L}(E, F)$ une application linéaire continue telle que

$$f(x_0 + h) = f(x_0) + L(h) + \underset{h \to 0}{o}(\|h\|_E). \tag{1.1}$$

Dans ce cas, l'application L est unique, et on l'appelle différentielle de f en x_0, et on la note $df(x_0)$[6].

Enfin on dira que f est différentiable sur Ω si elle est différentiable en tout point $x_0 \in \Omega$, et on appelle $df : \Omega \to \mathcal{L}(E, F)$ la différentielle[7] de f.

Démonstration. Soit L_1 et L_2 des éléments de $\mathcal{L}(E, F)$ pour lesquels (1.1) est satisfaite. Alors pour $\varepsilon > 0$, il existe $\delta > 0$ tel que

$$\forall h \in B(0, \delta), \quad \|L_1(h) - L_2(h)\|_F \leq \varepsilon \|h\|_E,$$

ce qui donne (en considérant $\frac{\delta h}{2\|h\|_E}$ pour $h \in E \setminus \{0\}$) $\|L_1 - L_2\|_{\mathcal{L}(E,F)} \leq \varepsilon$, d'où le résultat en faisant tendre ε vers 0. □

Remarque 1.2. L'équation (1.1) a un sens car pour h proche de 0, on a bien $x_0 + h \in \Omega$ puisque Ω est supposé ouvert. On donne trois versions équivalentes de cette équation :

1. on a la limite
$$\lim_{\|h\|_E \to 0, h \neq 0} \frac{\|f(x_0 + h) - f(x_0) - L(h)\|_F}{\|h\|_E} = 0,$$
2. ce qui se quantifie par :
$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \ \forall h \in B(0, \delta),\ \ \|f(x_0 + h) - f(x_0) - L(h)\|_F \leq \varepsilon \|h\|_E,$$
3. et se reformule aussi par : il existe ε une fonction définie sur un voisinage de 0 dans E et à valeurs dans F telle que
$$f(x_0 + h) = f(x_0) + L(h) + (\|h\|_E)\varepsilon(h), \qquad \text{avec} \quad \lim_{h \to 0} \|\varepsilon(h)\|_F = 0.$$

Remarque 1.3. Le programme est explicitement restreint au cas où les espaces de départ et d'arrivée E et F sont de dimension finie, disons $\mathbb{R}^n$ et $\mathbb{R}^p$ avec $(p, q) \in (\mathbb{N}^*)^2$. Dans ce cas, on peut seulement demander à l'application L d'être linéaire, puisqu'on a vu au Chapitre III que toute application linéaire définie sur un espace de dimension finie est continue.

On a préféré mettre le cas général, parce qu'il nous semble que ça ne coûte pas grand chose, et aussi parce qu'il faut parfois éviter de travailler en coordonnées, même quand les espaces sont de dimension finie : c'est le cas par exemple pour certaines fonctions définies sur l'espace de matrices $\mathcal{M}_n(\mathbb{R})$.

Remarque 1.4. La définition 1.1 est parfois appelée différentielle de Fréchet (dans cette remaque, on utilisera l'expression "Fréchet-différentiable", mais on l'omettra par la suite), et c'est la "bonne" notion au sens où les propriétés attendues (vues dans le cas de la dérivation au Chapitre I) seront satisfaites, comme on le verra dans la suite de ce paragraphe. Abordons deux autres notions néanmoins importantes : soit $f : \Omega \to F$ et $x_0 \in \Omega$:

1. Pour $h \in E$, on dit que f admet une dérivée directionnelle en x_0 dans la direction h si[8]
$$\lim_{t \in \mathbb{R}^*, t \to 0} \frac{f(x_0 + th) - f(x_0)}{t} \quad \text{existe dans } F. \tag{1.2}$$
Dans ce cas, on peut par exemple noter $f'(x_0; h)$ la valeur de cette limite.

6. Il existe de nombreuses notations : $Df(x_0)$, ou df_{x_0} ou $f'(x_0)$. Elles sont toutes légitimes, mais sauf si vous êtes vraiment à l'aise, je vous déconseille d'utiliser la dernière car vous pourriez être tenté de la confondre avec la notion de dérivée.

7. Parfois aussi appelée "application tangente".

8. Attention au choix fait ici : on n'impose pas de signe pour t dans la limite. Certains auteurs demanderont $t > 0$ pour différentier la direction h et $-h$, alors qu'avec notre définition, si $f'(x_0, h)$ existe alors nécessairement $f'(x_0; -h)$ existe et vaut $-f'(x_0; h)$. Avec notre convention, si on voulait évoquer le cas des limites pour lesquelles $t > 0$ et $t < 0$, on parlerait de dérivabilité directionnelle à droite ou à gauche dans la direction h.

2. Si f admet une dérivée directionnelle en x_0 dans toute direction $h \in E$, et que l'application $h \mapsto f'(x_0;h)$ est linéaire [9] et continue [10], on dit que f est Gateaux-différentiable en x_0.

Il est facile de voir [11] que si f est Fréchet-différentiable en x_0 (au sens de la définition 1.1 donc), alors elle admet une dérivée directionnelle en x_0 dans toute direction $h \in E$, et que $\forall h \in E$, $df(x_0)(h) = f'(x_0;h)$. Ainsi, elle est donc Gateaux-différentiable en x_0, puisque $df(x_0)$ est linéaire continue par définition.

Ces notions sont donc moins fortes que la différentiabilité au sens de Fréchet. Néanmoins la formule (1.2) peut être très utile car elle se ramène au cas d'une limite avec un paramètre réel (ce qu'on a l'habitude de manipuler) et elle permet d'identifier le candidat L dans la définition 1.1, qui n'est pas toujours facile à identifier. Il arrive donc souvent qu'on procède de la façon suivante pour montrer la différentiabilité d'une fonction :

1. on calcule $f'(x_0;h)$ pour tout $h \in E$, ce qui montre la dérivabilité directionnelle dans toute direction,
2. on montre que $h \mapsto f'(x_0;h)$ est linéaire et continue,
3. on montre que
$$\lim_{\|h\|_E \to 0, h \neq 0} \frac{\|f(x_0+h) - f(x_0) - f'(x_0;h)\|_F}{\|h\|_E} = 0,$$
ce qui prouve que f est Fréchet-différentiable et que $df(x_0) = f'(x_0;\cdot)$.

Cela peut aussi bien sûr servir à montrer qu'une application n'est pas différentiable, par exemple en montrant qu'elle n'est pas directionnellement dérivable dans une direction bien choisie.

Attention néanmoins, ces notions ne sont pas équivalentes à la Fréchet-différentiabilité d'une part, et d'autre part elle ne sont pas de "bonnes" notions au sens où elles n'ont pas de bonnes propriétés [12] : typiquement, il n'est pas vrai qu'une fonction ayant une dérivée directionnelle en x_0 dans toutes les directions $h \in E$ (et même Gateaux-différentiable en x_0) est continue en ce point, comme le montre le contre-exemple suivant :

$$\begin{array}{ccl} \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x,y) & \longmapsto & \begin{cases} 1 & \text{si } y = x^2 \text{ et } (x,y) \neq (0,0) \\ 0 & \text{sinon.} \end{cases} \end{array}$$

pour laquelle on vérifie facilement que $f'((0,0);h) = 0$ pour tout $h \in \mathbb{R}^2$, mais qui n'est pas continue (et donc par Fréchet-différentiable par la proposition 1.12) en $(0,0)$.

Remarque 1.5. Faites bien attention, et ce sera fondamental dans tout le chapitre, à ne pas confondre les objets
$$df : \Omega \to \mathcal{L}(E,F) \tag{1.3}$$
et
$$df(x_0) : E \to F. \tag{1.4}$$
Notamment, la continuité de l'une ou l'autre de ces applications n'a rien à voir : la continuité de (1.4) (en tant que fonction (linéaire) dont la variable est la direction $h \in E$) est requise dans la définition 1.1 (et automatique si $\dim(E) < +\infty$), alors que si f est telle que l'application (1.3) est continue (en tant que fonction dont la variable est le point $x_0 \in \Omega$), on obtient la définition suivante :

Définition 1.6. Soit $f : \Omega \to F$. On dit que f est de classe C^1 sur Ω si f est différentiable sur Ω et si sa différentielle $df : \Omega \to \mathcal{L}(E,F)$ est continue.

Notons que cette définition a un sens car $\mathcal{L}(E,F)$ est un espace vectoriel normé.

9. Remarquons qu'on a automatiquement avec la définition donnée dans (1.2) la formule $f'(x_0;\lambda h) = \lambda f'(x_0;h)$ pour tout $h \in E$ et tout $\lambda \in \mathbb{R}$; donc pour avoir la linéarité il faut demander en plus $f'(x_0;h+k) = f'(x_0;h) + f'(x_0;k)$ pour tout $(h,k) \in E^2$.

10. Encore une fois, si E est de dimension finie, il suffit de demander la linéarité.

11. Faites-le ! Voir aussi la proposition 1.20.

12. Notons néanmoins qu'elles peuvent s'avérer être de "bonnes" notions au sens où elles sont satisfaites par plus de fonctions, notamment quand E est de dimension infinie.

1.2 Exemples

Exemple 1.7 (Constante)**.** Soit $f : \Omega \to F$, et $x_0 \in F$. Si f est constante au voisinage de x_0, alors il existe $\delta > 0$ tel que

$$\forall h \in B(0,\delta), f(x_0+h) = f(x_0) + 0 + 0$$

et comme l'application $h \mapsto 0$ est linéaire continue (de E dans F), f est différentiable en x_0 de $df(x_0) = 0$[13].

Exemple 1.8 (Cas où $E = \mathbb{R}$)**.** Dans le cas où $f : \mathbb{R} \to F$, comme on l'a suggéré en introduction, on peut définir la dérivée[14] en $x_0 \in \mathbb{R}$ comme la limite

$$f'(x_0) := \lim_{t\to 0} \frac{f(x_0+t) - f(x_0)}{t}$$

si elle existe dans F, auquel cas $f'(x_0) \in F$. Dit autrement, il s'agit de la dérivée directionnelle dans la direction $1 \in \mathbb{R}$, qui est en fait la "seule direction" puisque $\mathbb{R}$ est de dimension 1. Il est alors facile de constater

$$f \text{ est dérivable en } x_0 \iff f \text{ est différentiable en } x_0$$

et dans ce cas[15]

$$\forall h \in \mathbb{R}, \quad df(x_0)(h) = f'(x_0).h$$

où "." désigne la multiplication entre un scalaire et un élément de F (c'est la structure d'espace vectoriel). En fait se cache ici une identification naturelle entre les applications linéaires de $\mathbb{R}$ dans F et les éléments de F, via l'isomorphisme :

$$\begin{array}{ccc} \mathcal{L}(\mathbb{R},F) & \longrightarrow & F \\ L & \longmapsto & L(1) \end{array} \quad \text{ou son inverse} \quad \begin{array}{ccc} F & \longrightarrow & \mathcal{L}(\mathbb{R},F) \\ y & \longmapsto & \left\{ \begin{array}{l} \mathbb{R} \to F \\ \lambda \mapsto \lambda.y \end{array} \right. \end{array}. \tag{1.5}$$

Lorsque $F = \mathbb{R}^p$ est de dimension finie, l'identification précédente consiste à identifier $\mathcal{L}(\mathbb{R},\mathbb{R}^p)$ (qui se représente dans une base par une matrice à une colonne et p lignes) à un vecteur de $\mathbb{R}^p$, ce qui correspond à la représentation habituelle des vecteurs comme des vecteurs colonnes.

Exemple 1.9. L'exemple qui suit est souvent mal compris : on retient parfois qu'une application linéaire continue est différentiable, et que "sa différentielle est elle-même". En effet, si $L \in \mathcal{L}(E,F)$, alors

$$\forall x_0 \in E, \ \ \forall h \in E, \ \ L(x_0+h) = L(x_0) + L(h) + 0$$

donc L est différentiable en tout $x_0 \in E$, et $dL(x_0)(h) = L(h)$ pour tout $h \in E$, ce qui signife

$$dL(x_0) = L \text{ pour tout } x_0 \in E.$$

Ceci n'est pas tout-à-fait la même chose que de dire que la différentielle de L est L, mais plus précisément que la valeur de la différentielle de L en x_0 est l'application L elle-même, voir aussi la remarque 1.5.

Exercice 1.10 (Applications bilinéaires)**.** Soient (E,F,G) trois espaces vectoriels normés, et $a : E \times F \to G$ une application bilinéaire. On la suppose de plus continue, c'est-à-dire (voir l'exercice 2.44 au Chapitre III) qu'il existe $M > 0$ tel que

$$\forall (x,y) \in E \times F, \ \ \|a(x,y)\|_G \le M \|x\|_E \|y\|_F.$$

13. Ce "0" est l'application linéaire nulle de E dans F. On peut aussi écrire $\forall h \in E$, $df(x_0)(h) = 0$ où "0" est cette fois le vecteur nul de F.

14. Attention ; nous faisons le choix ici, par souci de clarté, de distinguer le contexte des mots "différentielle" et "dérivée" et leurs notations. Mais il arrive dans la littérature que les mots soient considérés synonymes, et même d'utiliser la même notation pour l'un et l'autre (le plus souvent f' par commodité). Comme souvent, on conseille de distinguer les notations, au moins jusqu'à atteindre une aisance suffisante pour ne jamais être amené à mélanger les concepts. Une fois cette aisance acquise, on se rend compte qu'il est assez commode d'unifier les notations.

15. On peut lire la formule dans l'autre sens, et retenir $f'(x_0) = df(x_0)(1)$ qui suffit par linéarité à caractériser $df(x_0)$.

1. Montrer que a est différentiable en tout $(x, y) \in E \times F$, et
$$\forall (h,k) \in E \times F, \ \ da(x,y)(h,k) = a(x,k) + a(h,y).$$
2. Si on a $E = F$ et $L \in \mathcal{L}(E, G)$, montrer que l'application
$$\forall x \in E, \quad J(x) = \frac{1}{2}a(x,x) - L(x)$$
est différentiable sur E et calculer sa différentielle.
3. Si on suppose que $E = F = \mathbb{R}^n$ et $G = \mathbb{R}$, $A \in \mathcal{M}_n(\mathbb{R})$ et $b \in \mathbb{R}^n$, montrer que l'application
$$\forall x \in \mathbb{R}^n, \ \ J(x) = \frac{1}{2}x^T Ax - b^T x$$
est différentiable est calcul sa différentielle.

Exercice 1.11 (Applications matricielles)**.** Voir par exemple [Rou03, Exercice 16, 26] Soit $n \in \mathbb{N}^*$. On considère dans cet exercice des fonctions définies sur des ouverts de $\mathcal{M}_n(\mathbb{R})$. Par commodité, on munit $\mathcal{M}_n(\mathbb{R})$ d'une norme $\|\cdot\|$ qui satisfait
$$\forall (A,B) \in \mathcal{M}_n(\mathbb{R}), \ \ \|A.B\| \leq \|A\|.\|B\|.$$
On sait (voir le paragraphe 2.5 au Chapitre III) que les normes subordonnées satisfont cette propriété.

1. On pose $\forall A \in \mathcal{M}_n(\mathbb{R}), \ \ f(A) = A^2$. Montrer que f est différentiable sur $\mathcal{M}_n(\mathbb{R})$, et que
$$\forall H \in \mathcal{M}_n(\mathbb{R}), \ \ df(A)(H) = AH + HA.$$
2. On pose $\forall A \in GL_n(\mathbb{R}), \ \ g(A) = A^{-1}$.
   (a) Montrer que si $H \in \mathcal{M}_n(\mathbb{R})$ et $\|H\| < 1$, alors la matrice $I_n - H$ est inversible et
$$(I_n - H)^{-1} = \sum_{k \in \mathbb{N}} H^k.$$
   (b) En déduire que $GL_n(\mathbb{R})$ est un ouvert de $\mathcal{M}_n(\mathbb{R})$ [16], et que la différentielle de g en $A \in GL_n(\mathbb{R})$ est
$$dg(A)(H) = -A^{-1}HA^{-1}$$
3. On pose $\forall A \in \mathcal{M}_n(\mathbb{R}), \ \ h(A) = \det(A).$ [17]
   (a) Montrer que h est différentiable en I_n et
$$\forall H \in \mathcal{M}_n(\mathbb{R}), \ \ h'(I_n)(H) = \mathrm{tr}(H).$$
   (b) Montrer que si $A \in GL_n(\mathbb{R})$ alors h est différentiable en A et
$$\forall H \in \mathcal{M}_n(\mathbb{R}), \ \ h'(A)(H) = \det(A)\mathrm{tr}(A^{-1}H) = \mathrm{tr}(\mathrm{Com}(A)^T H).$$
   (c) Montrer que h est de classe C^1 sur $\mathcal{M}_n(\mathbb{R})$ [18], et en déduire que
$$\forall A \in \mathcal{M}_n(\mathbb{R}), \ \ \forall H \in \mathcal{M}_n(\mathbb{R}), \quad dh(A)(H) = \mathrm{tr}(\mathrm{Com}(A)^T H).$$

16. Rappelons qu'on peut aussi montrer cela avec la notion de déterminant, mais la preuve proposée ici est valable dans le cadre plus général des algèbres de Banach, voir aussi l'exercice 2.48 au Chapitre III.
17. Notons qu'il existe plusieurs méthodes pour calculer la différentielle du déterminant ; on peut aussi faire un calcul direct en coordonnées en prenant la base canonique $(E_{ij})_{1 \leq i,j \leq n}$ de $\mathcal{M}_n(\mathbb{R})$.
18. On aperçoit ici une problématique classique du calcul différentiel : montrer qu'une fonction est régulière (ici on peut en fait montrer que det est de classe C^∞) et calculer sa différentielle (et ses différentielles d'ordre supérieur) peuvent être deux choses très différentes. Dans la méthode ici proposée, on arrive à montrer que det est différentiable en revenant à la définition, mais seulement sur $GL_n(\mathbb{R})$ (question 3b), alors qu'à la question 3c, on démontre sans calcul que det est une fonction régulière (afin d'étendre le calcul de la question 3b à $\mathcal{M}_n(\mathbb{R})$ par un argument de densité, mais ce n'est pas le propos de la remarque), et on aurait pu le faire dès le début de l'exercice. Prenez donc conscience que de "calculer une différentielle" ou "montrer que la différentielle existe" peuvent soit être faits simultanément (si on revient à la définition par exemple), ou alors être deux raisonnements indépendants, suivant les situations. Sachant qu'il n'est pas rare qu'on ait besoin de l'existence de la différentielle, mais pas de son expression.

1.3 Premières propriétés

Proposition 1.12. *Soit $f : \Omega \to F$ et $x_0 \in \Omega$. Si f est différentiable en x_0 alors f est continue en x_0.*

Démonstration. On peut écrire

$$f(x_0 + h) = f(x_0) + df(x_0)(h) + o(\|h\|)$$

donc en faisant tendre h vers 0, en utilisant la continuité de $df(x_0)$, on obtient bien

$$\lim_{h\to 0} f(x_0 + h) = f(x_0).$$

□

Proposition 1.13. *Soit $x_0 \in \Omega$.*

1. *Soient $f, g : \Omega \to F$ deux fonctions différentiables en $x_0 \in \Omega$, et $\alpha \in \mathbb{K}$. Alors $f + g$ et αf sont différentiables en x_0 et*
$$d(f + g)(x_0) = df(x_0) + dg(x_0), \qquad d(\alpha f)(x_0) = \alpha df(x_0).$$
2. *Soit $f, g : \Omega \to \mathbb{R}$ deux fonctions différentiables en $x_0 \in \Omega$. Alors $f.g$ est différentiable en x_0 et*
$$d(f.g)(x_0) = f(x_0)dg(x_0) + g(x_0)df(x_0).$$
3. *Si $F = F_1 \times \ldots \times F_p$ est un produit d'espaces vectoriels normés[19], $f : \Omega \to F$, alors*
$$f \text{ est différentiable en } x_0 \iff \forall i \in [\![1, p]\!],\ f_i \text{ est différentiable en } x_0$$
et dans ce cas
$$\forall h \in E, \quad df(x_0)(h) = \big(df_1(x_0)(h), \ldots, df_p(x_0)(h)\big)$$
4. *Soit $f : \Omega \to F$ et $g : \Omega' \to G$ où Ω' est un ouvert de F et G est un espace vectoriel normé. On suppose que $f(\Omega) \subset \Omega'$. Si f est différentiable en x_0 et g est différentiable en $f(x_0)$, alors $g \circ f$ est différentiable en x_0 et*
$$d(g \circ f)(x_0) = dg(f(x_0)) \circ df(x_0). \tag{1.6}$$

Remarque 1.14. On conseille de passer un certain temps à se familiariser avec la formule (1.6), notamment de bien cerner la difficulté des objets en comparaison avec la formule habituelle $(g\circ f)'(x_0) = g'(f(x_0)f'(x_0)$ (voir la proposition 3.38 au Chapitre I) dans le cadre des fonctions réelles. Quelques conseils :

— Avec l'exemple 1.8, comprenez pourquoi la formule (1.6) généralise bien la dérivée d'une composée dans le cas réel.

— Comprenez pourquoi on ne peut pas écrire la formule dans l'autre sens, contrairement au cas réel où on pouvait écrire $f'(x_0)g'(f(x_0))$ du fait de la commutativité du produit réel.

— Comprenez que l'objet intervenant dans (1.6) a bien un sens car on compose des fonctions dont les espaces sont compatibles. D'ailleurs, le meilleur moyen d'éviter de nombreuses erreurs de calcul différentiel consiste à faire attention à ce que toutes les formules aient un sens.

— Si besoin, évitez les raccourcis et écrivez toujours la valeur en $h \in E$ d'une différentielle. Ainsi (1.6) signifie en fait

$$\forall h \in E, \ \ d(g \circ f)(x_0)(h) = dg(f(x_0))\Big(df(x_0)(h)\Big).$$

— Appliquez cette formule sur quelques exemples, voir l'exercice 1.16.

19. Donc F est muni de 'la' norme produit, voir la remarque 2.10 au Chapitre III.

Remarque 1.15. Ne perdons pas de vue que dans la proposition 1.13, on a deux types d'informations, la première étant que les opérations préservent le caractère différentiable des fonctions mises en jeu, la seconde donnant un formulaire. Il arrive qu'on utilise seulement le premier aspect, soit parce que c'est tout ce dont on a besoin, soit parce que le formulaire n'est pas toujours le moyen de plus commode d'arriver au résultat, et qu'il est parfois préférable de revenir à la définition. Voir aussi la note 18 et l'exercice 1.16.

Démonstration. 1. La première propriété s'obtient facilement en additionnant/multipliant par α les définitions de différentiabilité pour f et g.

2. La seconde s'obtient par un calcul similaire à la preuve de la même formule dans le cas réel, voir la proposition 3.37 au Chapitre I. Une autre méthode [20] consiste à considérer l'application $a : (x, y) \in \mathbb{R} \times \mathbb{R} \mapsto x.y \in \mathbb{R}$ qui est bilinéaire, et donc $f.g(x) = a(f(x), g(x)) = a \circ \Phi(x)$ où $\Phi : x \in \Omega \mapsto (f(x), g(x)) \in \mathbb{R} \times \mathbb{R}$, et on peut retrouver le résultat en appliquant l'exercice 1.10 et les résultats 3 et 4 de la proposition 1.13 démontrés ci-après.
3. Cette propriété suit des définitions et du fait qu'une application $L : E \to F_1 \times \dots F_p$ est linéaire continue si et seulement si ses fonctions composantes $(L_i)_{i\in[\![1,p]\!]}$ sont linéaires continues.
4. Par définition, il existe $\delta_1 > 0$, $\delta_2 > 0$, $\varepsilon_1 : B(0, \delta_1) \to F$ et $\varepsilon_2 : B(0, \delta_2) \to G$ telles que

$$\forall h \in B(0, \delta_1),\ f(x_0 + h) = f(x_0) + df(x_0)(h) + \|h\|\varepsilon_1(h), \quad \text{et} \quad \lim_{h\to 0} \varepsilon_1(h) = 0$$

$$\forall s \in B(0, \delta_2),\ g(f(x_0) + s) = g(f(x_0)) + dg(f(x_0))(s) + \|s\|\varepsilon_2(s), \quad \text{et} \quad \lim_{s\to 0} \varepsilon_2(s) = 0.$$

Par la proposition 1.12, il existe δ_3 tel que

$$\forall h \in B(0, \delta_3), \|f(x_0 + h) - f(x_0)\| < \delta_2$$

et on pose $\delta_1' = \min\{\delta_1, \delta_3\}$. Ainsi, en posant $s = f(x_0 + h) - f(x_0) = df(x_0)(h) + \|h\|\varepsilon_1(h)$, on obtient :

$$\begin{aligned} \forall h \in B(0, \delta_1'),\ g(f(x_0 + h)) &= g\Big(f(x_0) + \underbrace{df(x_0)(h) + \|h\|\varepsilon_1(h)}_{\in B(0,\delta_2)}\Big) \\ &= g(f(x_0)) + dg(f(x_0))(s) + \|s\|\varepsilon_2(s) \\ &= g(f(x_0)) + dg(f(x_0))(df(x_0)(h)) + \|h\|\varepsilon_3(h) \end{aligned}$$

où on a posé

$$\forall h \in B(0, \delta_1') \setminus \{0\},\ \ \varepsilon_3(h) = dg(f(x_0))(\varepsilon_1(h)) + \left\| df(x_0)\Big(\frac{h}{\|h\|}\Big) + \varepsilon_1(h) \right\| \varepsilon_2\Big(df(x_0)(h) + \|h\|\varepsilon_1(h)\Big)$$

qu'on peut prolonger par 0 quand $h = 0$, et est bien telle que $\lim_{h\to 0} \varepsilon_3(h) = 0$.

□

Exercice 1.16. On propose ici quelques calculs de différentielles. On suggère de faire systématiquement deux preuves, la première étant de revenir à la définition, la seconde étant de décomposer les fonctions via les opérations élémentaires (somme, composée etc) et d'appliquer la proposition 1.13.

1. Calculez la différentielle de la fonction : $\forall A \in \mathcal{M}_n(\mathbb{R}),\ \ f(A) = \operatorname{tr}(A^2)$.
2. Calculez la différentielle de la fonction : $\forall A \in \mathcal{M}_n(\mathbb{R}),\ \ g(A) = \operatorname{tr}(A)^2$.
3. On se place dans $(H, \langle\cdot, \cdot\rangle)$ un espace préhilbertien réel, c'est-à-dire un $\mathbb{R}$-espace vectoriel muni d'un produit scalaire.
   (a) Montrer que la fonction $x \in H \mapsto \|x\|^2 = \langle x, x\rangle$ est différentiable sur H et calculez sa différentielle.
   (b) Montrer que la fonction $x \in H \mapsto \|x\|$ est différentiable sur $H \setminus \{0\}$ et calculez sa différentielle. Montrer qu'elle n'est pas différentiable en 0.

20. Instructive à mettre en œuvre !

Exercice 1.17 (Interversion limite et différentielle)**.** Voir aussi [Rou03, Exercice 39] ; comparer à la proposition 4.13 du Chapitre I.

1. On suppose E, F deux espaces vectoriels normés, Ω un ouvert de E, et $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions de Ω dans F, qu'on suppose différentiables sur Ω. On suppose également
   — la suite $(f_n)_{n\in\mathbb{N}}$ converge simplement vers f sur Ω.
   — la suite des différentielles $(df_n)_{n\in\mathbb{N}}$ converge uniformément sur Ω.
   (a) Montrer que f est différentiable sur Ω et que
   $$\forall x \in \Omega, \quad df(x) = \lim_{n\to\infty} df_n(x).$$
   (b) En déduire que si les $(f_n)_{n\in\mathbb{N}}$ sont de classe C^1, alors f également.
2. On suppose $E = F = \mathcal{M}_n(\mathbb{R})$, et on considère
   $$\forall A \in \mathcal{M}_n(\mathbb{R}), \ \ \exp(A) = \sum_{k=0}^{+\infty} \frac{A^k}{k!}.$$
   (a) Justifier que la série précédente est convergente.
   (b) Montrer que l'application $\exp : \mathcal{M}_n(\mathbb{R}) \to \mathcal{M}_n(\mathbb{R})$ est de classe C^1.

1.4 Dérivées partielles

Nous commençons ce paragraphe par le cas de la dimension finie, en supposant que E est un produit de copies de $\mathbb{R}$, de sorte que $E = \mathbb{R}^n$. En effet, dans la pratique, E sera très souvent $\mathbb{R}^n$ ou $\mathcal{M}_n(\mathbb{R})$, ce dernier étant vu comme $\mathbb{R}^{n^2}$; on travaillera avec la base canonique de ces espaces.

Néanmoins, afin de clarifier le théorème des fonctions implicites, nous parlerons ensuite de différentielles partielles dans le cas où E s'écrit comme produit d'espace vectoriel normés $E = E_1 \times \ldots \times E_n$.

1.4.1 Cas de la dimension finie, écriture de la différentielle en coordonnées

Cette partie est en quelque sorte un cas particulier du paragraphe suivant, plus simple à comprendre et qui intervient dans la majorité des applications ; la simplification vient de l'exemple 1.8, ce qui permet de parler de "dérivée partielle", là où on parlera de "différentielle partielle" au paragraphe 1.4.2.

Soit donc $n \geq 2$; on munit $\mathbb{R}^n$ de la norme $\|\cdot\|_\infty$ par exemple, et on pose $(e_i)_{i\in[\![1,n]\!]}$ la base canonique de cet espace, c'est-à-dire $\forall i \in [\![1,n]\!],\ e_i = (0,\ldots,0,1,0,\ldots,0)$ où le 1 est en i-ième position.
On considère Ω un ouvert de $\mathbb{R}^n$, et F un espace vectoriel normé.

Définition 1.18. Soit $f : \Omega \to F$, $x = (x_1,\ldots,x_n) \in \Omega$, et $i \in [\![1,n]\!]$. On dit que f admet une dérivée partielle en la i-ième variable en x si

$$y \in \mathbb{R} \longmapsto f(x_1,\ldots,x_{i-1},y,x_{i+1},\ldots,x_n) \tag{1.7}$$

est dérivable en x_i. Dans ce cas, on note[21]

$$\partial_i f(x) \quad \text{ou} \quad \frac{\partial f}{\partial x_i}(x) \quad \text{ou} \quad \partial_{x_i} f(x) \quad \text{ou} \quad D_i f(x)$$

la valeur de la dérivée de la fonction définie en (1.7).

21. Pédagogiquement, on conseille d'utiliser préférablement les notations $\partial_i f$ ou $D_i f$ qui font référence à l'indice de la variable et non à son nom. En effet, le nom des variables est susceptible de changer, et même si la notation $\frac{\partial f}{\partial x_i}(x)$ est certainement la plus utilisée, elle apporte son lot d'ambiguïté. À titre d'exemple, les fonctions $f : (x,y) \in \mathbb{R}^2 \mapsto x - y$ et $f : (y,x) \in \mathbb{R}^2 \mapsto y - x$ sont a priori rigoureusement identique par le principe des variables muettes, mais $\frac{\partial f}{\partial x}$ aura probablement un résultat différent dans les deux cas. Alors que $\partial_1 f$ devrait fournir le même résultat car on fait référence à la position de la variable et non à son nom, qui peut être amené à changer. Si vous avez des difficultés avec les notations dans ce chapitre, on conseille la lecture de ce lien.

Remarque 1.19. Autrement dit,

$$\partial_i f(x) = \lim_{h\to 0, h\neq 0} \frac{f(x_1,\dots,x_{i-1},x_i+h,x_{i+1},\dots,x_n) - f(x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_n)}{h}$$

quand cette limite existe. C'est un élément de F (voir l'exemple 1.8).

Remarquons également qu'il s'agit en fait de la dérivée directionnelle de f dans la direction e_i, qu'on avait noté $f'(x; e_i)$.

L'énoncé suivant explique que si une fonction est différentiable, alors elle admet des dérivées partielles :

Proposition 1.20. *Soit $x = (x_1,\dots,x_n) \in \Omega$. Si $f : \Omega \to F$ est différentiable en x, alors f admet des dérivées partielles en x en toutes les variables, et*

$$\forall i \in [\![1,n]\!],\ \ \partial_i f(x) = df(x)(e_i).$$

En conséquence si $h = (h_1,\dots,h_n) \in \mathbb{R}^n$ alors

$$df(x)(h) = \sum_{i=1}^{n} \partial_i f(x) h_i. \tag{1.8}$$

Remarque 1.21. Attention, la réciproque est fausse, c'est-à-dire qu'une fonction peut admettre des dérivées partielles en toutes les variables en un point x, mais ne pas être différentiable en ce point. On peut recycler le contre-exemple donné à la remarque 1.4 puisque ce dernier admettait des dérivées directionnelles dans toutes les directions, et donc en particulier dans les directions de la base canoniques, c'est-à-dire des dérivées partielles selon toutes les coordonnées.

Démonstration. Soit $i \in [\![1,n]\!]$. Comme Ω est ouvert, il existe $\delta > 0$ tel que $B(x,\delta) \subset \Omega$. Alors pour tout $t \in]-\delta,\delta[$, par différentiabilité de f en x

$$f(x+te_i) = f(x) + df(x)(te_i) + o(\|te_i\|) = f(x) + t df(x)(e_i) + \underset{t\to 0}{o}(t),$$

où on a utilisé la linéarité de $df(x)$. On en déduit que f admet une dérivée partielle en la i-ième variable, et $\partial_i f(x) = df(x)(e_i)$. Par linéarité, si $h \in \mathbb{R}^n$, alors $h = \sum_{i=1}^{n} h_i e_i$ et donc

$$df(x)(h) = \sum_{i=1}^{n} h_i df(x)(e_i) = \sum_{i=1}^{n} \partial_i f(x) h_i.$$

□

Remarque 1.22 (Jacobienne)**.** Supposons que $F = \mathbb{R}^p$, et que $f = (f_1,\dots,f_p) : \Omega \to F$ est différentiable en x. On travaille dans les bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^p$. Alors $df(x) \in \mathcal{L}(\mathbb{R}^n,\mathbb{R}^p)$ est une application linéaire, et peut être représentée par une matrice. Comme d'habitude, un vecteur de $\mathbb{R}^n$ est vu comme un vecteur colonne, et la formule (1.8) se réécrit :

$$\forall i \in [\![1,p]\!],\ \forall h \in \mathbb{R}^n,\ \ df_i(x)(h) = \sum_{k=1}^{n} \partial_k f_i(x) h_k.$$

On reconnaît un produit matrice/vecteur, et donc on est amené à définir

$$J_f(x) := \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & \vdots & \vdots \\ \partial_1 f_p(x) & \cdots & \partial_n f_p(x) \end{pmatrix} = \Big(\partial_j f_i(x)\Big)_{1\le i\le p, 1\le j\le n}$$

qu'on appelle Jacobienne de f en x, et on a

$$df(x)(h) = \begin{pmatrix} df_1(x)(h) \\ \vdots \\ df_p(x)(h) \end{pmatrix} = \begin{pmatrix} \partial_1 f_1(x) & \cdots & \partial_n f_1(x) \\ \vdots & \vdots & \vdots \\ \partial_1 f_p(x) & \cdots & \partial_n f_p(x) \end{pmatrix} \times \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = J_f(x) \times h.$$

Remarquons que les résultats de la proposition 1.13 se transposent bien avec le calcul matriciel, plus précisément, si $f, g : \Omega \to \mathbb{R}^p$ sont différentiables en x et $\alpha \in \mathbb{R}$, alors

$$J_{f+\alpha g}(x) = J_f(x) + \alpha J_g(x)$$

et si $f : \Omega \to \mathbb{R}^p$ et $g : \Omega' \to \mathbb{R}^q$, avec Ω' un ouvert de $\mathbb{R}^p$ et $f(\Omega) \subset \Omega'$, sont respectivement différentiable en x et $f(x)$, alors

$$J_{g\circ f}(x) = J_g(f(x)) \times J_f(x)$$

puisque la composition des applications linéaires se traduit par un produit des matrices représentatives de ces applications dans des bases compatibles.

Remarque 1.23 (Gradient)**.** On suppose désormais que $F = \mathbb{R}$. Dans ce cas la remarque 1.22 s'applique et $J_f(x)$ est un vecteur ligne, mais on peut aussi interpréter (1.8) via le produit scalaire usuel $\langle \cdot, \cdot \rangle$ sur $\mathbb{R}^n$: en effet si on pose

$$\nabla f(x) = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{pmatrix} = J_f(x)^T$$

qu'on appelle gradient de f en x, alors on a

$$\forall h \in \mathbb{R}^n, \quad df(x)(h) = \langle \nabla f(x), h \rangle.$$

Ce résultat peut être interprété comme un résultat de dualité : comme $df(x) \in \mathcal{L}(E, \mathbb{R}) = E'$ et que E est de dimension finie, il existe un unique vecteur, à savoir $\nabla f(x)$ tel que $df(x)$ est en fait le produit scalaire avec ce vecteur, c'est-à-dire $df(x) = \langle \nabla f(x), \cdot \rangle$.

*Plus généralement, on peut donc parler de gradient si E est un espace de Hilbert, car alors le théorème de Riesz (voir le Chapitre IX) affirme que toute forme linéaire continue sur E est représenté de façon unique comme le produit scalaire avec un vecteur de E, qu'on appellera gradient de f en x et qu'on notera $\nabla f(x)$.

Exercice 1.24. Dans cet exercice, nous allons retrouver la différentielle du déterminant en travaillant en coordonnées. On pose $\forall A \in \mathcal{M}_n(\mathbb{R}), \quad f(A) = \det(A)$, et $(E_{ij})_{1\leq i,j\leq n}$ la base canonique de $\mathcal{M}_n(\mathbb{R})$.

1. Justifier que f est différentiable sur $\mathcal{M}_n(\mathbb{R})$.
2. Pour $A \in \mathcal{M}_n(\mathbb{R})$ et $(i,j) \in [\![1,n]\!]^2$, calculer

$$\lim_{t\to 0, t\neq 0} \frac{f(A + tE_{ij}) - f(A)}{t}.$$

3. En déduire que pour tout $A \in \mathcal{M}_n(\mathbb{R})$,

$$\forall H \in \mathcal{M}_n(\mathbb{R}), \quad df(A)(H) = \mathrm{tr}(\mathrm{Com}(A)^T H).$$

1.4.2 Différentielles partielles

Dans ce paragraphe, nous allons généraliser à un cadre un peu plus abstrait le paragraphe précédent ; la motivation principale est le théorème des fonctions implicites [22] ; on suppose que $E = E_1 \times \ldots \times E_n$ où $E_1, \ldots, E_n$ sont des espaces vectoriels normés, et E est muni de la norme produit. Ω désigne encore un ouvert de E.

22. La motivation n'est pas vraiment de manipuler des espaces de dimension infinie, mais le fait de voir $\mathbb{R}^n$ non pas comme n copies de $\mathbb{R}$, mais comme $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ avec $n_1 + n_2 = n$. Que le lecteur ne se décourage pas avec les premières définitions ; s'il a du mal à jongler avec les concepts ici présents, on lui conseille de se concentrer sur l'interprétation de ceux-ci en dimension finie, ce que nous décrivons à la remarque 1.28 et où nous allons seulement extraire des sous-matrices de la jacobienne.

Définition 1.25. Soit $f : \Omega \to F$, $x = (x_1, \dots, x_n) \in \Omega$, et $i \in [\![1, n]\!]$. On dit que f admet une différentielle partielle en la i-ième variable en x si

$$y \in \; E_i \longmapsto f(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n) \tag{1.9}$$

est différentiable en x_i. Dans ce cas, on note

$$\partial_i f(x) \quad \text{ou} \quad \frac{\partial f}{\partial x_i}(x) \quad \text{ou} \quad \partial_{x_i} f(x) \quad \text{ou} \quad D_i f(x)$$

la valeur de la différentielle de la fonction définie en (1.9), qui est un élément de $\mathcal{L}(E_i, F)$.

Remarque 1.26. Si pour $i \in [\![1, n]\!]$ l'espace E_i est de dimension 1, c'est-à-dire $E_i = \mathbb{R}$, alors on se retrouve dans le cadre de l'exemple 1.8 et alors la notion de dérivée partielle en la i-ième variable et de différentielle partielle en la i-ième variable peuvent être confondus via les isomorphismes (1.5). C'est ce qui justifie le fait qu'on a utilisé les mêmes notations.

Remarque 1.27. On peut constater facilement que :

1. si f admet une différentielle partielle en la i-ième variable, alors f admet des dérivées directionnelles en x dans toute direction $(0, \dots, 0, h_i, 0 \dots, 0) \in E$ où $h_i \in E_i$, et alors
$$\partial_i f(x)(h_i) = f'(x; h_i) = \lim_{t \to 0, t \neq 0} \frac{f(x_1, \dots, x_{i-1}, x_i + th_i, x_{i+1}, \dots, x_n) - f(x)}{t}.$$
2. si f est différentiable en x, alors f admet des différentielles partielles en toutes les variables en x, et
$$\forall h \in E, \;\; df(x)(h) = \sum_{i=1}^{n} \partial_i f(x)(h_i).$$

Remarque 1.28 (Cas de la dimension finie)**.** Supposons que $E_1 = \mathbb{R}^{n_1}$, $E_2 = \mathbb{R}^{n_2}$, $E = E_1 \times E_2$ et $F = \mathbb{R}^p$. Si $f : \Omega \to F$ est différentiable en $x \in \Omega$, alors la jacobienne de f s'écrit

$$J_f(x, y) = \Big(\partial_j f_i(x, y)\Big)_{1 \le i \le p, 1 \le j \le n_1 + n_2} = \Big(J_{1,f}(x, y) \quad J_{2,f}(x, y)\Big)$$

où

$$J_{1,f}(x, y) = \Big(\partial_j f_i(x, y)\Big)_{1 \le i \le p, 1 \le j \le n_1} \text{ et } J_{2,f}(x, y) = \Big(\partial_j f_i(x, y)\Big)_{1 \le i \le p, n_1 + 1 \le j \le n_1 + n_2}.$$

Ces matrices sont les représentations dans la base canonique des différentielles partielles $\partial_1 f(x) \in \mathcal{L}(\mathbb{R}^{n_1}, \mathbb{R}^p)$ et $\partial_2 f(x) \in \mathcal{L}(\mathbb{R}^{n_2}, \mathbb{R}^p)$.

1.5 Théorème des accroissements finis

Commençons par remarquer que si $\dim(F) > 1$, l'égalité des accroissements finis (corollaire 3.41 au Chapitre I) n'est pas valable : le contre-exemple classique est

$$\begin{array}{rcl} f : \mathbb{R} & \longrightarrow & \mathbb{C} = \mathbb{R}^2 \\ t & \longmapsto & e^{it} = (\cos(t), \sin(t)) \end{array}$$

car $f(0) = f(2\pi)$ mais il n'existe pas $c \in]0, 2\pi[$ tel que $f'(c) = \frac{f(2\pi) - f(0)}{2\pi - 0}$ puisque $\forall t \in \mathbb{R}, f'(t) = ie^{it}$ ne s'annule pas.

Remarque 1.29. Si par contre $F = \mathbb{R}$, $f : \Omega \to \mathbb{R}$, $a, b \in \Omega$ sont tels que $[a, b] \subset \Omega$, alors on peut considérer

$$\forall t \in [0, 1], \;\; \varphi(t) = f(a + t(b - a))$$

qui est à reparamétrisation près la restriction de f au segment $[a, b] \subset \Omega$, et lui appliquer l'égalité des accroissements finis qui donne l'existence de $\theta \in]0, 1[$ tel que

$$\varphi'(\theta) = \frac{\varphi(1) - \varphi(0)}{1}, \qquad \text{i.e.} \qquad f(b) - f(a) = df(c)(b - a)$$

où $c = a + \theta(b-a) \in \Omega$. Donc à part qu'on abandonne la forme "taux de variation" $f'(c) = \frac{f(b)-f(a)}{b-a}$ qu'on réécrit $f(b) - f(a) = f'(c)(b-a)$, on peut généraliser l'égalité des accroissements au cas d'un espace de départ de dimension supérieure à 1 (en s'assurant que l'ouvert de définition de f contient le segment $[a,b]$), si l'espace d'arrivée est de dimension 1.

Dans le cas où $\dim(F) > 1$ donc, nous allons néanmoins pouvoir généraliser l'inégalité des accroissements finis vue au corollaire 3.43 du Chapitre I :

Théorème 1.30. *Soit Ω un ouvert de E, $(a,b) \in \Omega^2$ tels que $[a,b] \subset \Omega$, et $f : \Omega \to F$ une fonction continue en tout point de $[a,b]$* [23] *et différentiable en tout point de $]a,b[$. Alors*

$$\|f(b) - f(a)\|_F \leq \Big(\sup_{z \in]a,b[} \|df(z)\|_{\mathcal{L}(E,F)}\Big)\|b-a\|_E.$$

Pour la preuve de ce résultat, on va utiliser le lemme suivant, qui a un intérêt en soit :

Lemme 1.31. *Soient $(\alpha,\beta) \in \mathbb{R}^2$, F un espace vectoriel normé, $g : [\alpha,\beta] \to F$ et $\varphi : [\alpha,\beta] \to \mathbb{R}$. On suppose que (g,φ) sont continues sur $[\alpha,\beta]$ et dérivables sur $]\alpha,\beta[$, et que* [24]

$$\forall t \in]\alpha,\beta[, \quad \|g'(t)\|_F \leq \varphi'(t). \tag{1.10}$$

Alors

$$\|g(\beta) - g(\alpha)\|_F \leq \varphi(\beta) - \varphi(\alpha).$$

Remarque 1.32. Le résultat précédent aurait une preuve très simple si on avait supposé les fonctions g et φ de classe C^1 sur $[\alpha,\beta]$, car il aurait suffit d'intégrer [25] (1.10) et d'appliquer le théorème fondamental de l'analyse 5.12 du Chapitre I.

Démonstration. Soit $\varepsilon > 0$. On pose

$$I := \Big\{y \in [\alpha,\beta], \quad \forall x \in [\alpha,y], \|g(x) - g(\alpha)\| \leq \varphi(x) - \varphi(\alpha) + \varepsilon(x-\alpha) + \varepsilon\Big\}.$$

On constate que I est un intervalle tel que $\alpha \in I \subset [\alpha,\beta]$. On veut montrer que $I = [\alpha,\beta]$; en effet, en écrivant alors

$$\|g(\beta) - g(\alpha)\| \leq \varphi(\beta) - \varphi(\alpha) + \varepsilon(\beta-\alpha) + \varepsilon,$$

on pourra faire tendre ε vers 0 et obtenir le résultat.

Posons alors $c = \sup I$. Remarquons d'abord que $c > \alpha$: par continuité de g et φ en α, il existe $\delta > 0$ tel que

$$\forall x \in [\alpha,\alpha+\delta[, \quad \|g(x) - g(\alpha)\| \leq \frac{\varepsilon}{2}, \qquad \text{et} \qquad \varphi(x) - \varphi(\alpha) \geq -\frac{\varepsilon}{2}$$

et donc

$$\forall x \in [\alpha,\alpha+\delta[, \quad \|g(x) - g(\alpha)\| - \big(\varphi(x) - \varphi(\alpha)\big) \leq \varepsilon \leq \varepsilon + \varepsilon(x-\alpha),$$

ce qui montre bien $c > \alpha$.

On constate également que $c \in I$: en effet par définition de c on a

$$\forall x \in [\alpha,c[, \ \|g(x) - g(\alpha)\| \leq \varphi(x) - \varphi(\alpha) + \varepsilon(x-\alpha) + \varepsilon$$

23. Rappelons que $[a,b] = \{(1-t)a + tb, t \in [0,1]\}$, et bien sûr vous aurez deviné que $]a,b[= \{(1-t)a + tb, t \in]0,1[\}$.
24. Voir l'exemple 1.8 pour la définition de $g'(t)$; on aurait aussi pu écrire $\|dg(t)\|_{\mathcal{L}(\mathbb{R},F)}$.
25. Rappelons qu'il n'est pas difficile de définir et d'étudier l'intégrale de Riemann des fonctions à valeurs dans un espace vectoriel normé complet (même si en pratique on se restreint souvent à $F = \mathbb{R}^p$), et qu'on a l'inégalité

$$\left\|\int_\alpha^\beta g(t)dt\right\| \leq \int_\alpha^\beta \|g(t)\|dt.$$

L'intégrale de Lebesgue n'est pas non plus difficile à définir si les fonctions sont à valeurs dans $\mathbb{R}^p$; pour le cas d'un espace vectoriel normé F quelconque, voir par exemple [Wag12a, 2.C]. Mais ici de toute façon, on n'a rien supposé sur g' ou φ', donc en particulier on n'a pas supposé leur mesurabilité, donc la preuve donnée dans cette remarque ne peut fonctionner dans le cas général.

mais par continuité de g et φ, on obtient que l'inégalité est valable pour $x = c$ ce qui montre bien $c \in I$.

Supposons donc par l'absurde que $c \in]\alpha, \beta[$. Par définition de la dérivabilité de g et φ en c, il existe $\delta > 0$ tel que

$$\forall x \in]c, c+\delta], \quad \left\| \frac{g(x)-g(c)}{x-c} - g'(c) \right\| \leq \frac{\varepsilon}{2} \quad \text{et} \quad \left| \frac{\varphi(x)-\varphi(c)}{x-c} - \varphi'(c) \right| \leq \frac{\varepsilon}{2},$$

d'où

$$\forall x \in]c, c+\delta], \;\; \|g(x)-g(c)\| \leq \|g'(c)\|(x-c) + \frac{\varepsilon}{2}(x-c) \quad \text{et} \quad \varphi(x)-\varphi(c) \geq \varphi'(c)(x-c) - \frac{\varepsilon}{2}(x-c).$$

On en déduit

$$\forall x \in]c, c+\delta], \quad \|g(x)-g(c)\| \leq \varphi(x)-\varphi(c) + \big(\underbrace{\|g'(c)\| - \varphi'(c)}_{\leq 0}\big)(x-c) + \varepsilon(x-c).$$

D'autre part, $c \in I$ donc

$$\|g(c)-g(a)\| \leq \varphi(c)-\varphi(a) + \varepsilon(c-a) + \varepsilon$$

et les deux inégalités précédentes impliquent par addition et inégalité triangulaire que $c+\delta \in I$, ce qui est une contraction ; ainsi $c = \beta$, ce qui conclut la preuve de ce lemme. □

Démonstration du théorème 1.30 : On applique le lemme 1.31 aux fonctions

$$\forall t \in [0,1], g(t) = f((1-t)a+tb) \quad \text{et} \quad \varphi(t) = Mt\|b-a\|$$

où $M = \sup\{\|df(z)\|_{\mathcal{L}(E,F)}, z \in]a,b[\}$. En effet g et φ sont continues sur $[0,1]$ et dérivables sur $]0,1[$ avec

$$\forall t \in]0,1[, \;\; g'(t) = df((1-t)a+tb)(b-a), \quad \|g'(t)\| \leq M\|b-a\| = \varphi'(t).$$

On obtient alors

$$\|f(b)-f(a)\| = \|g(1)-g(0)\| \leq \varphi(1)-\varphi(0) = M\|b-a\|.$$

□

Corollaire 1.33. *Soit Ω un ouvert convexe de E et $f : \Omega \to F$ une fonction différentiable sur Ω. Si*

$$\forall x \in \Omega, \quad \|df(x)\|_{\mathcal{L}(E,F)} \leq k$$

pour $k \in \mathbb{R}$, alors f est k-lipschitzienne sur Ω, c'est-à-dire

$$\forall (x,y) \in \Omega, \quad \|f(y)-f(x)\|_F \leq k\|y-x\|_E.$$

Remarque 1.34. Une autre façon de retenir le résultat est d'écrire, si $f : \Omega \to F$ avec Ω convexe, alors

$$\forall (x,y) \in \Omega, \quad \|f(y)-f(x)\|_F \leq \Big(\sup_{z\in\Omega} \|df(z)\|_{\mathcal{L}(E,F)}\Big)\|y-x\|_E.$$

Démonstration. Par convexité de Ω, si $(x,y) \in \Omega^2$ alors $[x,y] \subset \Omega$ donc on peut appliquer le théorème 1.30. □

Exemple 1.35. Soit $f : \Omega \to F$ une fonction de classe C^1 où Ω est un ouvert de E et supposons $\dim(E) < \infty$. Alors f est localement lipschitzienne, c'est-à-dire [26] que pour tout $x \in \Omega$, il existe $\alpha > 0$ tel que f est lipschitzienne sur $B(x,\alpha)$. En effet, il suffit de prendre $\alpha > 0$ tel que $\overline{B}(x,\alpha) \subset \Omega$ (un tel $\alpha > 0$ existe puisque Ω est ouvert), et par continuité de df et compacité de $\overline{B}(x,\alpha)$ (car on est en dimension finie), on sait que df est bornée sur le convexe $\overline{B}(x,\alpha)$, donc le corollaire 1.33 s'applique.

26. Notons qu'une autre formulation équivalente est la suivante :

$$f \text{ est localement lipschitzienne sur } \Omega \iff \text{ pour tout compact } K \subset \Omega, f \text{ est lipschitzenne sur } K.$$

On laisse la preuve en exercice.

Exercice 1.36 (Contre-exemple quand Ω n'est pas convexe, voir ici (remarque 1.8.22) ou [SR08]). On pose $\Omega = \{(x, y) \in \mathbb{R}^2, x > 0, 1 < x^2 + y^2 < 4\}$ et

$$\forall (x, y) \in \Omega, \ \ f(x, y) = \arctan\left(\frac{y}{x}\right).$$

On munit $\mathbb{R}^2$ de la norme euclidienne $\|\cdot\|_2$.

1. Montrer que Ω est connexe mais non convexe.
2. Montrer que f est différentiable sur Ω et pour $(x, y) \in \Omega$, calculer $df(x, y) \in \mathcal{L}(\mathbb{R}^2, \mathbb{R})$.
3. Montrer qu'on peut identifier $\mathcal{L}(\mathbb{R}^2, \mathbb{R})$ muni de la norme subordonnée avec $\mathbb{R}^2$ muni de la norme $\|\cdot\|_2$. On peut ainsi voir $df(x, y) \in \mathbb{R}^2$.
4. Montrer
$$\forall (x, y) \in \Omega, \ \ \|df(x, y)\|_2 \leq 1,$$
où la norme $\|\cdot\|_2$ est la norme euclidienne sur $\mathbb{R}^2$.
5. Montrer que f n'est pas 1-lipschitzienne.

Remarque 1.37. Il est naturel de se demander si on peut quand même obtenir un résultat quand Ω n'est pas convexe, mais seulement connexe : on renvoit par exemple à [Car67, page 48] et à l'exercice suivant.

Exercice 1.38. Soit Ω un ouvert connexe. Rappelons que d'après la proposition 2.22, Ω est connexe par ligne brisée.

1. Pour $(x, y) \in \Omega$ on pose
$$d_\Omega(x, y) := \inf \left\{ \sum_{i=0}^{n-1} \|x_{i+1} - x_i\|, \ \ n \in \mathbb{N}, \ (x_0, \ldots, x_n) \text{ ligne brisée entre } x \text{ et } y \text{ dans } \Omega \right\}.$$
   (a) Montrer que d_Ω est une distance sur Ω.
   (b) Montrer que les ouverts pour d_Ω sont les mêmes que pour la distance induite par $\|\cdot\|$ [27].
2. Montrer que si $f : \Omega \to F$ est une fonction différentiable sur Ω telle que
$$\forall x \in \Omega, \quad \|df(x)\|_{\mathcal{L}(E,F)} \leq k$$
pour $k \in \mathbb{R}$, alors
$$\forall (x, y) \in \Omega, \qquad \|f(y) - f(x)\|_F \leq k d_\Omega(x, y).$$

Rappelons qu'au Chapitre I, on a utilisé ces résultats pour établir le lien entre le signe de la dérivée et le sens de variation des fonctions ; ces mots n'ont pas a priori de sens en dimension supérieure, mais le résultat suivant, fondamental, persiste :

Corollaire 1.39. *Soit Ω un ouvert connexe*[28]*, et $f : \Omega \to F$ une fonction différentiable sur Ω. On suppose que*

$$\forall x \in \Omega, \ \ df(x) = 0.$$

Alors f est constante sur Ω.

Démonstration. Si Ω est vide, le résultat est trivial, sinon il existe $x_0 \in \Omega$, et on peut considérer $A = \{x \in \Omega, \ f(x) = f(x_0)\}$. L'ensemble A est non vide il contient x_0 et fermé par continuité de f. Pour conclure, du fait de la connexité de Ω, il reste à voir que A est ouvert : si $x \in A$, comme Ω est ouvert il existe $\delta > 0$ tel que $B(x, \delta) \subset \Omega$ et cette boule est un ouvert convexe. On peut donc appliquer le corollaire 1.33 à la restriction de f sur $B(x, \delta)$ avec la constante $k = 0$, ce qui donne $\forall y \in B(x, \delta), f(y) = f(x)$ et donc que $B(x, \delta) \subset A$, ce qui conclut la preuve. $\square$

27. Rappelons qu'on dit que les distances sont topologiquement équivalentes.
28. Attention à bien lire, on a bien écrit connexe et non convexe. Bien sûr le résultat est vrai si Ω est convexe puisqu'un ensemble convexe est connexe, mais on a mis ici une hypothèse beaucoup plus faible qu'au corollaire 1.33.

Exercice 1.40. Soit $f : \Omega \to F$ différentiable sur Ω, et (x, y) tels que $[x, y] \subset \Omega$. En appliquant le théorème 1.30 à une fonction bien choisie, montrer que

$$\|f(y) - f(x) - df(x)(y - x)\|_F \leq \sup_{z \in]x,y[} \|df(z) - df(x)\|_{\mathcal{L}(E,F)} \|y - x\|_E.$$

Citons une dernière application importante du théorème des accroissements finis :

Théorème 1.41. *Supposons que $E = \mathbb{R}^n$, et $f : \Omega \to F$ où Ω est un ouvert de E. Alors f est continument différentiable sur Ω si et seulement si f admet des dérivées partielles dans toutes les directions sur Ω et que $x \mapsto \partial_i f(x)$ est continue sur Ω.*

Remarque 1.42. On a bien insisté précédemment sur le fait que la différentiabilité d'une fonction f en un point x est un résultat plus fort que l'existence de dérivées partielles en toutes les coordonnées en ce point x. De la même manière, la notion de différentiabilité sur un ouvert Ω est une notion plus forte que l'existence de dérivées partielles en toutes les coordonnées sur Ω. Mais si on suppose dans un cas la différentielle continue en tant que fonction de x, et les dérivées partielles $(\partial_i f)_{1 \leq i \leq n}$ continues en tant que fonctions de x, alors les deux notions coïncident. Dit autrement, les concepts se "rejoignent" au niveau des fonctions de classe C^1.

Démonstration. Si on suppose f est différentiable et df est continue, alors on sait que f admet des dérivées partielles dans toutes directions et que pour tout $i \in [\![1, n]\!]$ et tout $x \in \Omega, \partial_i f(x) = df(x)(e_i)$ qui est donc continue sur Ω.

On suppose réciproquement que f admet des dérivées partielles en toutes directions et en tout $x \in \Omega$. Pour éviter les technicités, on suppose $n = 2$ [29] ; aussi on munit $\mathbb{R}^2$ de la norme $\| \cdot \|_\infty$ de sorte que $B(0, r) =]-r, r[^2$ pour $r > 0$. Soit $x = (x_1, x_2) \in \Omega$, et $\varepsilon > 0$. Comme Ω est ouvert, il existe $\delta_0 > 0$ tel que $B(x, \delta_0) \subset \Omega$. Soit $\delta \leq \delta_0$ qui sera précisé plus loin. Pour $h = (h_1, h_2) \in B(0, \delta)$, on remarque

$$\begin{aligned} f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) - \partial_1 f(x)(h_1) - \partial_2 f(x)(h_2) \\ = \Big[f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2 + h_2) - \partial_1 f(x)(h_1)\Big] \\ + \Big[f(x_1, x_2 + h_2) - f(x_1, x_2) - \partial_2 f(x)(h_2)\Big] \end{aligned}$$

Pour le premier morceau, on applique le résultat de l'exercice 1.40 à la fonction $h_1 \in]-\delta, \delta[\mapsto f(x_1 + h_1, x_2 + h_2)$ qui est bien différentiable [30], et donc [31]

$$\Big|f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2 + h_2) - \partial_1 f(x)(h_1)\Big| \leq \sup_{z \in B(x,\delta)} \|\partial_1 f(z) - \partial_1 f(x)\| |h_1|$$

De même en appliquant l'exercice 1.40 à la fonction $h_2 \in]-\delta, \delta[\mapsto f(x_1, x_2 + h_2)$ on obtient

$$\Big|f(x_1, x_2 + h_2) - f(x_1, x_2) - \partial_2 f(x)(h_2)\Big| \leq \sup_{z \in B(x,\delta)} \|\partial_2 f(z) - \partial_2 f(x)\| |h_2|.$$

Par continuité de $\partial_1 f$ et $\partial_2 f$, il existe $\delta > 0$ tel que

$$\forall z \in B(x, \delta), \quad \sup_{z \in B(x,\delta)} \|\partial_1 f(z) - \partial_1 f(x)\| \leq \varepsilon \quad \text{et} \quad \sup_{z \in B(x,\delta)} \|\partial_2 f(z) - \partial_2 f(x)\| \leq \varepsilon$$

29. La conviction de l'auteur est que le lecteur ayant compris le cas $n = 2$ sera convaincu que le cas $n \geq 2$ aurait pu être obtenu avec une preuve ayant la même logique. Notez néanmoins que la formulation choisie du théorème 1.41 ne se prête pas bien à une récurrence, contrairement à l'exercice 1.43 qui s'y prêtera très bien : on pourra en effet commencer par traiter le cas d'un produit de deux espaces vectoriels normés, et dans la preuve de l'hérédité, on pourra voir $E_1 \times \cdots \times E_{n+1}$ comme le produit de deux espaces à savoir $\Big(E_1 \times \cdots E_n\Big) \times E_{n+1}$, ce qui ramène aux cas d'un produit de n espaces (hypothèse de récurrence) et de deux espaces (traité en premier lieu).

30. On peut en fait dire dérivable ici puisque $h_1 \in \mathbb{R}$.

31. Dans le terme de droite de l'inégalité obtenue, on pourrait restreindre l'ensemble des z pris en compte au segment formé par les points $(x_1, x_2 + h_2)$ et $(x_1 + h_1, x_2 + h_2)$, mais on obtient un nombre plus grand en prenant en compte toute la boule $B(x, \delta)$.

et donc après addition on obtient

$$\forall h \in B(0,\delta), \quad \|f(x+h) - f(x) - \partial_1 f(x)(h_1) - \partial_2 f(x)(h_2)\| \leq \varepsilon(|h_1| + |h_2|)$$

avec $(h_1, h_2) \mapsto \partial_1 f(x)(h_1) - \partial_2 f(x)(h_2)$ linéaire (continue), donc f est différentiable en x avec $df(x)$: $h \in \mathbb{R}^2 \mapsto \partial_1 f(x)(h_1) - \partial_2 f(x)(h_2)$. □

Exercice 1.43. *Le théorème 1.41 peut être adapté sans réelle différence au cas des différentielles partielles. On suppose que $E = E_1 \times \cdots E_n$ avec $n \in \mathbb{N}^*$ et $(E_i)_{i \in [\![1,n]\!]}$ des espaces vectoriels normés. Soit Ω un ouvert de E et $f : \Omega \to F$.

Montrer que f est de classe C^1 sur Ω si et seulement si f admet des différentielles partielles en toutes les variables sur Ω, et que ces différentielles partielles sont continues sur Ω.

2 Différentielle seconde

Dans cette section, on montre comment itérer la notion de différentielle. Il faut être plus vigilant que pour la notion de dérivation, car l'espace d'arrivée de la fonction df n'est pas le même que l'espace d'arrivée de la fonction f.

2.1 Définition

On se donne à nouveau E et F deux espaces vectoriels normés, et Ω un ouvert de E. Si une fonction est une fois différentiable sur l'ouvert Ω, alors on peut considérer $df : \Omega \to \mathcal{L}(E,F)$ et comme $\mathcal{L}(E,F)$ est également un espace vectoriel normé, on peut se demander si cette application est elle-même différentiable. Plus précisément [32] :

Définition 2.1. Soit $x \in \Omega$, et $f : \Omega \to F$. On dit que f est deux fois différentiable en x si

— il existe un voisinage $\mathcal{V}$ de x sur lequel f est différentiable,

— la fonction $df : \mathcal{V} \to \mathcal{L}(E,F)$ est différentiable en x.

Dans ce cas, on définit $d^2 f(x) = d(df)(x) \in \mathcal{L}(E, \mathcal{L}(E,F))$ la différentielle seconde de f en x.

Remarque 2.2 (Identification)**.** Étant donnés E, F, G trois espaces vectoriels normés, nous allons réfléchir à l'espace

$$\mathcal{L}(E, \mathcal{L}(F,G))$$

qui n'est pas très commode à manipuler. En fait on peut l'identifier à l'espace

$$\mathcal{L}_2(E \times F, G) := \Big\{ a : E \times F \to G, \ \ a \text{ bilinéaire continue} \Big\}$$

de façon algébrique et analytique, c'est-à-dire qu'on peut définir entre ces deux espaces un isomorphisme qui préserve la norme [33], c'est-à-dire une isométrie :

Lemme 2.3. *Les applications*

$$\begin{array}{rccc} \Phi : & \mathcal{L}(E, \mathcal{L}(F,G)) & \longrightarrow & \mathcal{L}_2(E \times F, G) \\ & L & \longmapsto & \left\{ \begin{array}{ccc} E \times F & \to & G \\ (x,y) & \mapsto & L(x)(y) \end{array} \right. \end{array} \quad \textit{et} \quad \begin{array}{rccc} \Psi : & \mathcal{L}_2(E \times F, G) & \longrightarrow & \mathcal{L}(E, \mathcal{L}(F,G)) \\ & a & \longmapsto & \left\{ \begin{array}{ccc} E & \to & \mathcal{L}(F,G) \\ x & \mapsto & a(x,\cdot) \end{array} \right. \end{array}$$

sont des bijections réciproques l'une de l'autre, sont linéaires, et isométriques au sens

$$\forall L \in \mathcal{L}(E, \mathcal{L}(F,G)), \quad \|\Phi(L)\|_{\mathcal{L}_2(E \times F, G)} = \|L\|_{\mathcal{L}(E, \mathcal{L}(F,G))}$$

32. Comme pour la notion de dérivée, il faut être vigilant à cette étape : la différentiabilité seconde en un point ne peut avoir de sens que si la fonction est une fois différentiable sur un voisinage du point considéré.

33. Rappelons la norme mise sur $\mathcal{L}_2(E \times F, G)$ est

$$\forall a \in \mathcal{L}_2(E \times F, G), \quad \|a\|_{\mathcal{L}_2(E \times F, G)} := \sup_{\|x\|_E \leq 1, \|y\|_F \leq 1} \|a(x,y)\|_G,$$

voir l'exercice 2.44 au Chapitre III.

Démonstration. [34] Commençons par clarifier les mécanismes de Φ et Ψ : soit $L \in \mathcal{L}(E, \mathcal{L}(F, G))$ et $a \in \mathcal{L}_2(E \times F, G)$: alors

$$\forall (x, y) \in E \times F, \;\; \Phi(L)(x, y) = L(x)(y), \quad \text{et} \quad \Psi(a)(x)(y) = a(x, y).$$

Ainsi on obtient

$$\forall (x, y) \in E \times F, \;\; \Phi(\Psi(a))(x, y) = \Psi(a)(x)(y) = a(x, y), \quad \text{et} \quad \Psi(\Phi(L))(x)(y) = \Phi(L)(x)(y) = L(x)(y)$$

ce qui signifie

$$\forall L \in \mathcal{L}(E, \mathcal{L}(F, G)), \;\; \Psi \circ \Phi(L) = L \quad \text{et} \quad \forall a \in \mathcal{L}_2(E \times F, G), \;\; \Phi \circ \Psi(a) = a.$$

Ainsi ces fonctions sont des bijections réciproques l'une de l'autre. Leurs linéarités sont claires. Quant à l'égalité des normes, elle s'obtient en écrivant :

$$\begin{aligned} \forall L \in \mathcal{L}(E, \mathcal{L}(F, G)), \quad \|\Phi(L)\|_{\mathcal{L}_2(E \times F, G)} &= \sup_{\|x\|_E \le 1, \|y\|_F \le 1} \|\Phi(L)(x, y)\|_G = \sup_{\|x\|_E \le 1} \sup_{\|y\|_F \le 1} \|L(x)(y)\|_G \\ &= \sup_{\|x\|_E \le 1} \|L(x)\|_{\mathcal{L}(F,G)} = \|L\|_{\mathcal{L}(E, \mathcal{L}(F,G))}. \end{aligned}$$

□

Ainsi, si f est deux fois diférentiable en x, on écrira

$$\forall (h, k) \in E^2, \quad d^2 f(x)(h, k) \text{ à la place de } d^2 f(x)(h)(k).$$

Le résultat suivant permet de voir qu'une différentielle seconde n'est pas seulement une application bilinéaire, elle est aussi symétrique :

Théorème 2.4 (Théorème de Schwarz). *Soit $f : \Omega \to F$ deux fois différentiable en $x \in \Omega$. Alors*

$$\forall (h, k) \in E^2, \quad d^2 f(x)(h, k) = d^2 f(x)(k, h),$$

c'est-à-dire que $d^2 f(x)$ est une application bilinéaire symétrique.

Remarque 2.5. Ce résultat a pour conséquence qu'il suffit de connaître $h \mapsto d^2 f(x)(h, h)$ (c'est-à-dire l'application quadratique associée) pour identifier $d^2 f(x)$: en effet, on a [35]

$$d^2 f(x)(h, k) = \frac{1}{2}\Big[d^2 f(x)(h + k, h + k) - d^2 f(x)(h, h) - d^2 f(x)(k, k)\Big].$$

Pour cette raison, si on sait déjà que f est deux fois différentiables [36], alors on peut identifier la valeur de $d^2 f(x)$ en faisant un développement limité à l'ordre 2 de f en x, voir la proposition 2.12. [37]

34. Ce lemme et sa preuve sont absolument formels ; on conseille le lecteur de s'entraîner à faire ce genre de preuve avec un maximum d'automatisme et de rapidité (et donc, faites-là avant de la lire !!!). S'il n'y parvient pas, c'est que les concepts liés aux fonctions (incluant les notions d'espace de départ, d'arrivée, d'égalité entre fonction, de composition) ne sont pas parfaitement maîtrisés.

35. Une telle identité s'appelle identité de polarisation ; il en existe deux autres :

$$d^2 f(x)(h, k) = \frac{1}{2}\Big[d^2 f(x)(h, h) + d^2 f(x)(k, k) - d^2 f(x)(h - k, h - k)\Big].$$

$$d^2 f(x)(h, k) = \frac{1}{4}\Big[d^2 f(x)(h + k, h + k) - d^2 f(x)(h - k, h - k)\Big].$$

36. Information qu'on peut obtenir sans avoir à calculer.

37. Attention, comme on l'a remarqué pour la notion de dérivée, si une fonction admet un développement limité à l'ordre 2 en un point, cela ne permet pas de dire qu'elle est deux fois différentiable en ce point. C'est pourquoi on demande ici de savoir a priori que la fonction est deux fois différentiable.

Démonstration. Soit $x \in \Omega$ et $\varepsilon > 0$. Par définition de la différentiabilité de df en x, il existe $\delta > 0$ tel que $B(x,\delta) \subset \Omega$ et pour tout $h \in B(0,\delta)^2$,

$$\|df(x+h) - df(x) - d^2f(x)(h)\| \leq \varepsilon\|h\|.$$

On en déduit

$$\forall (h,k) \in B(0,\frac{\delta}{2})^2, \forall t \in [0,1],\ \|df(x+th+k) - df(x) - d^2f(x)(th+k)\| \leq \varepsilon\|th+k\| \leq \varepsilon(\|h\|+\|k\|).$$

On fixe $(h,k) \in B(0,\delta)^2$, et on regarde la fonction

$$\varphi : t \in [0,1] \mapsto f(x+th+k) - f(x+th)$$

qui d'après la proposition 1.13 est dérivable et telle que

$$\begin{aligned}\forall t \in [0,1],\ \varphi'(t) &= df(x+th+k)(h) - df(x+th)(h)\\ &= \big[df(x+th+k)(h) - df(x)(h)\big] - \big[df(x+th)(h) - df(x)(h)\big].\end{aligned}$$

On en déduit :

$$\begin{aligned}\forall t \in [0,1],\ \|\varphi'(t) - d^2f(x)(k,h)\| &\leq \Big\|df(x+th+k)(h) - df(x)(h) - d^2f(x)(k+th,h)\Big\|\\ &\quad + \Big\|df(x+th)(h) - df(x)(h) - d^2f(x)(th,h)\Big\|\\ &\leq 2\varepsilon(\|h\|+\|k\|)\|h\| \leq 2\varepsilon(\|h\|+\|k\|)^2.\end{aligned}$$

On applique le théorème 1.30 à la fonction $t \in [0,1] \mapsto \varphi(t) - td^2f(x)(k,h)$, ce qui donne

$$\|\varphi(1) - \varphi(0) - d^2f(x)(k,h)\| \leq 2\varepsilon(\|h\|+\|k\|)^2.$$

Mais comme $\varphi(1) - \varphi(0)$ est clairement symétrique en h et k, on obtient (en inversant les rôles de (h,k) et en utilisant l'inégalité triangulaire) :

$$\begin{aligned}\|d^2f(x)(h,k) - d^2f(x)(k,h)\| &\leq \|\varphi(1) - \varphi(0) - d^2f(x)(h,k)\| + \|\varphi(1) - \varphi(0) - d^2f(x)(k,h)\|\\ &\leq 4\varepsilon(\|h\|+\|k\|)^2.\end{aligned}$$

Ceci est vrai pour tout $(h,k) \in B(0,\delta)^2$, mais par bilinéarité, cela reste valable sur E^2, et on obtient le résultat en faisant tendre ε vers 0. □

Définition 2.6. Soit $f : \Omega \to F$. Si f est deux fois différentiable en tout point $x \in \Omega$ et la fonction $x \in \Omega \mapsto d^2f(x) \in \mathcal{L}_2(E \times E, F)$ est continue, on dit que f est de classe C^2 sur Ω.

2.2 Dérivées partielles d'ordre 2

Comme dans le paragraphe 1.4.1, on suppose pour simplifier que $E = \mathbb{R}^n$.

Définition 2.7. Soit $x \in \Omega$, et $(i,j) \in [\![1,n]\!]^2$. On dit que f admet une dérivée partielle seconde d'indice (i,j) en x si

— il existe un voisinage $\mathcal{V}$ de x sur lequel f admet une dérivée partielle par rapport à la j-ième variable,

— l'application $y \in \mathcal{V} \mapsto \partial_j f(y)$ admet une dérivée partielle par rapport à la i-ième variable en x.

Dans ce cas, on pose $\partial_i(\partial_j f(x)) \in F$ est la dérivée partielle seconde de f d'indice (i,j) en x, qu'on peut noter

$$\partial^2_{ij}f(x) \quad \text{ou} \quad \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \quad \text{ou} \quad D_{ij}f(x) \quad \text{ou} \quad \partial_{x_i x_j}f(x).$$

Remarque 2.8. Si f est deux fois différentiable en x, alors f admet des dérivées partielles secondes de tout indice en x, et

$$\partial_{ij} f(x) = d^2 f(x)(e_j, e_i).$$

Ainsi, par le théorème de Schwarz 2.4, on a

$$\partial_{ij} f(x) = \partial_{ji} f(x).$$

Mais attention, une fonction peut admettre des dérivées partielles secondes de tout indice sans être deux fois différentiable. Aussi, sans supposer la différentiabilité seconde, le théorème de Schwarz peut être mis en défaut, voir l'exercice 2.10.

Remarque 2.9. En appliquant le théorème 1.41, on en déduit que si f admet des dérivées partielles secondes de tout indice et en tout point $x \in \Omega$, et que ces dérivées partielles sont continues en la variable x, alors f est de classe C^2 sur Ω, donc en particulier est deux fois différentiable en tout $x \in \Omega$.

Exercice 2.10. Le contre-exemple suivant a été donné par Peano en 1884 : on pose

$$\forall (x,y) \in \mathbb{R}^2, \quad f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{sinon} \end{cases}.$$

Montrer que f admet des dérivées partielles secondes d'indice $(1,2)$ et $(2,1)$ en $(0,0)$ et que

$$\partial^2_{12} f(0,0) = 1, \qquad \text{et} \qquad \partial^2_{21} f(0,0) = -1.$$

Remarque 2.11 (Hessienne)**.** Supposons $E = \mathbb{R}^n$ et $F = \mathbb{R}$, et $f : \Omega \to \mathbb{R}$ deux fois différentiable en x. Alors $d^2 f(x)$ est une forme bilinéaire symétrique sur $\mathbb{R}^n$, et on peut la représenter de façon matricielle si on travaille avec la base canonique de $\mathbb{R}^n$: en effet

$$\forall (h,k) \in (\mathbb{R}^n)^2, \;\; d^2 f(x)(h,k) = \sum_{i,j=1}^n d^2 f(x)(e_i, e_j) h_i k_j = \sum_{i,j=1}^n \partial^2_{ij} f(x) h_i k_j,$$

donc on est amené à définir

$$H_f(x) := \begin{pmatrix} \partial^2_{11} f(x) & \dots & \partial^2_{1n} f(x) \\ \vdots & \vdots & \vdots \\ \partial^2_{n1} f(x) & \dots & \partial^2_{nn} f(x) \end{pmatrix} = \Big(\partial_{ij} f(x)\Big)_{1 \leq i,j \leq n}$$

qu'on appelle matrice Hessienne de f en x, et on a alors

$$\forall (h,k) \in (\mathbb{R}^n)^2, \quad d^2 f(x)(h,k) = k^T \times H_f(x) \times h = \langle H_f(x) h, k \rangle.$$

2.3 Formules de Taylor

À nouveau, E et F désignent des espaces vectoriels normés, et Ω un ouvert de E.

Proposition 2.12 (Formule de Taylor-Young)**.** *Soit $f : \Omega \to F$, deux fois différentiable en $x \in \Omega$. Alors*

$$f(x+h) = f(x) + df(x)(h) + \frac{1}{2} d^2 f(x)(h,h) + o(\|h\|^2)$$

Démonstration. Soit $\varepsilon > 0$. On définit sur un voisinage ouvert $\mathcal{V}$ de 0

$$h \mapsto R(h) = f(x+h) - f(x) - df(x)(h) - \frac{1}{2} d^2 f(x)(h,h).$$

Alors R est différentiable sur $\mathcal{V}$ et

$$\forall h \in \mathcal{V}, \forall k \in E, \; dR(h)(k) = df(x+h)(k) - df(x)(k) - d^2 f(x)(h,k)$$

où on a utilisé le théorème de Schwarz 2.4. Par définition de la différentiabilité seconde de f en x, il existe $\delta > 0$ tel que

$$\forall h \in B(0,\delta),\ \|dR(h)\|_{\mathcal{L}(E,F)} = \|df(x+h) - df(x) - d^2f(x)(h,\cdot)\|_{\mathcal{L}(E,F)} \leq \varepsilon\|h\|_E.$$

On applique le théorème 1.30 à la fonction R et on obtient :

$$\forall h \in B(0,\delta),\ \|R(h) - R(0)\| \leq \sup_{z\in[0,h]} \|dR(z)\|_{\mathcal{L}(E,F)}\|h\|_E \leq \varepsilon\|h\|_E^2$$

d'où le résultat puisque $R(0) = 0$. □

Comme il n'y a plus d'égalité des accroissements finis, il n'y a plus de formule de Taylor-Lagrange : mais comme on l'a fait avec l'inégalité des accroissements finis, on voit à l'exercice suivant (qui adapte à l'ordre 2 le résultat de l'exercice 1.40) qu'on donner une estimation de l'erreur entre f et son développement limité à l'ordre 2 :

Exercice 2.13. Soit $f : \Omega \to F$, deux fois différentiable sur Ω, $x \in \Omega$ et $h \in E$ tel que $[x, x+h] \subset \Omega$ On définit R :

$$\forall u \in [0,h], R(u) = f(x+u) - f(x) - df(x)(u) - \frac{1}{2}d^2f(x)(u,u).$$

1. En appliquant l'exercice 1.40, montrer que

$$\forall u \in [0,h],\ \ \|dR(u)\|_{\mathcal{L}(E,F)} \leq \sup_{z\in[x,x+h]} \|d^2f(z) - d^2f(x)\|_{\mathcal{L}_2(E\times E,F)}\|h\|_E.$$

2. En déduire que

$$\|f(x+h) - f(x) - df(x)(h) - \frac{1}{2}d^2f(x)(h,h)\|_F \leq \sup_{z\in[x,x+h]} \|d^2f(z) - d^2f(x)\|_{\mathcal{L}_2(E\times E,F)}\|h\|^2.$$

Enfin, on a encore la formule de Taylor avec reste intégral :

Proposition 2.14. *On suppose que F est un espace de Banach*[38] *et $f : \Omega \to F$ de classe C^2 sur Ω. Pour tout $x \in \Omega$ et $h \in E$ tel que $[x, x+h] \subset \Omega$, on a*

$$f(x+h) = f(x) + df(x)(h) + \int_0^1 (1-t)d^2f(x+th)(h,h)dt.$$

Démonstration. On pose $\varphi(t) = f(x+th)$ pour $t \in [0,1]$. Alors avec une intégration par parties[39]

$$\begin{aligned} f(x+h) - f(x) &= \varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt \\ &= [(t-1)\varphi'(t)]_0^1 - \int_0^1 (t-1)\varphi''(t)dt = \varphi'(0) + \int_0^1 (1-t)\varphi''(t)dt \end{aligned}$$

d'où le résultat puisque

$$\varphi'(t) = df(x+th)(h) \quad \text{et} \quad \varphi''(t) = d^2f(x+th)(h,h).$$

□

38. C'est pour que l'intégrale de fonctions à valeurs dans F soit définie ; en pratique on a très souvent F de dimension finie, donc le lecteur peut se restreindre à ce cas où il suffit d'intégrer chaque fonction coordonnée pour définir une intégrale à valeurs dans F.

39. En fait, on pouvait aussi directement appliquer la formule de Taylor avec reste intégral prouvé à la proposition 3.55 au Chapitre I ; cette formule reste valable si on remplace l'espace d'arrivée $\mathbb{R}$ par F espace de Banach (ou plus simplement $\mathbb{R}^p$).

2.4 Ordre supérieur

Bien sûr, on peut itérer ce qui précède à tout ordre $n \in \mathbb{N}^*$. Cela est assez fastidieux, mais si on a compris les difficultés qui apparaissent à l'ordre 2, il n'y a pas de nouvelle difficulté à aller à des ordres plus élevés. De plus, les applications sont souvent restreintes à l'ordre 2, donc on ne fait pas l'effort d'étudier l'ordre $n \geq 3$, on renvoit à la littérature pour plus de détails (voir par exemple [Rou03, page 286-288]). Précisons seulement quelques éléments :

1. si f est p-fois différentiable en x pour $p \geq 1$, alors $d^p f(x)$ sera une application p-linéaire continue de E^p vers F, symétrique au sens où échanger deux coordonnées dans E^p ne change pas le résultat,
2. si $E = \mathbb{R}^n$, pour $\alpha \in \mathbb{N}^n$ avec $|\alpha| := \alpha_1 + \cdots \alpha_n \leq p$, on peut définir
$$\partial^\alpha f(x) = \partial_1^{\alpha_1}\partial_2^{\alpha_2}\cdots\partial_n^{\alpha_n} f(x) \qquad \text{aussi dénotées} \qquad \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}\dots\partial x_n^{\alpha_n}}(x),$$
et l'ordre des différentielles n'a pas d'importance (si f est p fois différentiable en x),
3. le lien entre différentielle et dérivée partielle est
$$d^k f(x)(h^{[k]}) = \sum_{i_1,\dots,i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_1}\partial x_{i_2}\dots\partial x_{i_k}}(x) h_{i_1}\dots h_{i_k} = \sum_{\alpha\in\mathbb{N}^n, |\alpha|=k} \frac{k!}{\alpha_1!\cdots\alpha_n!}\partial^\alpha f(x) h_1^{\alpha_1}\cdots h_n^{\alpha_n},$$
où $h^{[k]} = (h, h, \cdots, h) \in E^k$ pour $k \in \mathbb{N}$,
4. la formule de Taylor reste intégral s'écrit, si f est de classe C^{p+1} sur Ω et (x, h) sont tels que $[x, x+h] \subset \Omega$,
$$f(x+h) = f(x) + \sum_{k=1}^{p} \frac{1}{k!} d^k f(x)(h^{[k]}) + \int_0^1 \frac{(1-t)^p}{p!} d^{p+1} f(x+th)(h^{[p+1]})dt.$$

3 Inversion

3.1 Différentielle d'un inverse

Définition 3.1. Soit Ω un ouvert de E, Ω' un ouvert de F, et $f : \Omega \to \Omega'$.

— On dit que f est un homéomorphisme si f est bijective, continue, et que $f^{-1} : \Omega' \to \Omega$ est continue.

— On dit que f est un C^1-difféomorphisme si f est bijective, de classe C^1, et que $f^{-1} : \Omega' \to \Omega$ est de classe C^1.

Exemple 3.2. Attention, une fonction peut être un homéomorphisme et être différentiable, mais peut ne pas être un difféomorphisme : par exemple la fonction $x \in \mathbb{R} \mapsto x^3$.

Commençons par analyse la différentiabilité d'une fonction inverse, comme on l'a fait pour la dérivation à la proposition 3.39 au Chapitre I.

Proposition 3.3. *Soit Ω un ouvert de E, Ω' un ouvert de F, $f : \Omega \to \Omega'$ et $x_0 \in \Omega$. On suppose que :*

— *f est un homéomorphisme,*

— *f est différentiable en x_0 et $df(x_0)$ est un isomorphisme*[40].

40. On entend par là que $df(x_0)$, qui est une application linéaire continue, est bijective et que sont inverse est également une application linéaire continue ; la linéarité de cet inverse est gratuite, la continuité ne l'est pas *a priori*. Certains auteurs disent seulement "$df(x_0)$ est inversible" (i.e. bijective et son inverse jouit des mêmes propriétés) ; remarquons que [Gou08b] dit "$df(x_0)$ est inversible et bicontinu" pour être plus précis ; peu importe le vocabulaire, il faut surtout comprendre ce qu'on entend par là. Si on est en dimension finie bien sûr, la continuité des applications linéaires est automatique. Et pour ceux qui sont familiers avec le théorème d'isomorphisme de Banach (qui dit que si $L : E \to F$ est une application linéaire continue et bijective, et que E, F sont des espaces de Banach, alors $L^{-1} : F \to E$ est continue), il n'y a en fait pas d'ambiguité à mélanger les mots "bijectif" et "inversible" si E et F sont des espaces de Banach.

Alors f^{-1} est différentiable en $y_0 = f(x_0)$ et

$$d(f^{-1})(y_0) = (df(x_0))^{-1} = (df(f^{-1}(y_0)))^{-1}.$$

Remarque 3.4. Comparez l'énoncé et la preuve à ceux de la proposition 3.39 au Chapitre I. Notamment, ne manquez pas de remarquer qu'on utilise que la fonction f^{-1} est continue, ce qui découle du théorème de la bijection (théorème 3.14 au Chapitre I). Vous pouvez aussi faire le lien avec l'exercice 2.53 au Chapitre II où on calcule la dérivée de $\mathrm{Log} = \exp^{-1}$ (sur les bons ouverts) après avoir démontré que Log est continue.

Démonstration. L'application $df(x_0)^{-1}$ est linéaire et continue de F dans E par hypothèse. On pose $c = \|df(x_0)^{-1}\|$. Par différentiabilité de f en x_0, on a

$$\forall x \in \Omega, \;\; f(x) = f(x_0) + df(x_0)(x - x_0) + \|x - x_0\|\varepsilon(x)$$

où ε est une fonction définie sur Ω et telle que $\lim_{x \to x_0} \varepsilon(x) = 0$. Pour tout $x \in \Omega$, on pose $y = f(x)$ et alors la formule précédente se réécrit :

$$df(x_0)(x - x_0) = y - y_0 - \|x - x_0\|\varepsilon(x)$$

et donc en composant par $df(x_0)^{-1}$ on obtient

$$x - x_0 = df(x_0)^{-1}(y - y_0) - \|x - x_0\|(df(x_0))^{-1} \circ \varepsilon(x)$$

c'est-à-dire

$$\forall y \in \Omega' \setminus \{y_0\}, \; f^{-1}(y) - f^{-1}(y_0) = df(x_0)^{-1}(y - y_0) - \|y - y_0\| \frac{\|f^{-1}(y) - f^{-1}(y_0)\|}{\|y - y_0\|}(df(x_0))^{-1} \circ \varepsilon \circ f^{-1}(y).$$

On pose ainsi $\widetilde{\varepsilon} : y \in \Omega' \setminus \{y_0\} \mapsto \frac{\|f^{-1}(y) - f^{-1}(y_0)\|}{\|y - y_0\|}(df(x_0))^{-1} \circ \varepsilon \circ f^{-1}(y)$ qu'on peut prolonger en y_0 par $\widetilde{\varepsilon}(y_0) = 0$. Nous devons montrer que $\lim_{y \to y_0} \widetilde{\varepsilon}(y) = 0$: en effet si tel est le cas, alors par définition f^{-1} est différentiable en y_0 et sa différentielle en y_0 est $df(x_0)^{-1}$.

On remarque que

$$\|x - x_0\| = \Big\| df(x_0)^{-1}\big[df(x_0)(x - x_0)\big] \Big\| \leq c\Big(\|y - y_0\| + \|x - x_0\| \|\varepsilon(x)\| \Big).$$

Par définition de la limite de ε, il existe $\mathcal{V}$ un voisinage de x_0 tel que $\forall x \in \mathcal{V}, \; \|\varepsilon(x)\| \leq \frac{1}{2c}$ et alors

$$\forall x \in \mathcal{V}, \;\; \|x - x_0\| \leq c\|y - y_0\| + \frac{1}{2}\|x - x_0\| \quad \text{d'où } \|x - x_0\| \leq 2c\|y - y_0\|,$$

et donc

$$\forall y \in f(\mathcal{V}), \;\; \|\widetilde{\varepsilon}(y)\| \leq 2c^2 \|\varepsilon(f^{-1}(y))\|$$

et on conclut en utilisant la continuité de f^{-1} et le fait que $\lim_{x \to x_0} \varepsilon(x) = 0$. □

Corollaire 3.5. *Supposons que E, F sont des espaces de Banach. Soit Ω un ouvert de E, Ω' un ouvert de F. Si $f : \Omega \to \Omega'$ est un homéomorphisme de classe C^1, alors*

$$f \text{ est un } C^1 - \text{difféomorphisme} \iff \forall x \in \Omega, df(x) \text{ est un isomorphisme.}$$

Remarque 3.6. En conséquence, pour qu'il y ait un C^1-difféomorphisme entre un ouvert (non vide) de E et un ouvert de F, il est nécessaire que E et F soient isomorphes. S'ils sont de dimension finie, ils ont donc même dimension. [41]

41. Autrement dit, il n'existe pas de C^1-difféomorphisme entre un ouvert de $\mathbb{R}^n$ et un ouvert de $\mathbb{R}^p$ si $n \neq p$. La question de savoir s'il existe un homéomorphisme (une application continue, bijective, d'inverse continu) entre $\mathbb{R}^n$ et $\mathbb{R}^p$ (ou entre des ouverts non vides de $\mathbb{R}^n$ et $\mathbb{R}^p$) si $n \neq p$ est beaucoup plus délicate ; la réponse reste négative, et c'est une conséquence du théorème de Brouwer dans le cas général.

Démonstration. Comment par supposer que f est un C^1-difféomorphisme ; alors on peut différentier les égalités

$$f \circ f^{-1} = \mathrm{Id}_{\Omega'}, \quad f^{-1} \circ f = \mathrm{Id}_{\Omega}$$

ce qui donne

$$\forall y \in \Omega', \ df(f^{-1}(y)) \circ d(f^{-1})(y) = \mathrm{Id}_F, \quad \text{et} \quad \forall x \in \Omega \ \ d(f^{-1})(f(x)) \circ df(x) = \mathrm{Id}_E$$

la première égalité pouvant se réécrire

$$\forall x \in \Omega, \ df(x) \circ d(f^{-1})(f(x)) = \mathrm{Id}_F.$$

Ceci montre que pour tout $x \in \Omega$, $df(x)$ est un isomorphisme et $(df(x))^{-1} = d(f^{-1})(f(x))$.

Réciproquement, on suppose que $df(x)$ est un isomorphisme pour tout $x \in \Omega$. Par la proposition 3.3, l'application f^{-1} est différentiable sur Ω' et

$$\forall y \in \Omega', \ d(f^{-1})(y) = (df(f^{-1}(y)))^{-1}.$$

Pour conclure, il faut montrer la continuité de $d(f^{-1})$: on réécrit la formule précédente sous la forme $d(f^{-1}) = \Phi \circ df \circ f^{-1}$ où

$$\begin{array}{rcl} \Phi : Isom(E,F) & \longrightarrow & Isom(F,E) \\ L & \longmapsto & L^{-1} \end{array} \tag{3.1}$$

Par hypothèse, f^{-1} et df sont continues ; il est classique que l'application Φ est aussi continue[42], ce qui conclut la preuve. □

3.2 Théorème d'inversion locale

Théorème 3.7 (Théorème d'inversion locale)**.** *Soit (E, F) deux espaces de Banach, Ω un ouvert de E, $f : \Omega \to F$ de classe C^1 et $x_0 \in \Omega$. Si $df(x_0) \in \mathcal{L}(E, F)$ est un isomorphisme*[43] *alors il existe $\mathcal{U}$ voisinage de x_0 et $\mathcal{V}$ voisinage de $f(x_0)$ tels que $f : \mathcal{U} \to \mathcal{V}$ est un C^1-difféomorphisme.*

Remarque 3.8. Pour interpréter ce résultat, on revient à l'idée initiale du calcul différentiel : autour de x_0, la fonction f est proche de $x \mapsto f(x_0) + df(x_0)(x - x_0)$ qui est bijective car $df(x_0)$ l'est.

Remarque 3.9. Si $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$, pour que l'hypothèse "$df(x_0)$ inversible" soit satisfaite, il est nécessaire que $n = p$.

Démonstration. **Étape 1 :** Réduction au cas $x_0 = 0$, $E = F$, $f(x_0) = 0$ et $df(x_0) = \mathrm{Id}_E$.

On pose pour $x \in \Omega - x_0$[44],

$$g(x) = (df(x_0))^{-1}\big[f(x + x_0) - f(x_0)\big].$$

42. Voir l'exercice 2.48 au Chapitre III où l'on propose une preuve dans le cas $E = F$; le cas général est similaire. Voir aussi [Car67, Théorème 1.7.3]. Si E et F sont de dimension finie, on peut considérer $E = F = \mathbb{R}^n$, et la continuité de Φ peut se déduire de la formule de Cramer

$$\forall A \in GL_n(\mathbb{R}), \ \ \Phi(A) = \frac{1}{\det(A)} Com(A)^T.$$

En effet les coefficients qui apparaissent dans $Com(A)$ sont des déterminants de sous-matrice de A, donc sont des fonctions continues (polynomiale en fait) de ses coefficients, ce qui correspond bien à dire qu'on est une fonction continue de A (quelle que soit la norme mise sur $\mathcal{M}_n(\mathbb{R})$, puisqu'en dimension finie les normes sont équivalentes). En fait, cet argument montre directement que Φ est une fonction de classe C^∞ sur $GL_n(\mathbb{R})$.

43. Remarquons d'une part que si E est de dimension finie (ce qui sera le cas pour la plupart des applications raisonnables au niveau de l'agrégation) alors toutes les applications linéaires sont continues ; donc l'hypothèse "$df(x_0)$ est inversible" convient également. Si E est de dimension infinie, comme il est supposé être un espace de Banach dans ce théorème, le théorème d'isomorphisme de Banach (hors programme, voir le Chapitre IX) affirme que si $T : E \to F$ est une application linéaire continue bijective entre deux espaces de Banach, alors $T^{-1} : F \to E$ sera continue. Donc si on travaille en dimension finie ou si on accepte ce théorème d'isomorphisme de Banach, alors l'hypothèse "$df(x)$ isomorphisme" peut simplement être remplacée par "$df(x_0)$ est bijective (et continue par définition de la différentiabilité)".

44. L'ensemble $\Omega - x_0 = \{x - x_0, x \in \Omega\}$ est la translation de Ω par le vecteur $-x_0$.

Par les règles de la proposition 1.13, g est de classe C^1, $g(0) = 0$ et $dg(0) = \mathrm{Id}_E$. Il est de plus facile de se convaincre que si l'on montre le résultat pour g, alors il sera valable pour f également. On suppose donc désormais $x_0 = 0$, $f(0) = 0$ et $df(0) = \mathrm{Id}_E$.

Étape 2 : Posons $\varphi : x \in \Omega \mapsto x - f(x)$. Alors φ est de classe C^1 sur Ω et $d\varphi(0) = 0$. Donc par continuité de $d\varphi$, il existe $r > 0$ tel que $\|d\varphi(x)\| \leq \frac{1}{2}$ pour tout $x \in B(0,r)$. Par le corollaire 1.33, comme $B(0,r)$ est un ouvert convexe, on en déduit que φ est 1/2-lipschitzienne sur $B(0,r)$, et en particulier elle est à valeurs dans $B(0,r/2)$, puisque $\varphi(0) = 0$. Remarquons également que par continuité, φ est également 1/2-lipschitzienne sur $\overline{B}(0,r)$, et qu'on a $\varphi(\overline{B}(0,r)) \subset \overline{B}(0,r/2)$.

Étape 3 : Soit $y \in B(0,r/2)$. Posons $\varphi_y = \varphi + y$. Alors[45]

$$\forall x \in \Omega, \quad f(x) = y \;\Leftrightarrow\; \varphi_y(x) = x.$$

On va donc appliquer le théorème de point fixe pour trouver le nombre d'antécédents de y par f dans $\overline{B}(0,r)$:

1. Comme pour tout $(x_1,x_2) \in \overline{B}(0,r)$, $\varphi_y(x_2) - \varphi_y(x_1) = \varphi(x_2) - \varphi(x_1)$, par l'étape précédente, φ_y est contractante,
2. Pour tout $x \in \overline{B}(0,r)$,
$$\|\varphi_y(x)\| = \|\varphi_y(x) - \varphi_y(0) + \varphi_y(0)\| \leq \frac{1}{2}\|x\| + \|y\| < r \tag{3.2}$$
donc $\overline{B}(0,r)$ est stable par φ_y,
3. $\overline{B}(0,r)$ est un fermé de E qui est un Banach, donc $\overline{B}(0,r)$ muni de la distance associée à la norme de E est un espace métrique complet.

donc par le théorème de point fixe, φ_y admet un unique point fixe sur $\overline{B}(0,r)$. Comme φ_y est à valeur dans $B(0,r)$, ce point fixe est en fait dans $B(0,r)$[46]. Ainsi on a montré

$$\forall y \in B(0,r/2), \exists! x \in B(0,r), \quad f(x) = y. \tag{3.3}$$

Posons désormais $\mathcal{U} = f^{-1}(B(0,r/2)) \cap B(0,r)$, qui est un ouvert par continuité de f et il contient 0 puisque $f(0) = 0$. Alors par ce qui précède, $f : \mathcal{U} \to B(0,r/2)$ est une bijection[47], qu'on note abusivement f.

Étape 4 : f^{-1} est continue : soit $(y_1,y_2) \in B(0,r/2)^2$, et posons $x_1 = f^{-1}(y_1) \in \mathcal{U}$, $x_2 = f^{-1}(y_2) \in \mathcal{U}$[48]. On a donc $\varphi_{y_1}(x_1) = x_1$ et $\varphi_{y_2}(x_2) = x_2$, donc

$$\begin{aligned}\|f^{-1}(y_2) - f^{-1}(y_1)\| &= \|x_2 - x_1\| = \|\varphi_{y_2}(x_2) - \varphi_{y_1}(x_1)\| = \|\varphi(x_2) - \varphi(x_1) + y_2 - y_1\| \\ &\leq \|y_2 - y_1\| + \frac{1}{2}\|x_2 - x_1\| = \|y_2 - y_1\| + \frac{1}{2}\|f^{-1}(y_2) - f^{-1}(y_1)\|.\end{aligned}$$

d'où

$$\|f^{-1}(y_2) - f^{-1}(y_1)\| \leq 2\|y_2 - y_1\|$$

et donc f^{-1} est 2-lipschitzienne, donc continue.

45. C'est l'idée cruciale de la preuve, qu'il faut garder en tête ; résoudre une équation $f(x) = y$ peut se voir comme une équation de point fixe simplement en écrivant qu'elle est équivalente à $x - f(x) = x - y$ c'est-à-dire à $x - f(x) + y = x$. Cette simple réécriture peut s'avérer fructueuse car il existe de multiples résultats sur les équations de point fixe ; c'est le cas ici où on utilise le théorème de point fixe de Picard.

46. Attention au jonglage entre boule ouverte et boule fermée ici ; on voudrait travailler avec des boules ouvertes car on veut se placer sur des ouverts, mais on a aussi besoin du théorème de point fixe, qui nécessite de travailler sur des boules fermées.

47. Attention à la subtilité suivante : si $f : E \to F$ et $A \subset E$ et $B \subset F$ est telle que

$$\forall y \in B, \exists! x \in A, y = f(x),$$

cela ne signifie pas que $f : A \to B$ est bijective, car on ne sait pas que $f(A) \subset B$. Donc il faut faire attention en utilisant (3.3) ; avec les notations précédentes, on peut par contre affirmer que $f : f^{-1}(B) \cap A \to B$ est bijective, et c'est ce qu'on fait ici.

48. Attention, on répète que la notation f est ici abusive, car il se peut que y_1, y_2 aient plusieurs antécédents par f ; mais ils en ont exactement un dans $\mathcal{U}$.

Étape 5 : par définition de r, on a

$$\forall x \in B(0,r), \quad \|d\varphi(x)\| = \|Id - df(x)\| \leq \frac{1}{2}$$

donc $df(x)$ est aussi un isomorphisme[49]. On peut donc appliquer le corollaire 3.5 à f[50] qui donne que $f = \mathcal{U} \to B(0,r/2)$ est un C^1-difféomorphisme. $\square$

Corollaire 3.10. *Soit (E,F) deux espaces de Banach, Ω un ouvert de E, $f : \Omega \to F$ de classe C^1 telle que pour tout $x \in \Omega$, $df(x)$ est un isomorphisme.*

1. *Alors f est une application ouverte, c'est-à-dire que pour tout $U \subset \Omega$ ouvert, $f(U)$ est un ouvert.*
2. *Si on suppose de plus que f est une fonction injective, alors $f : \Omega \to f(\Omega)$ est un C^1-difféomorphisme.*

Remarque 3.11. La partie 2 du corollaire 3.10 est parfois appelé “théorème d'inversion globale”. Ce résultat n'est en rien “magique”, car on fait une hypothèse d'injectivité très forte. La surjectivité est imposée par le fait qu'on écrit $f(\Omega)$ comme espace d'arrivée ; l'intérêt du résultat réside dans le fait que f^{-1} est de classe C^1 sur $f(\Omega)$ (qui est ouvert d'après la première partie de l'énoncé).

Démonstration.
1. Par le théorème 3.7, pour tout $x \in U$ il existe un ouvert $\mathcal{U}_x \subset U$ contenant x, et $\mathcal{V}_x$ un ouvert de F contenant $f(x)$ tel que $f : \mathcal{U}_x \to \mathcal{V}_x$ est un C^1-difféomorphisme. En particulier $f(\mathcal{U}_x) = \mathcal{V}_x$, et donc on peut écrire :

$$f(U) = f\left(\bigcup_{x\in U} \mathcal{U}_x\right) = \bigcup_{x\in U} f(\mathcal{U}_x) = \bigcup_{x\in U} \mathcal{V}_x$$

qui est un ouvert comme union d'ouverts.

2. D'après le point précédent, f est ouverte. Vue comme une fonction de Ω dans $f(\Omega)$, elle est injective et sujective, et on la suppose continue. Elle est donc un homéomorphisme (voir l'exercice 1.62 au Chapitre I) de Ω sur $f(\Omega)$. On peut donc appliquer le corollaire 3.5 et f est un C^1-difféomorphisme.

$\square$

Exercice 3.12 (Existence d'une racine k-ième dans les matrices). [51] Tiré de [BMP05, page 11]. Montrer que si $k \in \mathbb{N}^*$ et $A \in \mathcal{M}_n(\mathbb{R})$ est assez proche de I_n, alors il existe $B \in \mathcal{M}_n(\mathbb{R})$ telle que $B^k = A$.

Exercice 3.13. Tiré de [Rou03, Exercice 63.2]. On montre ici que le théorème 3.7 n'est pas valable si on suppose seulement f différentiable et non C^1 : on travaille dans $\mathbb{R}$, et on considère

$$\forall x \in \mathbb{R}, \quad f(x) = \begin{cases} x + x^2 \sin\left(\frac{1}{x}\right) & \text{si } x \neq 0 \\ 0 & \text{sinon} \end{cases}.$$

1. Montrer que f est C^1 sur $\mathbb{R}^*$.
2. Montrer que f est dérivable en 0 avec $f'(0) \neq 0$.
3. Montrer que f n'est injective sur aucun voisinage de 0.
4. Constatez que f n'est pas C^1 autour de 0, ce qui explique que le théorème 3.7 ne s'applique pas.

Remarque 3.14. Le corollaire 3.5, le théorème 3.7 et le corollaire 3.10 sont valables en remplaçant C^1 par C^k où $k \in \mathbb{N}^* \cup \{\infty\}$. Cela repose sur le fait suivant : si $f : \Omega \to \Omega'$ est un C^1-difféomorphisme et que f est de classe C^k, alors f^{-1} est de classe C^k. La preuve se fait facilement une fois qu'on a montré le caractère C^∞ de l'application Φ définie à l'équation (3.1). Voir [Car67, page 73-74] pour plus de détails.

49. Voir l'exercice 2.48 au Chapitre III.
50. Ne passez pas à côté de l'importance de l'étape 4 qui montre que $f : \mathcal{U} \to B(0,r/2)$ est un homéomorphisme, ce qui est demandé pour appliquer le corollaire 3.5...
51. En fait, le théorème d'inversion locale donne aussi la dépendance C^1 de la matrice B qui est fonction de A ; mais l'existence de la fonction f^{-1} (sans son caractère C^1) peut déjà être intéressante en soit.

Exercice 3.15. Tiré de [Rou03, Exercice 66] On considère $S_n(\mathbb{R})$ l'espace des matrices symétriques de $\mathcal{M}_n(\mathbb{R})$. Soit $A_0 \in S_n(\mathbb{R}) \cap GL_n(\mathbb{R})$; on pose

$$\forall M \in \mathcal{M}_n(\mathbb{R}), \ \ \varphi(M) = M^T A_0 M.$$

1. Montrer que φ est de classe C^1, calculer $d\varphi(I_n)$.
2. Montrer que $d\varphi(I_n)$ est surjective, et préciser son noyau.
3. Montrer qu'il existe $\mathcal{U}$ voisinage de A_0 dans $S_n(\mathbb{R})$ et $\Phi : \mathcal{U} \to GL_n(\mathbb{R})$ une application C^1 telle que
$$\forall A \in \mathcal{U}, \ \ A = \Phi(A)^T A_0 \Phi(A).$$
On pourra trouver une restriction judicieuse de φ à laquelle on peut appliquer le théorème d'inversion locale.

Exercice 3.16. Tiré de [Rou03, Exercice 70] Soit $f : \mathbb{R}^n \to \mathbb{R}^n$ une fonction de classe C^1 et telle que

$$\forall (x,y) \in (\mathbb{R}^n)^2, \ \ \|f(y) - f(x)\| \geq k\|y - x\|, \tag{3.4}$$

où $k \in]0, +\infty[$.

1. Montrer que f est injective, et que $f(\mathbb{R}^n)$ est fermée.
2. Montrer que $df(x)$ est inversible pour tout $x \in \mathbb{R}^n$.
3. Montrer que $f(\mathbb{R}^n)$ est un ouvert.
4. Conclure que f est un C^1-difféomorphisme sur $\mathbb{R}^n$.[52]

Exercice 3.17 (*Application typique en dimension infinie)**.** On pose[53]

$$E = (\{u \in C^1([0,1]), \ u(0) = 0\}, \|\cdot\|_{C^1}), \qquad \text{et} \qquad F = (C^0([0,1]), \|\cdot\|_\infty),$$

et

$$\forall u \in E, \ \ G(u) = u' + u^2.$$

1. Montrer que G est de classe C^1 sur E, et calculer dG.
2. Montrer que $dG(0)$ est un isomorphisme.
3. En déduire qu'il existe δ tel que si $f \in F$ est telle que $\|f\|_\infty \leq \delta$, alors le problème
$$\begin{cases} u' + u^2 = f & \text{sur } [0,1] \\ u(0) = 0 \end{cases}$$
admet une solution u de classe C^1.[54].

3.3 Théorème des fonctions implicites

On peut voir le théorème d'inversion locale comme un moyen de voir que l'équation $f(x) = y$, dont y est un paramètre et x l'inconnnue, a une unique solution. Le résultat est une généralisation à un type d'équation plus général de la forme $f(x,y) = 0$, où pour chaque paramètre y, on attend qu'il y ait (localement) une unique solution x.

52. Un théorème plus fort, parfois appelé théorème de Hadamard-Lévy, montre qu'on peut remplacer l'hypothèse 3.4 (qui est une sorte d'hypothèse de dilatation) par les hypothèses suivantes :

(a) pour tout $x \in \mathbb{R}^n$, $df(x)$ est inversible,

(b) la fonction f est propre, c'est-à-dire $\lim_{\|x\| \to +\infty} \|f(x)\| = +\infty$.

Voir par exemple [QZ13, page 399] avec l'hypothèse supplémentaire que f est de classe C^2 ; cela peut faire l'objet d'un développement, mais qui est difficile. Voir aussi [Ave83, page 131] pour une version sur des espaces de Banach.

53. Rappelons que si $u : [0,1] \to \mathbb{R}$ est de classe C^1, alors $\|u\|_{C^1} = \|u\|_\infty + \|u'\|_\infty$.

54. Il y a une forme d'unicité au voisinage de 0, au sens où la solution u obtenue sera bien la seule telle que $\|u\|_{C^1} \leq \delta'$ pour un certain δ'. Mais à ce stade, on ne prouve pas l'unicité de la solution.

Théorème 3.18. *Soit E_1, E_2, F trois espaces de Banach, Ω un ouvert de E_1, Ω' un ouvert de E_2, et $f : \Omega \times \Omega' \to F$ de classe C^1. On suppose $(x_0, y_0) \in \Omega \times \Omega'$ est tel que $f(x_0, y_0) = 0$ et*[55]

$$\partial_2 f(x_0, y_0) \text{ est un isomorphisme de } E_2 \text{ vers } F.$$

Alors il existe $\mathcal{W} \subset \Omega \times \Omega'$ voisinage ouvert[56] *de (x_0, y_0), $\mathcal{U} \subset \Omega$ voisinage ouvert de x_0 et $\varphi : \mathcal{U} \to F$ de classe C^1 tels que pour tout $(x, y) \in E_1 \times E_2$,*

$$(x, y) \in \mathcal{W} \text{ et } \ f(x, y) = 0 \quad \Leftrightarrow \quad x \in \mathcal{U} \quad \text{et} \quad y = \varphi(x). \tag{3.5}$$

Remarque 3.19. Ce résultat est un moyen de comprendre des ensembles définis “implicitement”, c'est-à-dire de la forme

$$\{(x, y) \in \Omega \times \Omega', \ \ f(x, y) = 0\}.$$

La raison pour laquelle on parle d'expression implicite, c'est que le lien entre x et y n'est pas explicite, au sens où si on se donne une valeur de $x \in \Omega$, il faut résoudre une équation pour trouver les valeurs de y autorisées, or on sait que ce type de résolution ne peut pas en général être fait explicitement. L'équation $f(x, y) = 0$ est dite implicite.

Remarque 3.20. Considérons le cas où $E_1 = \mathbb{R}^n$ et $E_2 = \mathbb{R}^p$. Pour que l'hypothèse “$\partial_2 f(x_0, y_0)$ est inversible” soit satisfaite, il est nécessaire que $\dim(E_2) = \dim(F)$. On suppose donc que $F = \mathbb{R}^p$. Dans ce cas, en notant $(x, y) = (x_1, \dots, x_n, x_{n+1}, \dots, x_{n+p})$, l'application linéaire $\partial_2 f(x_0, y_0)$ est représentée par la matrice

$$\Big(\partial_j f_i(x_0, y_0)\Big)_{1 \le i \le p, n+1 \le j \le n+p}$$

qui est une sous-matrice de $J_f(x_0, y_0) \in \mathcal{M}_{p,n+p}(\mathbb{R})$ (plus précisément, le “bloc carré de droite”). Ainsi, l'hypothèse d'inversibilité peut s'écrire simplement $\det\left(\big(\partial_j f_i(x_0, y_0)\big)_{1 \le i \le p, n+1 \le j \le n+p}\right) \neq 0$.

Intuitivement, on voit qu'on essaie de résoudre un système à p équations et $n + p$ inconnues. Génériquement, on s'attend donc à ce que pour un $x \in E_1$ donné, on ait “en gros” une seule solution y. Supposons par exemple que f est une application affine, c'est-à-dire $f(x, y) = A \times \begin{pmatrix} x \\ y \end{pmatrix} + b$, et comme $f(x_0, y_0) = 0$, on a

$$f(x, y) = 0 \quad \Leftrightarrow \quad A \times \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} = 0 \quad \Leftrightarrow \quad A_1 \times (x - x_0) + A_2 \times (y - y_0) = 0,$$

où on a écrit $A = \begin{pmatrix} A_1 & A_2 \end{pmatrix}$ avec $A_1 \in \mathcal{M}_{n,p}(\mathbb{R})$ et $A_2 \in \mathcal{M}_p(\mathbb{R})$. Ainsi, si A_2 est inversible, il y a bien une unique solution y, à chaque x donné. Il se trouve que dans ce cas particulier A_2 est bien la différentielle partielle de f par rapport à y.

Ainsi, l'algèbre linéaire permet de voir comment résoudre un tel système dans le cas où f est affine. L'idée du calcul différentiel est de transposer les résultats de l'algèbre linéaire aux cas où f n'est plus supposée affine ; mais cela ne fonctionne alors que localement [57], ce qui est logique puisque la proximité entre f et son développement limité à l'ordre 1 (qui est affine) n'est valable que sur un voisinage. Le théorème 3.18 formalise cette idée pour la question de la résolution d'une équation ou d'un système d'équations non linéaires.

Remarque 3.21. Il faut passer un peu de temps à comprendre (3.5). Celà implique en particulier :

1. $\varphi(x_0) = y_0$,
2. pour tout $x \in \mathcal{U}$, on a

$$f(x, \varphi(x)) = 0, \tag{3.6}$$

55. En terme de notation, vous pouvez préférer (x_1, x_2) à la place de (x, y) ; ou encore conserver (x, y) mais écrire $\partial_y f$ à la place de $\partial_2 f$. De toute façon, dans la pratique, il n'y a pas de raison que la variable qu'on exprime en fonction de l'autre soit notée y ou soit en seconde position ; passez du temps sur les exemples pour comprendre la mise en pratique de ce résultat.

56. Remarquez qu'on pourrait prendre $\mathcal{W}$ de la forme $\mathcal{U} \times \mathcal{V}$ où $\mathcal{U}$ est le voisinage ouvert de x_0 et $\mathcal{V}$ est un voisinage ouvert de y_0 dans F.

57. Le caractère local du résultat vient de l'apparition des voisinages ouverts $\mathcal{U}$ et $\mathcal{W}$.

3. plus précisément, si $x \in \mathcal{U}$, l'équation
$$f(x,y) = 0$$
d'inconnue y possède une et une seule solution telle que $(x,y) \in \mathcal{W}$[58], à savoir $\varphi(x)$,
4. on peut différentier l'équation (3.6), ce qui donne
$$\partial_1 f(x,\varphi(x)) + \partial_2 f(x,\varphi(x)) \circ d\varphi(x) = 0,$$
et donc comme x est proche de x_0 et que l'ensemble des isomorphismes de $\mathcal{L}(E_2,F)$ est un ouvert (voir l'exercice 2.48 au Chapitre III),
$$d\varphi(x) = -\big(\partial_2 f(x,\varphi(x))\big)^{-1} \circ \partial_1 f(x,\varphi(x)). \tag{3.7}$$
Deux remarques importantes sur ce point :
   (a) la première est qu'on déconseille d'apprendre par cœur la formule précédente, qui doit pouvoir se retrouver facilement à partir de (3.6) si vous maîtrisez les outils qui interviennent (différentielle partielle, différentielle d'une composée). D'ailleurs, écrivez les espaces de départ et d'arrivée des applications linéaires mises en jeu pour vérifier que la composition est bien possible,
   (b) la seconde est que ce calcul est un excellent moyen mnémotechnique pour savoir si c'est $\partial_1 f$ ou $\partial_2 f$ qui doit être supposée inversible dans le théorème 3.18 : l'hypothèse porte sur la différentielle partielle qu'on a besoin d'inverser pour pouvoir mener le calcul ![59]

Pour conclure sur ce point, si $E_1 = \mathbb{R}^n, E_2 = F = \mathbb{R}^p$, on peut écrire
$$J_f(x,y) = \Big(\partial_j f_i(x,y)\Big)_{1\le i\le p, 1\le j\le n+p} = \Big(J_{1,f}(x,y) \quad J_{2,f}(x,y)\Big)$$
où
$$J_{1,f}(x,y) = \Big(\partial_j f_i(x,y)\Big)_{1\le i\le p, 1\le j\le n} \quad \text{et } J_{2,f}(x,y) = \Big(\partial_j f_i(x,y)\Big)_{1\le i\le p, n+1\le j\le n+p}$$
et alors
$$\forall x \in \mathcal{U}, \quad J_\varphi(x) = -J_{2,f}(x,\varphi(x))^{-1} \times J_{1,f}(x,\varphi(x)).$$

Démonstration. On va déduire le théorème des fonctions implicites du théorème d'inversion locale : on pose
$$\begin{array}{rcl} G : \Omega \times \Omega' & \longrightarrow & E_1 \times F \\ (x,y) & \longmapsto & (x, f(x,y)) \end{array}$$
D'après la proposition 1.13, G est de classe C^1 et
$$\forall (x,y) \in \Omega\times\Omega', \forall (h,k) \in E_1 \times E_2, \ \ dG(x,y)(h,k) = (h, \partial_1 f(x,y)(h) + \partial_2 f(x,y)(k)).$$
En particulier, pour $(x,y) = (x_0,y_0)$, on constate que
$$\begin{aligned} \forall(\alpha,\beta) \in E_1\times F, \ \ dG(x_0,y_0)(h,k) = (\alpha,\beta) \ &\Leftrightarrow\ (h, \partial_1 f(x_0,y_0)(h) + \partial_2 f(x_0,y_0)(k)) = (\alpha,\beta) \\ &\Leftrightarrow \left\{ \begin{array}{l} h = \alpha \\ k = (\partial_2 f(x_0,y_0))^{-1}\big[\beta - \partial_1 f(x_0,y_0)(\alpha)\big] \end{array}\right. \end{aligned}$$
et donc $dG(x_0,y_0)$ est un isomorphisme de $E_1\times E_2$ sur $E_1 \times F$. Par le théorème d'inversion locale 3.7, il existe $\mathcal{W} \subset \Omega\times\Omega'$ un voisinage ouvert de (x_0,y_0) et $\widetilde{U}$ un voisinage ouvert de $(x_0,0) \in E_1\times F$ tels que $G : \mathcal{W} \to \widetilde{U}$ est un C^1-difféomorphisme. La fonction $G^{-1} : \widetilde{U} \to \mathcal{W}$ s'écrit alors $G^{-1}(x,y) = (x,\psi(x,y))$ avec $\psi : \widetilde{U} \to G$ de classe C^1. On pose $\mathcal{U} = \widetilde{U} \cap (E_1 \times \{0\})$ et $\varphi(x) = \psi(x,0)$ pour $x \in \mathcal{U}$. Comme $(x_0,0) \in \widetilde{U}$, l'ensemble $\mathcal{U}$ est un voisinage ouvert de x_0 ; de plus, par définition on a
$$\left\{\begin{array}{l} (x,y)\in\mathcal{W} \\ f(x,y) = 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} (x,y)\in\mathcal{W} \\ G(x,y) = \begin{pmatrix} x \\ 0 \end{pmatrix} \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} (x,0)\in\widetilde{U} \\ \begin{pmatrix} x \\ y \end{pmatrix} = G^{-1}(x,0) = \begin{pmatrix} x \\ \psi(x,0) \end{pmatrix} \end{array}\right. \Leftrightarrow \left\{\begin{array}{l} x\in\mathcal{U} \\ y = \varphi(x) \end{array}\right.$$
d'où le résultat. □

58. Attention, l'unicité est valable seulement sous cette condition $(x,y) \in \mathcal{W}$, c'est-à-dire qu'il pourrait y avoir d'autres solutions $y \in \Omega'$.
59. On ne fait l'hypothèse qu'en (x_0,y_0), mais cela reste vrai sur un voisinage.

Remarque 3.22. Comme à la remarque 3.14, le théorème 3.18 est valable en remplaçant C^1 par C^k où $k \in \mathbb{N}^* \cup \{\infty\}$. On peut le voir en reprenant la preuve et en utilisant une version C^k du théorème d'inversion locale 3.7, ou également en regardant l'équation 3.7.

Remarque 3.23. On a montré d'abord le théorème d'inversion locale, puis on en a déduit le théorème des fonctions implicites. On peut procéder dans l'autre sens : voir par exemple [Wag12a] où il est montré d'abord le théorème des fonctions implicites (voir aussi [Rou03, Exercice 86]), et on en déduit le théorème d'inversion locale : en effet, on peut voir ce dernier comme la résolution de l'équation $y - f(x) = 0$ dans un voisinage de (x_0, y_0).

Remarque 3.24. On s'est concentré sur une équation de la forme $f(x, y) = 0$, mais bien sûr on peut traiter le cas d'une équation de la forme $f(x, y) = c$ avec $c \in G$ en considérant la fonction $\widetilde{f}(x, y) = f(x, y) - c$.

4 Sous-variétés de $\mathbb{R}^n$

[60]

On cherche à définir la notion de sous-ensemble de classe C^1 de $\mathbb{R}^n$ de dimension $p \in [\![0, n]\!]$, qui s'appellera "sous-variété de $\mathbb{R}^n$ de dimension p". Un exemple naturel est bien sûr la notion de sous-espace vectoriel de dimension p dans $\mathbb{R}^n$, ou plus naturellement encore, de sous-espace affine ; on cherche donc généraliser ce concept issu de l'algèbre linéaire, grâce au calcul différentiel.

Commençons par penser au cas du cercle dans $\mathbb{R}^2$ qui correspond à un cas particulier de $n = 2$ et $p = 1$. On voit qu'il y a plusieurs façons de le définir :

$$\begin{aligned} C &= \{(x, y) \in \mathbb{R}^2,\ x^2 + y^2 = 1\} \\ &= \{(\cos\theta, \sin\theta), \theta \in [0, 2\pi]\} \\ &= \left\{(x, \sqrt{1 - x^2}), x \in [-1, 1]\right\} \cup \left\{(x, -\sqrt{1 - x^2}), x \in [-1, 1]\right\}. \end{aligned}$$

Ces trois définitions ont leurs avantages et leurs inconvénients ; on appellera la première la définition implicite de C, la seconde la version paramétrique de C, la 3ième la description sous forme de graphe.

On pourrait être tenté de généraliser chacune de ces définitions pour définir une courbe de classe C^1 dans $\mathbb{R}^2$:

- l'ensemble $\{f(x, y) = 0, (x, y) \in \mathbb{R}^2\}$ où f est de classe C^1,
- l'ensemble $\{(x(t), y(t)), t \in I\}$ où x et y sont de classe C^1 sur un intervalle $I \subset \mathbb{R}$,
- une union finie de k ensembles de la forme $\{(x, \varphi_i(x)), x \in [a_i, b_i]\}$ où les fonctions $\varphi_i : [a_i, b_i] \to \mathbb{R}$ sont de classe C^1 pour $i \in [\![1, k]\!]$.

Mais l'analyse d'exemples montrent que ces définitions ne sont pas correctes, car elles peuvent faire apparaître des singularités [61] ; un autre problème est de faire référence à des fonctions $(f, (x(\cdot), y(\cdot)), (\varphi_i)_{i \in [\![1,k]\!]})$ (qui n'ont aucune raison d'être unique) qui permettent de décrire l'ensemble qu'on considère, ce qui donne à ces tentatives de définitions un côté non intrinsèque. On souhaiterait plutôt faire référence directement à l'ensemble lui-même pour donner une définition qui ne dépend pas d'une paramétrisation.

L'objectif de ce paragraphe est double :

60. Le lecteur pourrait se demander pourquoi on parle de sous-variété, mais pas de variété. En fait, les notions de ce paragraphe se généralisent, et on peut parler de sous-variété d'une variété M donnée : ici on se restreint au cas où la variété M est simplement $\mathbb{R}^n$. C'est une restriction du programme qu'on conseille de suivre ; disons simplement que la notion de variété permet de voir un objet de "dimension" n sans faire référence à un espace qui le contiendrait (ici on définit un objet de dimension p "vu dans $\mathbb{R}^n$", donc on fait référence à un espace ambiant) ; c'est un peu point de vue "encore plus" intrinsèque. Mais dans le cadre de l'agrégation, ça mènerait à des considérations certainement trop éloignées du programme.

61. À titre d'exemple, tracez les ensembles

$$\{xy = 0, (x, y) \in \mathbb{R}^2\}, \quad \{(t^2, t^3), t \in \mathbb{R}\}, \quad \{(x, \sqrt{1 - x^2}), x \in [-1, 1]\} \cup \{(x, 0), x \in [1, 2]\}$$

1. proposer une définition plus conforme à l'intuition géométrique d'un objet régulier de dimension p dans $\mathbb{R}^n$,
2. voir comment on peut effectivement donner des définitions équivalentes sous forme implicite, paramétrique, ou en tant que graphe.

4.1 Définition

Définition 4.1. Soit $X \subset \mathbb{R}^n$ non vide, et $p \in [\![0, n]\!]$. On dit que X est une sous-variété de $\mathbb{R}^n$, de dimension p et de classe C^1, si pour tout $x_0 \in X$,

— il existe $\mathcal{U}_{x_0}$ voisinage ouvert de x_0 dans $\mathbb{R}^n$,

— il existe $\mathcal{V}_0$ voisinage ouvert de 0 dans $\mathbb{R}^n$,

— il existe $\psi : \mathcal{V}_0 \to \mathcal{U}_{x_0}$ un C^1-difféomorphisme,

tels que

$$\psi(0) = x_0, \qquad \text{et} \qquad \psi\Big[(\mathbb{R}^p \times \{0\}^{n-p}) \cap \mathcal{V}_0\Big] = X \cap \mathcal{U}_{x_0}.$$

Remarque 4.2. Dit en des termes moins techniques, un ensemble C^1 de dimension p dans $\mathbb{R}^n$ est vu comme étant localement la déformation régulière (représentée par la notion de C^1-difféomorphisme) d'un ensemble "plat" de la forme $\mathbb{R}^p \times \{0\}^{n-p}$.

Remarque 4.3. Les cas $p = 0$ et $p = n$ sont "extrêmes" mais autorisés :

— dans le cas $p = 0$, on voit que X est fait de points isolés (car la condition devient $X \cap \mathcal{U}_{x_0} = \psi((0, \ldots, 0))$ et donc X ne contient qu'un seul point (à savoir x_0) dans $\mathcal{U}_{x_0}$.

— dans le cas $p = n$, on voit que X est simplement un ouvert de $\mathbb{R}^n$ (la condition devient $X \cap \mathcal{U}_{x_0} = \psi(\mathcal{V}_0) = \mathcal{U}_{x_0}$).

Remarque 4.4. Si on prend $\mathcal{U}_{x_0} = \mathbb{R}^n = \mathcal{V}_0$ et $\psi = Id$, l'espace $\mathbb{R}^p \times \{0\}^{n-p}$ est bien une sous-variété de dimension p dans $\mathbb{R}^n$.

Remarque 4.5. Quand $p = 1$ (et $n \geq 2$), on parle de courbe de classe C^1, et quand $p = 2$ (et $n \geq 3$), on parle de surface de classe C^1.

4.2 Caractérisations

On en vient à la caractérisation d'une sous-variété avec les différents types de descriptions évoqués précédemment :

Théorème 4.6 (Caractérisation des sous-variétés de $\mathbb{R}^n$)**.** *Soit $X \subset \mathbb{R}^n$ non vide : les propriétés suivantes sont équivalentes :*

1. *X est une sous-variété de $\mathbb{R}^n$ de dimension p et de classe C^1,*
2. *pour tout $x_0 \in X$, il existe $\mathcal{U}_{x_0}$ voisinage ouvert de x_0 dans $\mathbb{R}^n$ et $f : \mathcal{U}_{x_0} \to \mathbb{R}^{n-p}$ de classe C^1 telle que*[62]
$$df(x_0) \text{ est surjective, et } X \cap \mathcal{U}_{x_0} = f^{-1}(0) = \{x \in \mathcal{U}_{x_0}, f(x) = 0\},$$
3. *pour tout $x_0 \in X$, il existe une décomposition $\mathbb{R}^n = \mathbb{R}^p \oplus \mathbb{R}^{n-p}$ dans laquelle on écrit $x_0 = (x_0^1, x_0^2)$, il existe $\mathcal{U}^1$ voisinage ouvert de $x_0^1 \in \mathbb{R}^p$, $\mathcal{V}^2$ voisinage ouvert de x_0^2 et $\varphi : \mathcal{U}^1 \to \mathcal{V}^2$ de classe C^1 telle que*
$$\varphi(x_0^1) = x_0^2, \quad \text{et} \quad X \cap (\mathcal{U}^1 \times \mathcal{V}^2) = \{(x^1, \varphi(x^1)), x^1 \in \mathcal{U}^1\},$$
4. *pour tout $x_0 \in X$, il existe $\mathcal{U}_{x_0}$ voisinage ouvert de $x_0 \in \mathbb{R}^n$, $\mathcal{V}_0$ voisinage ouvert de 0 dans $\mathbb{R}^p$ et $\Gamma : \mathcal{V}_0 \to \mathbb{R}^n$ de classe C^1 telle que*[63]
$$\Gamma(0) = x_0, \ \ \Gamma \text{ est un homéomorphisme de } \mathcal{V}_0 \text{ sur } X \cap \mathcal{U}_{x_0} \text{ et } d\Gamma(0) \text{ est injective.}$$

62. On dit que f est une submersion C^1 en x_0.
63. On dit que Γ est une immersion C^1 en x_0.

Remarque 4.7. Cet énoncé permet de voir comment corriger les problèmes évoqués précédemment :

— pour la représentation implicite, il faut ajouter que $df(x_0)$ soit surjective en tout x_0 pour avoir un ensemble de classe C^1

— avec la représentation en graphe, il faut choisir un système de coordonnées qui peut évoluer en fonction du point autour duquel on travaille,[64]

— pour la représentation paramétrique, il faut ajouter que $d\Gamma(0)$ soit injective si Γ est la paramétrisation définie sur un voisinage de $0 \in \mathbb{R}^p$ et dont l'image est X intersectée avec un voisinage de x_0.

Remarque 4.8. Pour ne pas s'y perdre sur les variables et leurs dimensions, pensez au sens de ces variables et au parallèle avec l'algèbre linéaire :

— dans la représenation implicite, on a $x \in \mathbb{R}^n$, et comme on veut "réduire" cette variable de dimension n afin de décrire un objet de dimension p, on doit mettre $n-p$ contraintes sur x, contraintes qui sont modélisées par f. La condition de surjectivité de $df(x_0)$ correspond au fait que ces contraintes sont libres, c'est-à-dire non-redondantes[65],

— dans la représentation graphe comme dans la représentation paramétrique[66], on veut décrire l'ensemble X à l'aide des fonctions Γ et φ, qui sont donc naturellement définies sur un espace de départ de dimension p. Dans le cas graphe, c'est $(x^1, \varphi(x^1))$ qui décrit l'ensemble $X \subset \mathbb{R}^n$, donc naturellement φ est à valeurs dans $\mathbb{R}^{n-p}$; dans le cas paramétrique, X est vu (localement) comme l'image de Γ, qui est donc à valeurs dans $\mathbb{R}^n$. L'hypothèse d'injectivité de $d\Gamma(0)$ permet que les points $\Gamma(t)$ pour t proche de 0 soient bien distincts[67].

Remarque 4.9. On peut naturellement définir la notion de sous-variété de classe C^k de $\mathbb{R}^n$, avec $k \in \mathbb{N}^* \cup \{+\infty\}$, en demandant dans la définition 4.1 que la fonction ψ soit un C^k difféomorphisme. Dans ce cas le théorème 4.6 reste valable en précisant que les fonctions qui interviennent (f, φ, Γ) sont de classe C^k.

Remarque 4.10. Pour les conditions "$df(x_0)$ est surjective" et "$d\Gamma(0)$ est injective", on dit parfois que ces différentielles doivent être de rang maximal, ce qui correspond bien à l'hypothèse dans les deux cas. Notez qu'il existe un "théorème du rang constant" qui généralise les éléments de démonstration ci-dessous et permet de comprendre les fonctions f telles que $df(x)$ est de rang constant, et qui généralise les notions d'immersion et de submersion.

Éléments de démonstration

— 1 ⇒ 2 : on considère $\psi^{-1} : x \in \mathbb{R}^n \mapsto (h_1(x), h_2(x)) \in \mathbb{R}^p \times \mathbb{R}^{n-p}$. La fonction $f = h_2$ convient car
$$\psi^{-1}(X \cap \mathcal{U}_{x_0}) = \mathbb{R}^p \times \{0\}^{n-p} = \{x \in \mathbb{R}^n, h_2(x) = 0\}.$$

— 2 ⇒ 3 Il s'agit du théorème des fonctions implicites.

— 3 ⇒ 4 On peut considérer
$$\Gamma(x^1) = \begin{pmatrix} x^1 + x_0^1 \\ \varphi(x^1 + x_0^1) \end{pmatrix}.$$

— 4 ⇒ 1 On peut considérer
$$\psi(x^1, x^2) = \begin{pmatrix} \Gamma_1(x^1) \\ x^2 + \Gamma_2(x^2) \end{pmatrix},$$
mais pour que l'hypothèse soit satisfaite, il faudrait réordonner les variables au préalable (voir [Rou03, Exercice 73]).

64. Dans le cas du cercle, il avait suffit de mettre deux graphes (partie supérieure et partie inférieure du cercle), mais alors le "recollement" C^1 des deux morceaux était un peu miraculeux ; on aurait pu s'en sortir en donnant d'autres conditions, mais parce que le recollement se restreignait à un nombre fini de points (deux en l'occurrence). Cette approche ne se généralise pas bien aux cas de dimension supérieure, et donc on préfère (quitte à augmenter le nombre de graphes nécessaires) demander que les graphes décrivent des morceaux qui vont se chevaucher en partie.

65. Penser à $f(x, y, z) = (x + y + z, x + y + z)$ ou $f(x, y, z) = (x + y + z, x + y + z + 1)$ qui ne donneront pas des ensembles de dimension 1 dans $\mathbb{R}^3$

66. En fait, comme on l'évoquera dans le schéma de preuve ci-dessous, la version graphe est un cas particulier d'une représentation paramétrique.

67. Penser à $\Gamma : t \in \mathbb{R} \mapsto 0 \in \mathbb{R}^2$ qui ne décrit pas une courbe de $\mathbb{R}^2$, ou $\Gamma : (t_1, t_2) \in \mathbb{R}^2 \mapsto (t_1, t_1, 0)$ qui ne décrit pas une surface de $\mathbb{R}^3$.

4.3 Exemples

En premier lieu, on conseille fortement le lecteur à se familiariser avec le théorème 4.6 dans les cas de petites dimensions c'est-à-dire $(n,p)=(2,1)$, $(n,p)=(3,1)$ et $(n,p)=(3,2)$ et de les mettre en pratique dans des cas simples ; voir à ce sujet [Rou03, Exercices 88,90,91]. On conseille surtout de le faire avant d'aborder des exemples plus élaborés.

Exemple 4.11. L'ensemble

$$\mathbb{S}^{n-1}=\Big\{(x_1,\dots,x_n)\in\mathbb{R}^n,\ \sum_{k=1}^n x_k^2=1\Big\}$$

est une sous-variété de $\mathbb{R}^n$ de dimension n^2-1. On l'appelle sphère (unité) de $\mathbb{R}^n$.

Pour justifier ce fait, il suffit de voir que la fonction $f : x\in\mathbb{R}^n\mapsto\sum_{k=1}^n x_k^2$ est telle que pour tout $x\in\mathbb{S}^{n-1}$, $df(x)$ est surjective. En l'occurrence, $df(x)\in\mathcal{L}(\mathbb{R}^n,\mathbb{R})$ est surjective si et seulement si elle est non nulle ; on peut représenter $df(x)$ par son gradient si on travaille dans la base canonique, ce qui donne

$$\forall x\in\mathbb{S}^{n-1},\quad \nabla f(x)=\begin{pmatrix}2x_1\\ \vdots\\ 2x_n\end{pmatrix}=2x\neq 0.$$

En effet le seul point où $\nabla f(x)$ s'annule est 0 qui n'est pas dans $\mathbb{S}^n$.

Exemple 4.12. On peut montrer que les ensembles

$$SL_n(\mathbb{R})=\{A\in\mathcal{M}_n(\mathbb{R}),\ \det(A)=1\},\ \ O_n(\mathbb{R})=\{A\in\mathcal{M}_n(\mathbb{R}),\ A^TA=I_n\}$$

sont des sous-variétés de $\mathcal{M}_n(\mathbb{R})\simeq\mathbb{R}^{n^2}$ de dimension n^2-1 et $n(n-1)/2$ respectivement.

Ces exemples ouvrent la porte à un sujet bien vaste et délicat, que sont les groupes de Lie ; le livre [MT86] reste raisonnablement accessible et permet d'aborder le sujet, mais cela doit rester réservé au candidat très ambitieux, car on a vite fait de se faire avoir avec des notions délicates.

4.4 Espace tangent

Définition 4.13. Soit X une sous-variété de $\mathbb{R}^n$ de classe C^1. On dit que $h\in\mathbb{R}^n$ est tangent à X au point x_0 s'il existe $\delta>0$ et $\gamma :]-\delta,\delta[\to\mathbb{R}^n$ de classe C^1 et tel que

$$\gamma(]-\delta,\delta[)\subset X,\quad \gamma(0)=x_0,\quad \gamma'(0)=h.$$

On pose $T_{x_0}X$ l'ensemble des vecteurs tangents à X en x_0.

Proposition 4.14. *Soit X est une sous-variété de $\mathbb{R}^n$ de classe C^1 et de dimension p, et $x_0\in X$. Alors*

- *$T_{x_0}X$ est un sous-espace vectoriel de $\mathbb{R}^n$ de dimension p,*
- *avec les notations du théorème 4.6 :*

$$T_{x_0}X=d\psi(0)(\mathbb{R}^p\times\{0\}^{n-p})=Ker(df(x_0))=\{(x^1,x^2),\ x^2=d\varphi(x_0^1)x^1\}=Im(d\Gamma(0))$$

Remarquons que le premier point suit du second. Pour la preuve du second, on invite à regarder les références.

5 Optimisation

On aborde quelques éléments du domaine de l'optimisation qui sont liés au calcul différentiel.

5.1 Convexité

Soit E un espace vectoriel normé.

Définition 5.1. Soit C un ensemble convexe [68] de E, et $f : C \to \mathbb{R}$.

— La fonction f est dite convexe si

$$\forall (x,y) \in C^2,\ \forall t \in [0,1],\quad f((1-t)x + ty) \leq (1-t)f(x) + tf(y). \tag{5.1}$$

— La fonction f est dite strictement convexe si [69]

$$\forall (x,y) \in C^2 \text{ tels que } x \neq y,\ \forall t \in]0,1[,\quad f((1-t)x + ty) < (1-t)f(x) + tf(y).$$

Exercice 5.2. Une première utilisation simple de la stricte convexité est sur l'unicité du minimiseur. Soit E un espace vectoriel normé, et $C \subset E$. Soit $f : C \to \mathbb{R}$.

1. On suppose C un ensemble convexe, et f strictement convexe. Montrer qu'il y a unicité de $x_0 \in C$ tel que
$$f(x_0) = \inf \Big\{ f(x), x \in C \Big\}.$$
2. Donner un exemple où C est convexe, f est convexe, mais il n'y a pas unicité d'un minimiseur de f sur C.
3. Donner un exemple où $f : E \to \mathbb{R}$ est strictement convexe, mais il n'y a pas unicité d'un minimiseur de f sur C (d'où l'importance de supposer C convexe).

Exercice 5.3 (Lien convexité d'une fonction/convexité d'un ensemble)**.** Soit C un ensemble convexe de E, et $f : C \to \mathbb{R}$. On définit

$$Epi(f) := \{(x, \lambda) \in C \times \mathbb{R},\quad f(x) \leq \lambda\}$$

qu'on appelle épigraphe de f. Montrer que f est convexe si et seulement si $Epi(f)$ est un sous-ensemble convexe de $E \times \mathbb{R}$.

Proposition 5.4 (Caractérisation de la convexité)**.** *Soit Ω un ouvert convexe de E, et $f : \Omega \to \mathbb{R}$.*

1. *si f est différentiable sur Ω, alors*
$$\begin{aligned} f \text{ est convexe} \quad &\Longleftrightarrow \quad \forall (x,y) \in \Omega^2,\ f(y) \geq f(x) + df(x)(y-x) \\ &\Longleftrightarrow \quad \forall (x,y) \in \Omega^2, (df(y) - df(x))(y-x) \geq 0 \end{aligned}$$
2. *on suppose f deux fois différentiable sur Ω. Alors* [70]
   (a) *f est convexe $\Leftrightarrow$ $\forall x \in \Omega, d^2 f(x) \geq 0$,*
   (b) *$\forall x \in \Omega, d^2 f(x)$ est définie positive $\Rightarrow$ f est strictement convexe.*

Remarque 5.5. La caractérisation $\forall (x,y) \in \Omega^2, (df(y) - df(x))(y-x) \geq 0$ généralise le fait que f' est croissante dans le cas où $E = \mathbb{R}$ et $C = I$ est un intervalle de $\mathbb{R}$ [71]. En effet, la croissance de f' est bien équivalente au fait que $(f'(y) - f'(x)).(y-x)$ soit positif ou nul pour tout $(x,y) \in I^2$.

---

68. Attention, on a besoin que l'ensemble de départ soit convexe, car comme on va évaluer f en $(1-t)x + ty$, on a besoin que ce dernier soit dans C. Retenez que dès que vous souhaitez utiliser un argument de convexité, il faut que l'ensemble de travail soit convexe.
69. Soyez vigilant ! La formule (5.1) a en effet des cas d'égalité triviaux, à savoir $x = y$ ou $t \in \{0,1\}$. La stricte convexité consiste à demander que ces cas d'égalité soient les seuls.
70. Voir la remarque 5.6 pour les définitions de positivité d'une forme quadratique.
71. *D'ailleurs, on dit parfois d'un opérateur $T : \Omega \to E'$ (avec $\Omega \subset E$) qui satisfait
$$\forall (x,y) \in \Omega^2,\ (T(y) - T(x))(y-x) \geq 0$$
qu'il est monotone. Attention donc à la terminologie : une fonction $g : I \to \mathbb{R}$ est croissante si et seulement si l'opérateur associé $\widetilde{g} : x \in I \mapsto (h \mapsto g(x)h) \in \mathbb{R}'$ est monotone.
Si $E = \mathbb{R}^n$ est muni de son produit scalaire usuel, et si on assimile E' à $\mathbb{R}^n$ en travaillant dans la base canonique, la monotonie de df peut s'écrire :
$$\forall (x,y) \in \mathbb{R}^n,\ \langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0.$$

Remarque 5.6. Remarquons que $d^2f(x) \geq 0$ signifie

$$\forall h \in E, \;\; d^2f(x)(h,h) \geq 0.$$

Aussi, $d^2f(x)$ est définie positive signifie

$$\forall h \in E \setminus \{0\}, \;\; d^2f(x)(h,h) > 0.$$

Remarque 5.7. Il est classique que le dernier énoncé (2b) n'est pas une équivalence : la fonction $x \in \mathbb{R} \mapsto x^4$ est strictement convexe, mais sa dérivée seconde s'annule en 0.

Démonstration. 1. On suppose f différentiable sur Ω.

— Supposons f convexe sur Ω. Soit $(x,y) \in \Omega^2$. Pour $t \in]0,1[$, on a

$$f((1-t)x+ty) \leq (1-t)f(x)+tf(y)$$

ce qui se réécrit

$$\frac{f(x+t(y-x))-f(x)}{t} \leq f(y)-f(x)$$

En fait tendre t vers 0, et on obtient

$$df(x)(y-x) \leq f(y)-f(x).$$

— Supposons

$$\forall (x,y) \in \Omega^2, \quad df(x)(y-x) \leq f(y)-f(x).$$

En échangeant les rôles de x et y et en additionnant les inégalités, on obtient

$$\forall (x,y) \in \Omega^2, \quad (df(y)-df(x))(y-x) \geq f(x)-f(y)+f(y)-f(x) = 0.$$

— Supposons

$$\forall (x,y) \in \Omega^2, \quad (df(y)-df(x))(y-x) \geq 0.$$

Soit $x, y \in \Omega^2$. On pose $\varphi : t \in [0,1] \mapsto f(x+t(y-x))$. Alors φ est dérivable sur $[0,1]$ et

$$\forall t \in [0,1], \; \varphi'(t) = df(x+t(y-x))(y-x).$$

Ainsi pour tout $t \in [0,1]$, $\varphi'(t) \geq \varphi'(0)$. Par le théorème des accroissements finis appliqué à φ sur $[0,1]$, on obtient qu'il existe $c \in]0,1[$ tel que

$$\varphi(1)-\varphi(0) = \varphi'(c) \geq \varphi'(0), \qquad \text{i.e.} \qquad f(y)-f(x) \geq df(x)(y-x),$$

et ce pour tout $(x,y) \in \Omega^2$.

— [72]On suppose

$$\forall (x,y) \in \Omega^2, f(y) \geq f(x)+df(x)(y-x).$$

Soit $(x,y) \in \Omega^2$ et $t \in [0,1]$. On pose $x_t = (1-t)x+ty$ et on applique l'hypothèse aux couples (x,x_t) et (y,x_t) :

$$f(x) \geq f(x_t)+df(x_t)(x-x_t), \qquad \text{et} \qquad f(y) \geq f(x_t)+df(x_t)(y-x_t).$$

En multipliant la première inégalité par $(1-t)$ et la seconde par t puis en additionnant, on obtient

$$(1-t)f(x)+tf(y) \geq (1-t+t)f(x_t)+df(x_t)[(1-t)x+ty-(1-t+t)x_t) = f(x_t)+df(x_t)(x_t-x_t) = f(x_t)$$

d'où la convexité de f.

72. Remarquez qu'on n'a pas fait une preuve circulaire $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 1$ mais qu'on a montré $1 \Rightarrow 2 \Rightarrow 3$ et $3 \Rightarrow 2 \Rightarrow 1$, ou encore $1 \Leftrightarrow 2$ et $2 \Leftrightarrow 3$; on peut probablement faire une preuve plus directe de $3 \Rightarrow 1$, mais ce n'est pas sûr qu'on y gagne beaucoup.

2. On suppose f deux fois différentiables sur Ω.
   (a) Supposons f convexe. Soit $x \in \Omega$, et $h \in E$. Alors pour $t > 0$ petit, $x + th \in \Omega$, donc on peut écrire
   $$(df(x + th) - df(x))(h) \geq 0$$
   d'après la première partie de l'énoncé. En divisant par $t > 0$ et en faisant tendre t vers 0, on obtient
   $$d^2 f(x)(h, h) \geq 0.$$
   (b) On suppose
   $$\forall x \in \Omega, \forall h \in E^2, \ \ d^2 f(x)(h, h) \geq 0.$$
   On pose $\varphi : t \in [0, 1] \mapsto f(x + t(y - x))$. Alors φ est deux fois dérivables sur $[0, 1]$, et
   $$\forall t \in [0, 1], \ \varphi'(t) = df(x + t(y - x))(y - x), \quad \varphi''(t) = d^2 f(x + t(y - x))(y - x, y - x) \geq 0$$
   Par la formule de Taylor-Lagrange (voir la proposition 3.53 et la note 92 au Chapitre I [73]), il existe $c \in]0, 1[$ tel que
   $$\varphi(1) - \varphi(0) = \varphi'(0) + \frac{1}{2} \underbrace{\varphi''(c)}_{\geq 0}, \qquad \text{d'où} \qquad f(y) - f(x) \geq df(x)(y - x),$$
   d'où le résultat par la première partie de l'énoncé.

□

Le corollaire suivant est fondamental dans l'utilisation de la convexité.

Corollaire 5.8. *Soit Ω un ouvert convexe de E, et $f : \Omega \to \mathbb{R}$ convexe et différentiable sur Ω. Si $df(x_0) = 0$, alors x_0 est solution de*
$$f(x_0) = \inf \Big\{ f(x), x \in \Omega \Big\}.$$

Démonstration. On utilise la première caractérisation dans la proposition 5.4, ce qui donne
$$\forall x \in \Omega, \ f(x) \geq f(x_0) + df(x_0)(x - x_0) = f(x_0)$$
d'où le résultat. □

Remarque 5.9. Cet énoncé est caractéristique de l'importance de la convexité en optimisation ; il donne une réciproque à la proposition 5.14, ce qui donne une équivalence entre un problème d'optimisation et une équation, appelée équation d'optimalité d'ordre 1 ou dans certains contextes équations d'Euler-Lagrange. Un exemple célèbre d'une telle correspondance, est le problème de Dirichlet : étant donné un ouvert régulier [74] Ω de $\mathbb{R}^n$, on considère les problèmes

$$u \text{ solution de } \min \left\{ \int_\Omega \left[\frac{1}{2} |\nabla v(x)|^2 - f(x) v(x) \right] dx, v : \overline{\Omega} \to \mathbb{R}, \ \forall x \in \partial\Omega, v(x) = 0 \right\}$$
$$\overset{?}{\Leftrightarrow} \quad \left\{ \begin{array}{rcll} -\Delta u(x) & = & f(x) & \text{pour tout } x \in \Omega \\ u(x) & = & 0 & \text{pour tout } x \in \partial\Omega \end{array} \right.$$

où $f : \Omega \to \mathbb{R}$ est donnée, et $u : \overline{\Omega} \to \mathbb{R}$ est l'inconnue [75]. Formellement, les deux problèmes sont bien équivalents, mais il reste la difficulté de savoir dans quel espace se trouve u, ou plus précisément quelle est sa régularité. En effet, dans la première formulation, il semble naturel de supposer u de classe C^1 sur Ω, alors que dans la seconde, on aimerait avoir u de classe C^2.

73. En fait, en accord avec la remarque 1.29, nous montrons ici que la formule d'Euler-Lagrange est valable pour une fonction définie sur un ouvert d'un espace vectoriel normé, et à valeurs dans $\mathbb{R}$, à condition que le segment $[x, y]$ soit dans Ω. Remarquons aussi qu'on peut simplifier certaines des preuves présentées ici si on considère qu'on les a déjà montrées pour des fonctions définies sur un intervalle de $\mathbb{R}$ (voir le paragraphe 6.2.2 au Chapitre I.
74. On ne précise pas la signification de ce mot ici, vu qu'on fait une remarque plutôt d'ordre culturel.
75. Rappelons si besoin la notation classique $\Delta u(x) = \sum_{i=1}^n \partial_{ii}^2 u(x)$.

Si on suppose que $u : \overline{\Omega} \to \mathbb{R}$ est de classe C^2 sur Ω, et est continue sur $\overline{\Omega}$, alors on peut effectivement montrer (soit à la main, soit en utilisant la proposition 5.14 et le corollaire 5.8) que u est solution du problème d'optimisation si et seulement si elle est solution de l'équation aux dérivées partielles associée.

Ce problème est également historiquement important du fait des développements qu'il a amené en analyse fonctionnelle ; en effet, la convexité du problème d'optimisation a également son importance en ce qui concerne la question d'existence d'une solution au problème, et a motivé le développement de ce qu'on appelle aujourd'hui les espaces de Sobolev.

Notons qu'il est possible d'évoquer ce problème dans le cadre du programme de l'agrégation, mais qu'il est conseillé de le faire lorsque $n = 1$, c'est-à-dire que Ω est un intervalle de $\mathbb{R}$; voir le Chapitre IX.

Exercice 5.10. Soit $\alpha \in]0, +\infty[$. On peut définir la notion de α-convexité : étant donné C un ensemble convexe, on dit que $f : C \to \mathbb{R}$ est α-convexe si et seulement si

$$\forall (x,y) \in C, \forall t \in [0,1], \quad f((1-t)x + ty) \leq (1-t)f(x) + tf(y) - \alpha \frac{t(1-t)}{2} \|x-y\|^2.$$

1. Montrer que si $f : C \to \mathbb{R}$ est α-convexe, alors f est strictement convexe.
2. On suppose que $(E, \langle \cdot, \cdot \rangle)$ est un espace préhilbertien ; établir pour quelles valeurs de $\beta > 0$ la fonction $x \in E \mapsto \|x\|^2$ est β-convexe.
3. Soit Ω un ouvert convexe, et $f : \Omega \to \mathbb{R}$.
   (a) On suppose f différentiable sur Ω. Montrer que

$$\begin{aligned} f \text{ est } \alpha\text{-convexe} \quad &\Leftrightarrow \quad \forall (x,y) \in \Omega^2, \quad f(y) \geq f(x) + df(x)(y-x) + \frac{\alpha}{2}\|y-x\|^2, \\ &\Leftrightarrow \quad \forall (x,y) \in \Omega^2, (df(y) - df(x))(y-x) \geq \alpha \|y-x\|^2 \end{aligned}$$

   (b) On suppose f deux fois différentiable sur Ω. Montrer que [76]

$$f \text{ est } \alpha\text{-convexe} \quad \Leftrightarrow \quad \forall x \in \Omega, \ \forall h \in E^2 \ \ d^2 f(x)(h,h) \geq \alpha \|h\|^2.$$

5.2 Conditions d'optimalité

Définition 5.11. Soit $\Omega \subset E$, et $f : \Omega \to \mathbb{R}$.

— On dit que x_0 est un minimum (ou minimum global ou minimum absolu) de f si

$$\forall x \in \Omega, \ f(x_0) \leq f(x).$$

On peut aussi dire que x_0 est une solution du problème d'optimisation

$$\inf \Big\{ f(x), \quad x \in \Omega \Big\}.$$

— On dit que $x_0 \in \Omega$ est un minimum local de f s'il existe $\delta > 0$ tel que

$$\forall x \in B(x_0, \delta) \cap \Omega, \quad f(x) \geq f(x_0).$$

On peut définir de façon similaire un maximum global et un maximum local. On parlera d'extremum global ou local pour désigner les points qui sont soit des maxima soit minima globaux ou locaux respectivement.

Remarque 5.12. Remarquons qu'un minimum global est en particulier un minimum local.

Attention à l'emploi des mots minimum/maximum/extremum sans adjectif ; a priori on sous-entend qu'il est global si on ne précise pas ; mais alors avec cette terminologie un minimum local n'est pas forcément un minimum ! Si vous souhaitez éviter toute ambiguïté, n'utilisez pas ces mots sans adjectif.

76. On pourrait noter $d^2 f(x) \geq \alpha Id$ pour $\forall h \in E, \ d^2 f(x)(h,h) \geq \alpha \|h\|^2$.

Remarque 5.13. Remarquons que si x_0 est un maximum local/global de f, alors x_0 est un minimum local/global de $-f$. Pour cette raison, on présente en général les résultats uniquement pour les problèmes de minimisation (c'est ce qu'on fait dans la suite), mais à d'éventuels changements de signe près, tout s'applique aux problèmes de maximisation.

La recherche de minima est une démarche naturelle dans beaucoup de problèmes, mais est aussi souvent délicate. Le calcul différentiel permet de donner des moyens d'identification de tels minima ; mais sauf dans le cas d'hypothèse de convexité (voir le corollaire 5.8), on trouvera toujours des extrema locaux.

5.2.1 Cas sans contrainte

On parle d'optimisation sans contrainte si Ω, appelé ensemble des éléments admissibles dans le problème d'optimisation, est un ouvert de E. Remarquons que ce qui suit est aussi valable si Ω n'est pas ouvert mais que x_0 est supposé à l'intérieur de Ω.

Proposition 5.14 (Conditions nécessaires d'optimalité d'ordre 1 et 2)**.** *Soit Ω un ouvert de E, $f : \Omega \to \mathbb{R}$ et $x_0 \in \Omega$.*

1. *On suppose f différentiable sur Ω* [77] *; alors*

$$x_0 \in \Omega \text{ est un minimum local de } f \;\Rightarrow\; df(x_0) = 0.$$

2. *On suppose f deux fois différentiable sur Ω ; alors*

$$x_0 \in \Omega \text{ est un minimum local de } f \;\Rightarrow\; df(x_0) = 0 \quad \text{et} \quad d^2 f(x_0) \geq 0.$$

Remarque 5.15. Étant donnée une fonction $f : \Omega \to \mathbb{R}$, les solutions $x_0 \in \Omega$ de l'équation

$$df(x_0) = 0$$

sont appelés points critiques de f.

En pratique, pour trouver un minimum de f, on peut donc chercher ses points critiques, ce qui permet en quelque sorte de "réduire le champ de recherche", i.e. identifier les possibles candidats à la minimalité :

1. si par exemple il y a exactement un point critique, alors ce point est le seul candidat pour être minimum de f ; si par ailleurs on sait qu'il existe un minimum de f, alors c'est bien le point trouvé [78] ;
2. si f a plusieurs points critiques (en nombre fini ou infini) et qu'on sait l'existence d'un minimum de f, on peut décider de qui est le minimum global en comparant les valeurs de f en ces points, si cela est facile ;
3. remarquons que la condition d'optimalité d'ordre 1 ne distingue pas la recherche de minima et de maxima ;
4. on peut utiliser la condition d'ordre 2 pour réduire encore le nombre de candidats possibles à la minimalité ; contrairement à la condition d'ordre 1, la condition est bien sûr différente entre les minima et les maxima.

Démonstration. Supposons que x_0 est un minimum local de f.

Suppons également que f est différentiable. Alors il existe $\delta > 0$ tel que $\forall x \in B(x_0, \delta),\; f(x) \geq f(x_0)$, où on a éventuellement réduit δ de sorte à avoir $B(x_0, \delta) \subset \Omega$. Soit $h \in E \setminus \{0\}$: alors $x_0 + th \in B(x_0, \delta)$ si $|t| < \frac{\delta}{\|h\|}$, d'où en divisant par $t > 0$ petit,

$$df(x_0)(h) + o(1) \geq 0$$

77. En fait, on n'a besoin que de la différentiabilité de f en x_0 ; mais précisément, en général on applique ce résultat pour trouver x_0 qui est l'inconnue du problème.

78. Attention à ne pas négliger la question d'existence ! C'est une erreur classique de considérer que si f n'a qu'un point critique, alors ce dernier est le minimum de f ; en effet f peut très bien ne pas avoir de minimum. Néanmoins, si le problème est convexe (i.e. Ω est convexe et f est convexe, alors le fait de trouver un point critique montre qu'il est minimum (voir le corollaire 5.8, et donc cela prouve l'existence d'un minimum !

ce qui donne $df(x_0)(h) \geq 0$ en faisant tendre t vers 0. Si maintenant on divise par $t < 0$ petit, on obtient

$$df(x_0)(h) + o(1) \leq 0$$

d'où $df(x_0)(h) \leq 0$ en faisant tendre t vers 0. Au final, $df(x_0)(h) = 0$.

Supposons maintenant que f est deux fois différentiable. Par le point précédent, $df(x_0) = 0$. Donc en reprenant le calcul précédent et en appliquant la formule de Taylor-Young (proposition 2.12), on obtient

$$f(x_0 + th) = f(x_0) + t\underbrace{df(x_0)(h)}_{=0} + \frac{t^2}{2}d^2f(x_0)(h,h) + o(t^2) \geq f(x_0)$$

d'où en divisant par t^2 et en faisant tendre t vers 0, $d^2f(x_0)(h,h) \geq 0$. □

Proposition 5.16 (Conditions suffisante d'optimalité)**.** *Soit Ω un ouvert de E, $f : \Omega \to \mathbb{R}$ deux fois différentiable sur Ω, et $x_0 \in \Omega$. S'il existe $\alpha > 0$ tel que*[79]

$$df(x_0) = 0 \qquad \text{et} \qquad \forall h \in E,\ d^2f(x_0)(h,h) \geq \alpha\|h\|^2,$$

alors x_0 est un minimum local de f[80].

Remarque 5.17. Attention, la condition

$$\forall h \in E \setminus \{0\}, d^2f(x_0)(h,h) > 0. \tag{5.2}$$

ne suffit pas a priori. Néanmoins, si E est de dimension finie, cela suffit, car la condition (5.2) implique qu'il existe $\alpha > 0$ telle que

$$\forall h \in E \setminus \{0\}, d^2f(x_0)(h,h) \geq \alpha\|h\|^2.$$

En effet, du fait de la compacité de l'ensemble $\mathbb{S}_E = \{h \in E, \|h\| = 1\}$ si E est de dimension finie, le problème

$$\min\left\{d^2f(x_0)(h,h),\ h \in \mathbb{S}_E\right\}$$

admet une solution $h_0 \in \mathbb{S}$, donc la valeur α du minimum est bien strictement positive. Et alors pour tout $h \in E \setminus \{0\}$, on a

$$d^2f(x_0)(h,h) = d^2f(x_0)\left(\frac{h}{\|h\|}, \frac{h}{\|h\|}\right)\|h\|^2 \geq \alpha\|h\|^2.$$

Démonstration. D'après la formule de Taylor, on a pour tout h suffisamment petit,

$$f(x_0 + h) = f(x_0) + df(x_0)(h) + \frac{1}{2}d^2f(x_0)(h,h) + \|h\|^2\varepsilon(h),$$

où ε est définie sur un voisinage de 0 dans E, et a pour limite 0 quand h tend vers 0.

Par hypothèse $df(x_0) = 0$, et il existe $\delta > 0$ tel que

$$\forall h \in B(0,\delta),\ |\varepsilon(h)| \leq \frac{\alpha}{4}$$

d'où

$$f(x_0 + h) \geq f(x_0) + \frac{\alpha}{2}\|h\|^2 - \frac{\alpha}{4}\|h\|^2 \geq f(x_0),$$

d'où le résultat. □

Remarque 5.18. Les conditions nécessaires et les conditions suffisantes données précédemment diffèrent. C'est normal ; si la différentielle seconde a des directions d'annulation[81], alors il faudrait aller voir le comportement des dérivées 3ième ou supérieure pour analyser la situation. Par exemple, la fonction $x \in \mathbb{R} \mapsto x^4$ admet un minimum en 0, alors que la dérivée seconde de f en 0 s'annule.

79. La seconde condition s'énonce parfois par "$d^2f(x_0)$ est coercive". On laisse le lecteur se convaincre que la condition est bien la même que la coercivité d'une fonction au sens où elle tend vers $+\infty$ en $+\infty$.

80. Notez qu'on peut en fait montrer (la preuve est la même) que x_0 est un minimum local strict, c'est-à-dire qu'il existe $\delta > 0$ tel que $\forall x \in B(x_0,\delta) \setminus \{x\},\ f(x) > f(x_0)$.

81. On parle en dimension finie ; le cas de la dimension infinie, où on peut avoir $df(x_0) = 0$ et $d^2f(x_0)(h,h) > 0$ pour tout h non nul, mais x_0 n'est pas un minimum local, est encore plus complexe.

5.2.2 Cas avec contraintes d'égalité

On se place en dimension finie dans ce paragraphe [82] ; on considère $(n,p) \in \mathbb{N}^*$, Ω un ouvert de $\mathbb{R}^n$, $f:\Omega \to \mathbb{R}$, et $g:\Omega \to \mathbb{R}^p$; la fonction f est l'énergie qu'on souhaite minimiser, et g représentera les contraintes [83] : on s'intéresse au problème d'optimisation

$$\inf \Big\{ f(x),\ x \in \Omega,\ g(x) = 0 \Big\}, \tag{5.3}$$

qu'on appelle optimisation sous contraintes d'égalité.

Remarque 5.19. L'égalité $g(x) = 0$ peut se réécrire

$$\forall i \in [\![1,p]\!],\quad g_i(x) = 0,$$

d'où l'idée de p contraintes.

Bien sûr, étant donnés $c \in \mathbb{R}^p$, on pourrait aussi considérer la contrainte $g(x) = c$, simplement en remplaçant g par $g-c$. D'ailleurs, comme les énoncés qui suivent ne font qu'intervenir les différentielles de g, ça ne change rien de travailler avec g ou $g-c$.

Sauf si g est mal choisi, l'ensemble

$$\{x \in \Omega,\ g(x) = 0\}$$

n'est pas un ouvert, on ne peut donc pas appliquer les résultats du paragraphe précédent. Pour cette raison, on a le résultat suivant qui donne une condition nécessaire d'optimalité d'ordre 1 pour le problème (5.3).

Théorème 5.20 (Condition d'Euler-Lagrange à l'ordre 1)**.** *Soit $x_0 \in \Omega$ solution locale de* (5.3)*, c'est-à-dire qu'il existe $\delta > 0$ tel que*

$$g(x_0) = 0, \quad \text{et} \quad \Big[\,\forall x \in B(x_0,\delta) \text{ tels que } g(x) = 0, \quad f(x) \geq f(x_0)\,\Big]. \tag{5.4}$$

On suppose f différentiable, et g de classe C^1. Enfin on suppose

$$dg(x_0) \text{ est surjective.}$$

Alors il existe $\lambda \in \mathbb{R}^p$ tel que

$$df(x_0) = \lambda \cdot dg(x_0), \tag{5.5}$$

où $\lambda \cdot dg(x_0) = \sum_{i=1}^p \lambda_i dg_i(x_0)$.

Remarque 5.21. Ainsi, naturellement (l'optimalité est plus faible car n'est valable que parmi les x tels que $g(x) = 0$), la condition obtenue est plus faible que la condition d'optimalité sans contrainte ; elle signifie

$$\forall h \in \mathbb{R}^n,\ df(x_0)(h) = \sum_{i=1}^p \lambda_i dg_i(x_0)(h).$$

On peut encore l'écrire

$$\nabla f(x_0) = \sum_{i=1}^p \lambda_i \nabla g_i(x_0).$$

Les $(\lambda_i)_{i\in[\![1,p]\!]}$ sont appelés multiplicateurs de Lagrange. En pratique, il faut les voir comme des inconnues : plus précisément, pour chercher les candidats à l'optimalité, on cherche à résoudre le système

$$\left\{ \begin{array}{rcl} g(x_0) & = & 0 \\ \nabla f(x_0) & = & \displaystyle\sum_{i=1}^p \lambda_i \nabla g_i(x_0) \end{array} \right. \tag{5.6}$$

qui a $p+n$ équations et $p+n$ inconnues, à savoir les couples (x_0,λ) (même si au final c'est x_0 qui nous intéresse) ; on peut d'ailleurs raisonnablement appeler point critique un élément $x_0 \in \Omega$ tel qu'il existe $\lambda \in \mathbb{R}^p$ pour lequel les équations (5.6) sont satisfaites.

Ainsi le minimiseur, s'il existe, fait partie de l'ensemble des points critiques.

82. C'est une simplification ; il existe des versions de ces résultats dans des espaces de Banach.
83. C'est une question de préférence que de dire la ou les contraintes ; en général, on réfère à une condition réelle comme étant une condition. Avec cette terminologie, il y a p contraintes ici.

Remarque 5.22. La condition “$dg(x_0)$ est surjective” peut se réécrire[84]

$$(\nabla g_1(x_0), \ldots, \nabla g_p(x)) \text{ est une famille libre.}$$

Si $p = 1$, la condition devient simplement $\nabla g_1(x_0) \neq 0$.

Encore une fois, comme x_0 est l’inconnue, on est parfois amené à vérifier cette hypothèse en tous les points x_0 admissibles.

Il s’agit de la même condition qu’au théorème 4.6 pour caractériser les sous-variétés de $\mathbb{R}^n$ par une définition implicite.

Démonstration. **Étape analytique :** Nous allons montrer que[85]

$$\forall h \in \mathbb{R}^n \text{ tel que } dg(x_0)(h) = 0, \quad df(x_0)(h) = 0. \tag{5.7}$$

— **Première preuve, valable seulement si $p = 1$:**[86] On commence par écrire un développement limité de g au voisinage de x_0 à l’ordre 1 : pour tout $h \in \mathbb{R}^n$ assez proche de 0, on a

$$g(x_0 + h) = g(x_0) + dg(x_0)(h) + \|h\|\varepsilon(h)$$

où ε est définie sur un voisinage de 0, et tend vers 0 quand h tend vers 0.

Soit h tel que $dg(x_0)(h) = 0$. Bien sûr si $h = 0$, alors $df(x_0)(h) = 0$. On peut donc supposer que $h \neq 0$. Comme, par ailleurs, $dg(x_0) \neq 0$, il existe $e \in \mathbb{R}^n$ tel que $dg(x_0)(e) = 1$.

Par hypothèse sur ε, il existe $\alpha > 0$ (qu’on peut supposer inférieur à δ[87] sans perte de généralité) tel que

$$\forall w \in B(0, \alpha), \ |\varepsilon(w)| \leq \frac{1}{2(\|h\| + \|e\|)}.$$

Nous allons montrer que pour tout $0 < t < t_0 := \frac{\alpha}{\|h\|+\|e\|}$, on peut trouver $s(t) \in [-t, t]$ tel que

$$g(x_0 + th + s(t)e) = 0, \qquad \text{et} \qquad \frac{s(t)}{t} \xrightarrow[t \to 0]{} 0. \tag{5.8}$$

En effet, soit un tel $t \in]0, t_0[$; remarquons que pour tout $s \in [-t, t]$, on a

$$g(x_0 + th + se) = \underbrace{g(x_0) + tdg(x_0)(h)}_{=0} + sdg(x_0)(e) + \|th + se\|\varepsilon(th + se) = s + \|th + se\|\varepsilon(th + se).$$

On a pour tout $s \in [-t, t]$, $th + se \in B(0, \alpha)$, donc

$$\|th + se\||\varepsilon(th + se)| \leq |t|(\|h\| + \|e\|)|\varepsilon(th + se)| \leq \frac{t}{2},$$

et donc en prenant $s = t$ et $s = -t$ on a

$$g(x_0 + th + te) \geq t/2 > 0 \quad \text{ et } \quad g(x_0 + th - te) \leq -t/2 < 0.$$

Par le théorème des valeurs intermédiaires, il existe donc $s(t) \in [-t, t]$ tel que $g(x_0 + th + s(t)e) = 0$. De plus, $s(t) + \|th + s(t)e\|\varepsilon(th + s(t)e) = 0$, et comme $s(t) \xrightarrow[t \to 0]{} 0$ (du fait que $s \in [-t, t]$), par composition des limites on obtient

$$\frac{|s(t)|}{t} = \frac{1}{t}\|th + s(t)e\||\varepsilon(th + s(t)e)| \leq (\|h\| + \|e\|)|\varepsilon(th + s(t)e)| \xrightarrow[t \to 0]{} 0.$$

Nous allons en conclure que $df(x_0)(h) = 0$. En effet, comme x_0 est un minimum local de f sur $\{x \in \Omega, g(x) = 0\}$ et que par (5.8) $g(x_0 + th + s(t)e) = 0$, on peut écrire que pour tout $0 < t < t_0$,

$$f(x_0 + th + s(t)e) - f(x_0) \geq 0.$$

84. Vérifiez que c’est bien la même chose !

85. Remarquez qu’il s’agit bien d’une condition impliquée par (5.5).

86. On pourrait qualifier cette preuve d’être “à la main” ; je l’aime bien car il me semble qu’on comprend bien ce qu’il se passe sur un dessin ; je n’ai pas de référence, si quelqu’un en connaît une, ça serait super qu’il me l’envoie.

87. Venant de (5.4).

En effectuant un développement limité de f à l'ordre 1, il vient que

$$tdf(x_0)(h) + s(t)df(x_0)(e) + o(t) \geq 0.$$

En divisant par $t > 0$, puis par passage à la limite quand $t \to 0^+$, on obtient $df(x_0)(h) \geq 0$. En changeant h en $-h$, on obtient $df(x_0)(h) \leq 0$, d'où finalement $df(x_0)(h) = 0$.

— **Deuxième preuve, valable dans le cas général :** On pose $E_1 = Ker(dg(x_0))$ [88] ; il existe E_2 un supplémentaire de E_1 dans $\mathbb{R}^n$, c'est-à-dire que $E_1 \oplus E_2 = \mathbb{R}^n$. On note $x = (x^1, x^2)$ les coordonnées dans ce découpage de $\mathbb{R}^n$. Par construction, $\partial_1 g(x_0) = 0$. De plus $\partial_2 g(x_0)$ est une bijection de E_2 sur $Im(dg(x_0))$:
  — en effet, si $y \in Im(dg(x_0))$ alors il existe $h = (h^1, h^2) \in \mathbb{R}^n$ tel que $y = dg(x_0)(h^1, h^2) = \partial_2 g(x_0)(h^2)$ d'où la surjectivité de $\partial_2 g(x_0)$ de E_2 dans $Im(dg(x_0))$.
  — et si $\partial_2 g(x_0)(h) = 0$ avec $h \in E_2$, alors $h \in E_1 \cap E_2 = \{0\}$, donc $\partial_2 g(x_0)$ est injective.

Par hypothèse, $Im(\partial_2 g(x_0)) = \mathbb{R}^p$, ainsi $\partial_2 g(x_0)$ est une bijection de E_2 dans $\mathbb{R}^p$. Par le théorème des fonctions implicites, il existe $\mathcal{U}$ un voisinage de x_0^1 et $\varphi : \mathcal{U} \to E_2$ de classe C^1 tels que

$$\varphi(x_0^1) = x_0^2, \quad \text{et} \quad \forall x^1 \in \mathcal{U}, \;\; g(x^1, \varphi(x^1)) = 0.$$

En dérivant cette dernière relation et en prenant la valeur en x_0^1, on obtient

$$\partial_1 g(x_0) + \partial_2 g(x_0) \circ d\varphi(x_0^1) = 0$$

mais comme $\partial_1 g(x_0) = 0$ et $\partial_2 g(x_0)$ est bijective, finalement $d\varphi(x_0^1) = 0$.
Par optimalité locale de x_0 sur l'ensemble $\{x \in \Omega, g(x) = 0\}$, on a donc qu'il existe $\alpha > 0$ tel que

$$\forall x^1 \in B(x_0^1, \alpha), \;\; f(x^1, \varphi(x^1)) \geq f(x_0^1, \varphi(x_0^1)) = f(x_0).$$

On est donc ramené à une optimisation sans contrainte, donc par la proposition 5.14, la différentielle de $x^1 \mapsto f(x^1, \varphi(x^1))$ est nulle en x_0^1, c'est-à-dire

$$\partial_1 f(x_0) + \partial_2 f(x_0) \circ d\varphi(x_0^1) = 0$$

et comme $d\varphi(x_0^1) = 0$, on a finalement $\partial_1 f(x_0) = 0$. Autrement dit, on a

$$\forall h \in E_1, \;\; df(x_0)(h) = \partial_1 f(x_0)(h) = 0.$$

Étape algébrique : Pour conclure, on utilise le lemme algébrique ci-après pour $L = df(x_0)$ et $L_i = dg_i(x_0), i \in [\![1, p]\!]$. □

Lemme 5.23. *Soient E un espace vectoriel normé, $\{L_1, \ldots, L_p\}$ une famille libre dans E' et $L \in E'$. On note $\Gamma = \bigcap_{i=1}^p \mathrm{Ker} L_i$. Alors $L|_\Gamma = 0$ si et seulement s'il existe $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ tels que*

$$L = \sum_{i=1}^{p} \lambda_i L_i. \tag{5.9}$$

Démonstration. [89] Si $L \in \mathrm{vect}\{L_i, i = 1, \ldots, p\}$, alors L est combinaison linéaire des L_i donc tout vecteur qui annule chacun des L_i annule aussi L, d'où l'inclusion des noyaux.

Réciproquement, on suppose que $\bigcap_{i=1}^p \ker L_i \subset \ker L$ et l'on veut en déduire que $L \in \mathrm{vect}\{L_i, i = 1, \ldots, p\}$. On raisonne par récurrence sur p.

Pour $p = 1$, l'hypothèse est que $\ker L_1 \subset \ker L$. La famille (L_1) est libre, ce qui veut dire ici que $L_1 \neq 0$. Il existe donc $e_1 \in E$, $L_1(e_1) \neq 0$, c'est-à-dire $e_1 \notin \ker L_1$. L'espace se décompose en somme directe $E = \mathbb{R}e_1 \oplus \ker L_1$ c'est-à-dire que pour tout $u \in E$, il existe $\alpha \in \mathbb{R}$ et $v \in \ker L_1$ tels que $u = \alpha e_1 + v$. Par conséquent, $L(u) = \alpha L(e_1) + L(v) = \alpha L(e_1)$ vu que $L(v) = 0$ par hypothèse. De même, $L_1(u) = \alpha L_1(e_1)$, donc $L(u) = \frac{L(e_1)}{L_1(e_1)} L_1(u)$, soit $L = \lambda_1 L_1$ avec $\lambda_1 = \frac{L(e_1)}{L_1(e_1)}$.

88. Il s'agit du plan tangent à la variété $\{x \in \Omega, g(x) = 0\}$ en x_0.
89. On peut faire une preuve plus expéditive via la dualité, voir [Ave83].

Supposons le résultat vrai pour $p-1$ formes linéaires. Sous forme contraposée, on a donc que $M \notin \text{vect}\{M_k\}$ implique que $\cap_k \ker M_k \not\subset \ker M$, pour tout famille libre $\{M_k\}$ de cardinal $p-1$. Pour chaque $j = 1, \dots, p$, la famille $\{L_i, i \neq j\}$ est libre de cardinal $p-1$, et comme la famille $\{L_i\}$ est libre, $L_j \notin \text{vect}\{L_i, i \neq j\}$. L'hypothèse de récurrence s'applique, et il existe donc $e_j \in \cap_{i\neq j} \ker L_i$ tel que $e_j \notin \ker L_j$. En d'autres termes, $L_i(e_j) = 0$ pour tout $i \neq j$ et $L_j(e_j) \neq 0$. Pour tout $u \in E$, on pose alors

$$v = u - \sum_{j=1}^{p} \frac{L_j(u)}{L_j(e_j)} e_j.$$

Il vient donc

$$L_i(v) = L_i(u) - \sum_{j=1}^{p} \frac{L_j(u)}{L_j(e_j)} L_i(e_j) = L_i(u) - L_i(u) = 0,$$

pour tout i, si bien que $v \in \bigcap_{i=1}^{p} \ker L_i$. Par hypothèse, il s'ensuit que $L(v) = 0$, ce qui signifie que

$$L(u) = \sum_{j=1}^{p} \frac{L_j(u)}{L_j(e_j)} L(e_j) = \sum_{j=1}^{p} \frac{L(e_j)}{L_j(e_j)} L_j(u),$$

d'où

$$L = \sum_{j=1}^{p} \lambda_j L_j \text{ avec } \lambda_j = \frac{L(e_j)}{L_j(e_j)}.$$

□

Remarque 5.24. Dans la preuve précédente, on ignore un peu la "théorie" des sous-variétés de $\mathbb{R}^n$. En effet en considérant $\Gamma = \{x \in \Omega, g(x) = 0\}$ qui est une sous-variété au voisinage de x_0, et si on admet par exemple la proposition 4.14, voici une preuve bien plus rapide [90] de l'étape analytique : si h est tel que $dg(x_0)(h) = 0$, alors il existe $\gamma :]-\delta, \delta[\to \mathbb{R}^n$ de classe C^1 telle que

$$\gamma(]-\delta, \delta[) \subset \Gamma, \quad \gamma(0) = x_0, \quad \gamma'(0) = h.$$

Et donc on peut écrire

$$\forall t \in]-\delta, \delta[, \ f(\gamma(t)) \geq f(\gamma(0)).$$

La dérivée de $t \mapsto f(\gamma(t))$ est donc nulle en 0, d'où

$$0 = (f \circ \gamma)'(0) = df(\gamma(0))(\gamma'(0)) = df(x_0)(h).$$

90. Mais il n'y a rien de magique bien sûr, il faut démontrer cette proposition si on veut être complet ; voir par exemple [Ave83, Théorème 2.3 page 98] avec une preuve assez simple de l'égalité $T_{x_0}\Gamma = Ker(dg(x_0))$.

Chapitre VIII

Analyse de Fourier et Distributions

Ce chapitre me paraît particulièrement essentiel dans le programme de l'agrégation :

— ça "tombe" très fréquemment à l'écrit : entre 2000 et 2019, plus d'un sujet sur deux faisait intervenir l'analyse de Fourier, et dans plus d'un sur quatre, c'était un élément central du sujet, qui intervenait dès le début de celui-ci ;

— et pour cause, l'analyse de Fourier est très transverse : elle utilise (quand on veut bien la faire) la théorie de la mesure de Lebesgue, et fait souvent recours à l'analyse fonctionnelle ;

— il y a bien sûr les leçons **246-250** (éventuellement **209**) qui font directement référence à l'analyse de Fourier, mais d'une part il y a des leçons dans lesquelles on ne peut pas "couper" à Fourier (par exemple **213-222-239**), et d'autre part, l'analyse de Fourier sera souvent le meilleur moyen de donner des applications à vos résultats, que ceux-ci soient issus de la théorie de Lebesgue, ou de l'analyse fonctionnelle.

La théorie des distributions est apparue il y a environ 10 ans dans le programme de l'agrégation [1], ce qui fait toujours débat ; les rapports des dernières années s'efforcent à préciser les attentes du jury [2]. Ce n'est pas le lieu de discuter si la présence de cette théorie est légitime ou non ; mais on invite quand même le candidat à lire les rapports 2018-2017 pour mesurer ce qui est attendu par le jury sur ces notions. On l'a déjà mentionné dans le reste du cours : plus le niveau des notions considérées est élevé, moins le jury est exigeant dans la maîtrise des fondements de la théorie [3] ; pour la théorie des distributions, ce principe s'applique complément, et le jury répète à maintes reprises dans les rapports qu'il attend surtout une manipulation de ces outils sur des exemples, et qu'aucune subtilité théorique n'est attendue [4].

Comme toujours, mon conseil reste le suivant : si vous souhaitez faire une impasse sur le paragraphe "distributions" [5], ne faites pas une impasse à 100%, prenez quand même la peine d'en savoir un minimum. Dans les écrits 2018 et 2019, le jury mentionne la notion de distribution [6] ; même s'il ne fallait pas en savoir grand chose, il était probablement plus agréable d'aborder les questions concernées en ayant une idée, même un peu vague, de ce qu'est une distribution !

Pour les références, la plupart des livres d'intégration évoquent l'analyse de Fourier, voir par exemple [Wag12a, GH13]. Il y a également d'excellentes références en anglais, [Kat04] qui est un grand classique, concis mais de niveau élevé, [Zyg03] qui en dira beaucoup plus qu'ici, et [DM72].

1. Elle figurait dans pas moins de 3 titres de leçons au départ, pour ensuite finir par ne plus être mentionnée explicitement dans les leçons ; elle fait un retour pour le moins criticable pour la session 2020 en s'immiscant vicieusement dans le titre de **228**...

2. D'ailleurs, il me semble qu'entre les programmes 2019 et 2020, dans la section Analyse et Probabilités, seul le chapitre distribution voit une évolution ; évolution pas tellement négligeable puisque ces dernières années, la notion de distribution "tout court" avait disparu pour laisser place à la notion de distribution tempérée ; mais pour la session 2020, la notion de distribution refait son apparition au programme.

3. À titre d'exemple, rappelons que le jury n'attend pas grand chose sur la construction des mesures comme la mesure de Lebesgue, ou la notion de mesure produit. De même dans les applications en probabilités, les subtilités de théorie de la mesure sont souvent laissées en arrière plan

4. Notamment dans la définition même d'une distribution !

5. Ce qui se défend tout à fait.

6. Très tôt dans le sujet, dans le cas de 2019.

1 Séries de Fourier

On va s'intéresser ici aux fonctions définies de $\mathbb{R}$ et à valeurs dans $\mathbb{C}$[7], et qui sont 2π-périodiques[8]. On notera $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$, sans se soucier qu'il s'agit d'un espace quotient[9] :

— étant donné $k \in \mathbb{N} \cup \{\infty\}$, on notera $C^k(\mathbb{T})$ pour

$$C^k(\mathbb{T}) = \Big\{ f : \mathbb{R} \to \mathbb{C},\ f \text{ est de classe } C^k \text{ et } 2\pi\text{-périodique} \Big\}.$$

On peut munir ces espaces de leur norme naturelle : par exemple la norme $\|\cdot\|_\infty$ si $k = 0$. Comme l'ensemble des fonctions 2π-périodique est inclus et fermé dans $(C_b^0(\mathbb{R}), \|\cdot\|_\infty)$, $(C^0(\mathbb{T}), \|\cdot\|_\infty)$ est un espace de Banach.

— étant donné $p \in [1, +\infty[$, on notera $L^p(\mathbb{T})$ pour[10]

$$L^p(\mathbb{T}) = \left\{ f : \mathbb{R} \to \mathbb{C} \text{ mesurable, } 2\pi\text{-périodique, et telle que } \int_{-\pi}^{\pi} |f(t)|^p dt < +\infty \right\},$$

que l'on munit de la norme

$$\|f\|_{L^p(\mathbb{T})} := \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t)|^p dt \right)^{1/p}. \tag{1.1}$$

Attention à la présence du $\frac{1}{2\pi}$, qui sera commode notamment pour l'énoncé du théorème de Parseval. Bien sûr, ça n'est pas une obligation, mais soyez vigilant. On distinguera la notation $\|\cdot\|_p$ (sans le $\frac{1}{2\pi}$) et $\|\cdot\|_{L^p(\mathbb{T})}$ (avec).

Exercice 1.1. Soit $f : \mathbb{R} \to \mathbb{C}$ 2π-périodique, et $a \in \mathbb{R}$. Montrer que

$$\int_{-\pi}^{\pi} f(t)dt = \int_a^{a+2\pi} f(t)dt,$$

autrement dit que l'intégrale est la même sur tout intervalle de longueur 2π. On utilisera la notation

$$\int_{\mathbb{T}} f(t)dt.$$

Remarque 1.2 (Périodisation d'une fonction)**.** Étant donné une fonction $f :]-\pi, \pi] \to \mathbb{C}$[11], on peut considérer sa périodisée $\widetilde{f} : \mathbb{R} \to \mathbb{C}$, construite ainsi : pour tout $k \in \mathbb{Z}$ et tout $t \in]-\pi + 2k\pi, \pi + 2k\pi]$, on pose $\widetilde{f}(t) = f(t - 2k\pi)$.

---

7. On ne considèrera donc les séries de Fourier que comme somme d'exponentielles complexes ; il n'est pas sans intérêt de vouloir, notamment si la fonction est à valeurs réelles, ou ayant une parité particulière, développer celle-ci en somme de cos et/ou de sin. Dans ce cas, soit vous apprenez toutes les formules adaptées à ce cadre par cœur (ce que je déconseille de faire si on a une mémoire faillible), soit vous apprenez à passer d'une somme d'exponentielle à une somme de cos / sin en un temps très raisonnable.

8. Il y a essentiellement, dans le choix de la période, 3 conventions qui existent dans la littérature : 2π (que nous choisissons ici et qui, je pense, est le choix plus fréquent), 1 (qui a ses avantages et ses inconvénients en comparaison) et T où $T > 0$ est un réel quelconque, qui a clairement l'avantage d'être plus général, mais donne des formules un peu plus coûteuses à retenir. Tous les choix se défendent ; mais il faut savoir se ramener à un cas ou à un autre en un temps raisonnable. N'oubliez pas que si f est T périodique et $\lambda > 0$, alors $x \mapsto f(\lambda x)$ est T/λ-périodique.

9. L'espace $\mathbb{R}/2\pi\mathbb{Z}$ pourrait être considéré comme un vrai ensemble quotient (on quotiente le groupe additif (et commutatif) $\mathbb{R}$ par un de ses sous-groupes), mais cela pose le problème de la topologie, voire de la structure différentielle ou de la mesure qu'on met dessus, puisque nous allons parler de $C^0(\mathbb{T})$, $C^k(\mathbb{T})$, $L^p(\mathbb{T})$; on peut définir ces choses rigoureusement, mais cela sort un peu de cadre de l'agrégation, et serait surtout inutilement compliqué alors qu'on peut donner des définitions "à la main" comme on le fait ici sans souci. On prend donc l'emploi de $\mathbb{R}/2\pi\mathbb{Z}$ ici simplement comme une notation ; si vous maîtrisez le sens des notations que vous employez, notamment via une description claire des espaces fonctionnels que vous manipulez, je ne pense pas que le jury vous tiendrait rigueur d'utiliser une notation "quotient" alors que vous n'utilisez pas réellement l'objet quotient. Si la notation vous inquiète, ne vous forcez pas à l'utiliser ! Il n'y a pas de mal à utiliser $C^k_{2\pi}(\mathbb{R})$ par exemple ; quant aux espaces L^p, comme on l'explique à la remarque 1.2, il est tout-à-fait satisfaisant d'écrire $L^p(-\pi, \pi)$.

10. Plus précisément, il s'agit des classes d'équivalence pour l'égalité presque partout des fonctions 2π-périodiques et dont la puissance $p$ième est intégrable sur $]-\pi, \pi[$.

11. On pourrait le faire aussi pour des fonctions définies sur n'importe quel intervalle de longueur 2π.

On peut faire l'abus de noter f à la place de $\widetilde{f}$: auquel cas les espaces $L^p(-\pi,\pi)$ et $L^p(\mathbb{T})$ peuvent être confondus sans problème, à condition de faire attention au fait qu'on a renormalisé la définition de la norme (voir (1.1)) par commodité.

Si par contre on essaie de faire de même pour une fonction $f \in C^k(]-\pi,\pi])$, alors la fonction périodisée $\widetilde{f}$[12] n'est pas nécessaire de classe C^k sur $\mathbb{R}$. Il faut que le comportement de f en π à gauche soit compatible avec le comportement de f en $-\pi$ à droite.
Par exemple, étant donnée $f \in C^0(]-\pi,\pi])$, la fonction $\widetilde{f}$ sera continue sur $\mathbb{R}$ si et seulement si $f(\pi^-) = f(-\pi^+)$.

Définition 1.3. — On appelle polynome trigonométrique une fonction de la forme $x \in \mathbb{R} \mapsto \sum_{k=-N}^{N} a_k e^{ikx}$ où $N \in \mathbb{N}$ et $(a_{-N},\dots,a_N) \in \mathbb{C}^{2N+1}$.

— On appelle série trigonométrique une série de fonctions dont les sommes partielles s'écrivent

$$\left(x \in \mathbb{R} \mapsto \sum_{n=-N}^{N} a_n e^{inx}\right)_{N\in\mathbb{N}}$$

où $(a_n)_{n\in\mathbb{Z}} \in \mathbb{C}^{\mathbb{Z}}$.[13]

La question que nous allons aborder ici, est d'essayer de voir une fonction comme une limite d'une série trigonométrique bien particulière, qu'on appellera série de Fourier de f. Remarquons que cette méthode a été introduite et utilisée par Fourier pour étudier l'équation de la chaleur au début de 19e siècle. La justification mathématique des résultats de convergence des séries de Fourier motivera l'introduction de nombreux concepts d'analyse au 19e siècle (notamment avec l'intégrale de Riemann) et également au 20e siècle (l'intégrale de Lebesgue, l'introduction de l'analyse fonctionnelle...), au moins jusqu'au célèbre théorème de Carleson (1966).

1.1 Cadre $L^1(\mathbb{T})$

Remarque 1.4. Considérons le polynome trigonométrique $P : x \in \mathbb{R} \mapsto \sum_{k=-N}^{N} a_k e^{ikx}$. Une question naturelle est de savoir comment retrouver les coefficients a_k à partir de P[14] ; pour cela, on constate :

$$\forall k \in [\![-N,N]\!],\ \ a_k = \frac{1}{2\pi}\int_{\mathbb{T}} P(t)e^{-ikt}dt$$

puisque $\int_{\mathbb{T}} e^{int}dt = 2\pi\delta_{n,0}$ pour tout $n \in \mathbb{Z}$.

Définition 1.5. Soit $f \in L^1(\mathbb{T})$.

— on définit la suite des coefficients de Fourier de f par[15]

$$\forall n \in \mathbb{Z},\ \widehat{f}(n) = \frac{1}{2\pi}\int_{\mathbb{T}} f(t)e^{-int}dt$$

12. Il peut toujours s'autoriser l'abus d'écrire f au lieu de $\widetilde{f}$, mais il faut redoubler de vigilance, car comme on l'explique ici, la fonction $\widetilde{f}$ ne conserve pas toutes les propriétés de f.

13. Attention, ici on ne présage a priori de rien sur la convergence de la série trigonométrique.

14. Un peu comme le fait la formule de Taylor pour les polynomes, à savoir que si $P = \sum_{k=0}^{N} a_k X^k \in \mathbb{R}[X]$, alors $a_k = \frac{P^{(k)}(0)}{k!}$.

15. La notation $c_n(f)$ à la place de $\widehat{f}(n)$ est probablement plus répandue ; bien sûr, j'invite le lecteur à choisir la notation qu'il préfère, notamment dans le cadre du concours (il faut évidemment respecter les notations du sujet d'écrit si celui-ci les précise ; quant à l'oral, il est souvent plus commode d'utiliser au maximum les notations de sa référence favorite pour éviter de s'emméler les pinceaux dans l'écriture de son plan), mais je me permets de motiver un peu mon choix de notation :

— il est intéressant de voir l'opération $f \mapsto \widehat{f}$ comme un “changement de variable” ; ceci est plus facile à lire avec la notation “chapeau”. Cela donne tout de suite un point de vue “analyse fonctionnelle” ; notamment, on sera amené à se demander dans quel espace vit $\widehat{f}$, suivant les informations qu'on a sur f.

— il est vrai que la notation la plus répandue pour les suites, est d'écrire n en indice ; mais au fond ça n'est pas très logique. La fonction f aura pour variable $x \in \mathbb{T}$, et il est intéressant de voir $\widehat{f}$ également comme une fonction, mais dont la variable est $n \in \mathbb{Z}$. D'ailleurs, on parle parfois de “variable de Fourier” en désignant $n \in \mathbb{Z}$, voir aussi le point suivant.

— la raison la plus convaincante, mais aussi la plus dangereuse, est néanmoins le rapprochement que l'on peut faire avec la notion de transformation de Fourier, qu'on notera également $f \mapsto \widehat{f}$. Le danger est bien sûr de confondre

— On appelle série de Fourier associée à f la série trigonométrique

$$\left(S_N(f) := \left(x \in \mathbb{R} \mapsto \sum_{n=-N}^{N} \widehat{f}(n)e^{inx}\right)\right)_{N\in\mathbb{N}}.$$

Remarque 1.6. La question centrale de la théorie exposée dans ce paragraphe est donc :

la fonction f est-elle limite de sa série de Fourier $(S_N(f))_{N\in\mathbb{N}}$?

Mais n'oublions pas que cette question est imprécise, car une série de Fourier est une série de fonctions, donc il existe de multiples façons de converger.

Proposition 1.7. *L'application*

$$\begin{array}{rccc} \widehat{\cdot} : & L^1(\mathbb{T}) & \longrightarrow & \ell^\infty(\mathbb{Z}) \\ & f & \longmapsto & \widehat{f} \end{array}$$

est bien définie, linéaire continue, et

$$\|\widehat{f}\|_\infty \leq \|f\|_{L^1(\mathbb{T})}. \tag{1.2}$$

Démonstration. Pour montrer que $\widehat{\cdot}$ est bien définie, il faut montrer que pour tout $f \in L^1(\mathbb{T})$, la suite $\widehat{f}$ est bien définie et appartient à $\ell^\infty(\mathbb{Z})$ [16] : soit donc $f \in L^1(\mathbb{T})$. Comme pour tout $n \in \mathbb{Z}$ et tout $t \in \mathbb{R}, |f(t)e^{-int}| = |f(t)|$ et que $|f|$ est supposée intégrable sur $\mathbb{T}$, on en déduit que $\widehat{f}(n)$ a un sens pour tout $n \in \mathbb{Z}$. [17] De plus, on a

$$\forall n \in \mathbb{Z}, \quad |\widehat{f}(n)| = \left|\frac{1}{2\pi}\int_{\mathbb{T}} f(t)e^{-int}dt\right| \leq \frac{1}{2\pi}\int_{\mathbb{T}} \left|f(t)e^{-int}\right| dt = \frac{1}{2\pi}\int_{\mathbb{T}} |f(t)|\, dt = \|f\|_{L^1(\mathbb{T})}.$$

Ceci montre que $\widehat{f} \in \ell^\infty(\mathbb{Z})$, et donc que $\widehat{\cdot}$ est bien définie. La linéarité de $\widehat{\cdot}$ découle de la linéarité de l'intégrale de Lebesgue, et donc l'inégalité précédente prouve que $\widehat{\cdot}$ est continue, ainsi que l'inégalité (1.2) attendue. □

On peut en fait affiner l'espace d'arrivée proposé dans l'énoncé précédent. Rappelons qu'on note

$$c_0(\mathbb{Z}) = \left\{u : \mathbb{Z} \to \mathbb{C}, \quad \lim_{n\to\infty} u(n) = \lim_{n\to-\infty} u(n) = 0\right\}$$

et qu'il s'agit d'un sous-espace vectoriel fermé de $\ell^\infty(\mathbb{Z})$. Ainsi, $(c_0(\mathbb{Z}), \|\cdot\|_\infty)$ est un espace de Banach.

Lemme 1.8 (Lemme de Riemann-Lebesgue)**.** *Pour tout $f \in L^1(\mathbb{T})$, on a $\widehat{f} \in c_0(\mathbb{Z})$.*

Remarque 1.9. On a donc affiné l'espace d'arrivée de $\widehat{\cdot}$; on pourrait se demander si on pourrait encore affiner, ou dit autrement, si $f \in L^1(\mathbb{T}) \mapsto \widehat{f} \in c_0(\mathbb{Z})$ est surjective. La question n'est pas simple, et la réponse est non.

les deux notions ; si on comprend ce qu'on fait, il me semble que ce danger est limité puisque les cadres “classiques” des séries de Fourier et de la transformée de Fourier sont “d'intersection nulle” au sens où aucune fonction 2π-périodique n'est intégrable sur $\mathbb{R}$, si ce n'est bien sûr la fonction nulle. L'avantage est par contre de se rendre compte que plusieurs énoncés sur les séries de Fourier (à commencer par la définition des coefficients de Fourier) ont leur analogue dans l'étude de la transformée de Fourier ; à titre d'exemple, citons les théorèmes de Parseval et de Plancherel. Concluons en disant que cette similitude n'est pas anodine : il existe un cadre plus général dont les séries de Fourier et la transformée de Fourier seront deux cas particuliers, à savoir l'analyse harmonique sur les groupes abéliens localement compacts.

16. N'oubliez pas cette seconde partie qui dit que l'espace d'arrivée de l'opérateur est légitime ! C'est parfois la seule étape qui mérite justification.

17. Pour être plus rigoureux, il aurait fallu dire également que $\widehat{f}(n)$ (pour $n \in \mathbb{Z}$) ne dépend pas du représentant de f. En effet, f est mathématiquement une classe de fonctions, et l'objet $\widehat{f}(n)$ est la valeur commune de $\frac{1}{2\pi}\int_{\mathbb{T}} g(t)e^{-int}dt$ pour $g \in f$. C'est bien le cas ici vu qu'on manipule des intégrales. En pratique, on manipulera f comme si c'était une fonction, tout en restant vigilant sur le fait que les opérations qui font intervenir f sont inchangées si on modifie f sur un ensemble de mesure nulle. À titre d'exemple, on n'écrira jamais $f(0)$ pour $f \in L^1(\mathbb{T})$.

C'est d'ailleurs une difficulté non négligeable de notre changement de variable $f \mapsto \widehat{f}$, car étant doné $f \in L^1(\mathbb{T})$, il serait naturel de savoir où chercher exactement la nouvelle inconnue $\widehat{f}$. Plus généralement, si on considère un espace fonctionnel $X \subset L^1(\mathbb{T})$, il est naturel de se demander à quoi ressemble l'espace

$$\widehat{X} := \{\widehat{f},\ f \in X\}.$$

En pratique, il est rare de pouvoir identifier $\widehat{X}$ explicitement ; on trouve souvent des conditions nécessaires ou suffisantes. Dans certains cas exceptionnels et importants que nous détaillerons dans la suite, on peut effectivement identifier $\widehat{X}$.

Démonstration. [18] Soit $f \in C^1_c(]-\pi,\pi[)$ et $n \in \mathbb{Z}$. Alors $f \in L^1(-\pi,\pi)$ donc $\widehat{f}(n)$ est bien définie. De plus, par intégration par partie, on obtient

$$\int_{-\pi}^{\pi} f(t)e^{-int}dt = \left[f(t)\frac{e^{-int}}{-in}\right]_{-\pi}^{\pi} - \int_{-\pi}^{\pi} f'(t)\frac{e^{-int}}{-in}dt = \frac{1}{in}\int_{-\pi}^{\pi} f'(t)e^{-int}dt$$

d'où

$$\left|\int_{-\pi}^{\pi} f(t)e^{-int}dt\right| \leq \frac{1}{|n|}\int_{-\pi}^{\pi} |f'(t)|dt \xrightarrow[n\to\pm\infty]{} 0.$$

Soit maintenant $f \in L^1(-\pi,\pi)$ et $\varepsilon > 0$. Par le théorème de densité 5.45 du Chapitre IV, il existe $g \in C^1_c(]-\pi,\pi[)$ tel que $\|f-g\|_{L^1(\mathbb{T})} \leq \varepsilon$. D'après ce qui précède, il existe $n_0 \in \mathbb{N}$ tel que

$$\forall n \in \mathbb{Z} \text{ tel que } |n| \geq n_0,\quad |\widehat{g}(n)| \leq \varepsilon.$$

Ainsi on obtient

$$\forall n \in \mathbb{Z} \text{ tel que } |n| \geq n_0,\quad \left|\widehat{f}(n)\right| \leq \left|\widehat{f}(n) - \widehat{g}(n)\right| + |\widehat{g}(n)| \leq \|\widehat{f}-\widehat{g}\|_\infty + \varepsilon \leq \|f-g\|_{L^1(\mathbb{T})} + \varepsilon \leq 2\varepsilon.$$

Et donc par définition de la limite, on récupère bien

$$\widehat{f}(n) \xrightarrow[n\to\pm\infty]{} 0.$$

□

Voyons quelques propriétés élémentaires de $\widehat{\cdot}$. Rappelons qu'une bonne façon de voir cette opération est de la considérer comme un changement de variable, qui prend une fonction 2π-périodique et la transforme en une suite. Il est donc naturel de se demander comme se comporte ce changement de variable vis-à-vis des opérations sur les fonctions, ou de voir comment les propriétés des fonctions considérées se traduit sur leurs images par ce changement de variable.

Proposition 1.10. *Soit $(f,g) \in L^1(\mathbb{T})$.*

1. *pour $\alpha \in \mathbb{C}$, on a*

$$\widehat{\alpha f + g} = \alpha\widehat{f} + \widehat{g}.$$

18. Il existe plusieurs preuves de ce résultat, mais si vous faites une hypothèse assez faible sur f, on procède toujours par un argument de densité ; autrement dit, on procède toujours en deux étapes comme ici, la première étape pouvant varier, mais la seconde étape est toujours plus ou moins la même. L'autre méthode la plus classique pour la première étape consiste à prouver le résultat d'abord pour une fonction de la forme $\mathbb{1}_I$ où I est un intervalle, puis par linéarité on obtient le résultat pour les fonctions en escaliers. Ensuite, si vous êtes dans le cadre de l'intégrale de Riemann, vous pouvez approcher les fonctions continues ou continues par morceaux ou réglées par des fonctions en escaliers, pour la norme uniforme, et procéder comme la seconde étape ici ; si vous êtes dans le cadre de l'intégrale de Lebesgue, vous devez justifier que l'ensemble des fonctions en escaliers (et intégrable) est dense dans $L^1(-\pi,\pi)$ pour la norme $\|\cdot\|_{L^1(\mathbb{T})}$ (ce qui est vrai, mais est difficile aussi, à peu près au même titre que le théorème 5.45 du Chapitre IV qu'on utilise ici ; une argumentation pour voir la densité des fonctions en escalier est d'utiliser le théorème 5.38 du Chapitre IV pour approcher une fonction de L^1 par une fonction continue à support compact en norme 1, puis d'approcher cette dernière par une fonction en escalier en norme ∞, et de voir que ceci montre bien que la fonction initiale est proche d'une fonction en escalier en norme 1 (attention au jeu avec les deux normes)).

On peut aussi raisonner sur les polynomes trigonométriques si on sait qu'ils sont denses (voir la note 29), ou encore sur les fonctions continues (voir ce lien page 9).

2. *pour $\beta \in \mathbb{R}$, on considère $\tau_\beta f : x \mapsto f(x-\beta)$. Alors $\tau_\beta f \in L^1(\mathbb{T})$ et*

$$\forall n \in \mathbb{Z}, \quad \widehat{\tau_\beta f}(n) = e^{-in\beta}\widehat{f}(n).$$

3. *supposons $f \in C^1(\mathbb{T})$ [19]. Alors*

$$\forall n \in \mathbb{Z}, \quad \widehat{f'}(n) = in\widehat{f}(n). \tag{1.3}$$

En conséquence

$$\widehat{f}(n) = \underset{\pm\infty}{o}\left(\frac{1}{|n|}\right) \tag{1.4}$$

4. *On peut poser pour presque tout $x \in \mathbb{T}$,* [20]

$$f * g(x) = \frac{1}{2\pi}\int_{\mathbb{T}} f(t)g(x-t)dt$$

qu'on appelle convolution périodique de $(f,g) \in L^1(\mathbb{T})$. De plus, $f * g \in L^1(\mathbb{T})$ et

$$\|f*g\|_{L^1(\mathbb{T})} \leq \|f\|_{L^1(\mathbb{T})}\|g\|_{L^1(\mathbb{T})},$$

et

$$\forall n \in \mathbb{Z}, \quad \widehat{f*g}(n) = \widehat{f}(n)\widehat{g}(n).$$

5. *Pour $n \in \mathbb{Z}$, on pose $e_n : x \in \mathbb{T} \mapsto e^{inx}$. Alors*

$$e_n * f = \widehat{f}(n)e_n.$$

Plus généralement, si $P = \sum_{k=-N}^{N} a_k e_k$ est un polynome trigonométrique, alors

$$P * f = \sum_{k=-N}^{N} a_k \widehat{f}(k)e_k. \tag{1.5}$$

Remarque 1.11. La propriété 3 est centrale pour comprendre l'intérêt des séries de Fourier : en effet, l'opération de dérivation sur la fonction f, se traduit par une opération de multiplication (par in, $n \in \mathbb{Z}$) sur la suite $\widehat{f}$. Cette opération de multiplication est a priori beaucoup plus simple que l'opération initiale de dérivation.

Donc pour revenir à l'interprétation de $\widehat{\cdot}$ comme un changement de variable, on peut espérer que si on cherche à résoudre un "problème" dont l'inconnue est f et que ce problème fait intervenir la dérivée de f, la transformation de ce problème par Fourier sera plus simple à résoudre. Bien sûr, ceci se comprend sur les exemples, et dans le cas de l'équation de la chaleur, c'est une façon d'interpréter le succès des séries de Fourier dans ce problème.

Remarque 1.12. On peut itérer la propriété 3, et avec la note 19, on obtient : si $f \in C^{k-1}(\mathbb{T}) \cap C^k_m(\mathbb{T})$ avec $k \in \mathbb{N}^*$, alors

$$\forall n \in \mathbb{Z}, \quad \widehat{f^{(k)}}(n) = (in)^k\widehat{f}(n).$$

On en déduit également que [21]

$$\widehat{f}(n) = \underset{\pm\infty}{o}\left(\frac{1}{|n|^k}\right).$$

On peut retenir le principe suivant :

19. En fait, le résultat reste valable sous l'hypothèse plus faible que $f \in C^0(\mathbb{T}) \cap C^1_m(\mathbb{T})$ des fonctions continues, 2π-périodiques, et C^1 par morceaux ; la raison en est que l'intégration par partie faite dans la démonstration reste valable sous cette hypothèse plus faible. Notez que cette amélioration n'est pas inutile et sert régulièrement.

20. Attention, il ne faut pas confondre la convolution sur $\mathbb{R}$ et la convolution sur $\mathbb{T}$, c'est-à-dire des fonctions 2π-périodiques ; en effet, dans le premier cas le bon cadre est $L^1(\mathbb{R})$ et on intégère sur $\mathbb{R}$ (et il n'y a pas de coefficient $1/2\pi$), dans le second on intégre seulement sur un intervalle de longueur 2π, et on renormalise si on le souhaite (ce n'est pas une obligation). On ne prend pas la peine de distinguer les notations car aucune fonction (à part la fonction nulle) n'est à la fois intégrable sur $\mathbb{R}$ et 2π périodique.

21. Il est intéressant de noter que (1.3) et (1.4) ont chacun un intérêt en soit ; il arrive qu'on ait besoin de connaître explicitement le lien entre $\widehat{f}$ et $\widehat{f'}$, mais aussi qu'on ne s'intéresse qu'au comportement de $\widehat{f}$, mais dans ce dernier cas, il devient pertinent de réécrire $\widehat{f}$ en faisant intervenir $\widehat{f'}$ (donc en lisant la formule (1.3) à l'envers, si vous voulez, pour n'utiliser que le fait que $\widehat{f'} \in c_0(\mathbb{Z})$. Le fait que ça fait intervenir f' devient secondaire.

"La régularité de f se traduit par un comportement de $\widehat{f}$ en $\pm\infty$,"

ou encore

"Plus f est régulière, plus $\widehat{f}$ tend vite vers 0 en $\pm\infty$".

Ceci donne des informations sur $\widehat{X}$ quand X est un espace de fonctions régulières, voir la remarque 1.9.

Remarque 1.13. On peut voir la formule (1.5) comme un nouvel exemple du principe de "régularisation par convolution"[22] : en effet, en convolant un polynome trigonométrique avec une fonction seulement intégrable, on récupère un polynome trigonométrique, qui est bien sûr une fonction très régulière.

Remarque 1.14. La formule (1.5) montre qu'on peut aborder la question de la convergence des séries de Fourier comme une question sur la convolution. En effet, en l'appliquant au cas particulier de $S_N(f)$, on constate que

$$\forall f \in L^1(\mathbb{T}),\ \ \forall N \in \mathbb{N}, \quad S_N(f) = D_N * f$$

où

$$\forall N \in \mathbb{N},\ \ D_N = \sum_{k=-N}^{N} e_k.$$

Un calcul simple de série géométrique montre que[23]

$$\forall N \in \mathbb{N}, \forall x \in \mathbb{R} \setminus 2\pi\mathbb{Z}, \quad D_N(x) = \frac{\sin\left((N+1/2)x\right)}{\sin(x/2)}. \tag{1.6}$$

On l'appelle le noyau de Dirichlet. Il est naturel de se demander si $(D_N)_{N\in\mathbb{N}}$ constitue une identité approchée ; la réponse est non, ce qui explique la difficulté de la question posée par Fourier (voir la remarque 1.6) est délicate. Cela peut permettre aussi de suggérer que la réponse est non en général. On peut en fait rigoureusement déduire du fait que $(D_N)_N$ n'est pas une identité approchée, qu'il existe des fonctions dont la série de Fourier diverge (il faudrait préciser en quel sens), via l'utilisation du théorème de Baire (voir le Chapitre IX).

Démonstration de la proposition 1.10 :

1. C'est la linéarité de $\widehat{\cdot}$ qu'on a déjà évoqué à la proposition 1.7 et qui découle simplement de la linéarité de l'intégrale de Lebesgue.
2. En utilisant l'exercice 1.1, et avec un changement de variable, on constate que
$$\|\tau_\beta f\|_{L^1(\mathbb{T})} = \frac{1}{2\pi}\int_{\mathbb{T}} |f(t-\beta)|dt = \frac{1}{2\pi}\int_{\mathbb{T}} |f(t)|dt = \|f\|_{L^1(\mathbb{T})},$$
donc $\tau_\beta f \in L^1(\mathbb{T})$. De plus, par un calcul similaire, on obtient
$$\widehat{\tau_\beta f}(n) = \frac{1}{2\pi}\int_{\mathbb{T}} f(t-\beta)e^{-int}dt = \frac{1}{2\pi}\int_{\mathbb{T}} f(t)e^{-in(t+\beta)}dt = e^{-in\beta}\widehat{f}(n).$$
3. Soit $f \in C^1(\mathbb{T})$. Alors $f \in L^1(\mathbb{T})$ et $f' \in L^1(\mathbb{T})$. On peut donc écrire, par intégration par parties,
$$\forall n \in \mathbb{Z},\ \widehat{f'}(n) = \frac{1}{2\pi}\int_{\mathbb{T}} f'(t)e^{-int}dt = \underbrace{\left[\frac{1}{2\pi}f(t)e^{-int}\right]_{-\pi}^{\pi}}_{=0} - \frac{1}{2\pi}\int_{\mathbb{T}} f(t)(-in)e^{-int}dt = in\widehat{f}(n).$$
Pour la conséquence sur le comportement asymptotique de $\widehat{f}$, on utilise le lemme de Riemann-Lebesgue à $f' \in L^1(\mathbb{T})$.

22. Rappelons qu'on a vu dans le cas réel, et en faisant les bonnes hypothèses pour que les convolutions aient un sens, que $f * g$ a au moins la régularité de la plus régulière entre f et g.

23. On a écarté $x \in 2\pi\mathbb{Z}$ pour avoir un dénominateur non nul ; mais bien sûr la fonction D_N est régulière sur $\mathbb{R}$, ce qui ne saute pas aux yeux en regardant le terme de droite de l'équation. Je vous laisse quand même vous convaincre que cette formule définit bien une fonction continue en 0 (donc sur $2\pi\mathbb{Z}$ par périodicité), et qu'on retrouve bien la valeur de D_N en 0. Remarquez que ce genre de considération est un bon moyen de tester la véracité d'une telle formule.

4. La preuve de la première partie est la même que la preuve de la proposition 5.30 vue au Chapitre IV. La seconde partie s'obtient avec les théorèmes de Fubini-Tonelli et Fubini : en effet,

$$\int_{\mathbb{T}}\int_{\mathbb{T}} |f(t)g(x-t)e^{-inx}|dtdx = \int_{\mathbb{T}} |f(t)|dt \int_{\mathbb{T}} |g(t)|dt < +\infty$$

donc on peut appliquer Fubini et on obtient :

$$\widehat{f*g}(n) = \int_{\mathbb{T}}\int_{\mathbb{T}} f(t)g(x-t)e^{-inx}\frac{dt}{2\pi}\frac{dx}{2\pi} = \int_{\mathbb{T}} f(t)e^{-int}\underbrace{\left(\int_{\mathbb{T}} g(x-t)e^{-in(x-t)}\frac{dx}{2\pi}\right)}_{=\widehat{g}(n)}\frac{dt}{2\pi} = \widehat{f}(n)\widehat{g}(n).$$

5. Pour la première formule, il suffit d'un simple calcul ; la seconde formule s'obtient par linéarité de $*$.

1.2 Théorème de Féjer

Comme on l'a fait dans le cas de $\mathbb{R}^d$ au paragraphe 5.4 du Chapitre IV, on peut définir une notion d'identité approchée dans le cadre périodique :

Définition 1.15 (Identité approchée périodique)**.** Soit $(\rho_N)_{N\in\mathbb{N}}$ une famille de fonctions mesurables de $\mathbb{T}$ dans $\mathbb{C}$. On dit que $(\rho_N)_{N\in\mathbb{N}}$ est une identité approchée périodique si elle satisait :

1. il existe $M \in \mathbb{R}$ tel que
$$\forall N \in \mathbb{N},\quad \|\rho_N\|_{L^1(\mathbb{T})} \leq M,$$
2. pour tout $N \in \mathbb{N}$,
$$\frac{1}{2\pi}\int_{\mathbb{T}} \rho_N(t)dt = 1,$$
3. pour tout $\delta \in (0,\pi)$,
$$\lim_{N\to+\infty}\int_{]-\pi,\pi[\setminus]-\delta,\delta[} \rho_N(t)dt = 0.$$

Remarque 1.16. Comparons avec la définition 5.46 du Chapitre IV :

— pour la condition 2, il ne faut pas oublier le coefficient $1/2\pi$ de normalisation.

— on a indicé par $N \in \mathbb{N}$ au lieu de $\varepsilon > 0$; c'est juste un choix adapté aux exemples que nous allons le plus souvent rencontrer, on aurait tout-à-fait pu adapter la définition précédente au cadre d'un indice réel $\varepsilon > 0$ qui converge vers 0. Voir par exemple le noyau de Poisson ; voir aussi la remarque 5.48 au Chapitre IV.

— afin de pouvoir considérer la question "$(D_N)_{N\in\mathbb{N}}$ est-elle une identité approchée ?", nous avons donné une condition plus générale pour 1 ; la condition
$$\forall N \in \mathbb{N},\quad \rho_N \text{ est à valeur réelle et } \rho_N \geq 0$$
combinée à la condition 2 implique bien la condition 1. Voir aussi la remarque 5.49 au Chapitre IV.

On a déjà évoqué que $(S_N(f))_{N\in\mathbb{N}}$ peut se voir comme une convolution avec $(D_N)_{N\in\mathbb{N}}$. Mais on peut montrer [24] que $(D_N)_{N\in\mathbb{N}}$ n'est pas une identité approchée.

La solution proposée par le théorème de Féjer, consiste à étudier une notion plus faible de convergence de $S_N(f)$ vers f, qu'on appelle convergence au sens de Césaro. En effet, dans le cadre des suites réelles [25], on a vu que

$$\lim_{n\to+\infty} u_n = \ell \quad\Longrightarrow\quad \lim_{n\to+\infty}\frac{\sum_{k=0}^n u_k}{n+1} = \ell, \tag{1.7}$$

24. En montrant que $\|D_N\|_{L^1(\mathbb{T})}$ se comporte comme $\ln(N)$ en $+\infty$, à une constante multiplicative près.

25. Le cadre du théorème de Césaro est en fait beaucoup plus général, on peut l'énoncer dans un espace vectoriel normé quelconque, et la preuve est la même.

autrement dit, la notion de convergence classique est plus forte [26] que la notion de convergence des moyennes (partie de droite dans (1.7)), qu'on appelle convergence au sens de Césaro.

On définit donc, pour $f \in L^1(\mathbb{T})$,

$$\forall N \in \mathbb{N},\ \ \sigma_N(f) = \frac{S_0(f) + \cdots + S_N(f)}{N+1} = \frac{\sum_{n=0}^N S_n(f)}{N+1}.$$

Ainsi, grâce à la proposition 1.10 et à la remarque 1.14, on est amené à poser

$$\forall N \in \mathbb{N}, \quad K_N = \frac{\sum_{n=0}^N D_n}{N+1}$$

et alors

$$\forall f \in L^1(\mathbb{T}),\ \forall N \in \mathbb{N},\ \ \sigma_N(f) = K_N * f.$$

Remarque 1.17. On peut montrer

$$\forall N \in \mathbb{N},\ \ K_N = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) e_n,$$

et donc

$$\forall f \in L^1(\mathbb{T}),\ \ \sigma_N(f) = \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) \widehat{f}(n) e_n.$$

Mais contrairement au cas de $(D_N)_{N\in\mathbb{N}}$, on a le résultat suivant :

Théorème 1.18 (Théorème de Féjer)**.** *La famille $(K_N)_{N\in\mathbb{N}}$ est une identité approchée périodique.*

Remarque 1.19. Par rapport à la remarque 1.16 ci-dessus, $(K_N)_{N\in\mathbb{N}}$ est même à valeurs réelles (ce qui se voit facilement avec (1.6)) et positive (ce qui ne se voit pas facilement). D'ailleurs, pour cette raison, on peut complémenter le corollaire suivant 1.20 par la propriété :

$$\forall p \in [1, +\infty],\ \forall f \in L^p(\mathbb{T}), \forall N \in \mathbb{N}, \quad \|\sigma_N(f)\|_p \leq \|f\|_p,$$

voir le lemme 5.56 au Chapitre IV.

Démonstration. **Première étape : calcul de K_N :** Soit $N \in \mathbb{N}$. On a pour tout $x \in \mathbb{R} \setminus 2\pi\mathbb{Z}$,

$$\begin{aligned}
K_N(x) &= \frac{1}{N+1}\sum_{n=0}^{N} D_n(x) = \frac{1}{N+1}\sum_{n=0}^{N} \frac{\sin\left((n+1/2)x\right)}{\sin(x/2)} = \frac{1}{(N+1)\sin(x/2)} \mathrm{Im}\left(\sum_{n=0}^{N} e^{i(n+1/2)x}\right) \\
&= \frac{1}{(N+1)\sin(x/2)} \mathrm{Im}\left(e^{ix/2}\frac{e^{i(N+1)x}-1}{e^{ix}-1}\right) = \frac{1}{(N+1)\sin(x/2)} \mathrm{Im}\left(e^{i\frac{(N+1)x}{2}} \frac{\sin\left(\frac{(N+1)x}{2}\right)}{\sin(x/2)}\right) \\
&= \frac{1}{N+1}\left(\frac{\sin\left(\frac{(N+1)x}{2}\right)}{\sin(x/2)}\right)^2
\end{aligned}$$

Deuxième étape : vérification des 3 conditions d'une identité approchée. On a maintenant deux formules pour K_N, sa définition, et la formule montrée à l'étape précédente. On va utiliser l'une ou l'autre suivant les conditions à vérifier :

1. La formule montrée à l'étape 1 montre en particulier que K_N est réelle positive pour tout $N \in \mathbb{N}$, donc le fait que $(K_N)_{N\in\mathbb{N}}$ soit bornée dans $L^1(\mathbb{T})$ découle de la condition suivante.

26. Quand on dit "plus forte", comme souvent en mathématiques, on entend "plus forte au sens large", c'est-à-dire qu'on ne prétend pas que c'est strictement plus fort ; vous devez savoir néanmoins que le théorème de Césaro n'admet pas de réciproque en général, donc il existe des suites qui convergent au sens de Césaro, mais qui ne converge pas au sens classique.

2. Pour montrer que pour tout $N \in \mathbb{N}$, $\|K_N\|_{L^1(\mathbb{T})}$, le piège est qu'il ne faut pas utiliser la formule montrée à l'étape 1, mais la définition :

$$\forall N \in \mathbb{N}, \quad \frac{1}{2\pi}\int_{\mathbb{T}} K_N(t)dt = \frac{1}{N+1}\left(\sum_{n=0}^{N} \frac{1}{2\pi}\int_{\mathbb{T}} D_n(t)dt\right) = \frac{N+1}{N+1} = 1$$

où on a utilisé que $\frac{1}{2\pi}\int_{\mathbb{T}} D_n(t)dt = 1$ pour tout $n \in \mathbb{N}$, ce qui découle aussi de sa définition $D_n = \sum_{k=-n}^{n} e_k$.
3. Soit $\delta \in]0, \pi[$. Alors

$$\begin{aligned}\forall N \in \mathbb{N}, \quad \frac{1}{2\pi}\int_{]-\pi,\pi[\setminus]-\delta,\delta[} K_N(t)dt &= \frac{1}{\pi}\int_{\delta}^{\pi} K_N(t)dt \le \frac{1}{\pi(N+1)}\int_{\delta}^{\pi}\left(\frac{1}{\sin(\delta/2)}\right)^2 dt \\ &\le \frac{1}{(N+1)\sin^2(\delta/2)} \xrightarrow[N\to+\infty]{} 0,\end{aligned}$$

d'où le résultat.

□

Corollaire 1.20. — *Soit $f \in C^0(\mathbb{T})$. Alors*

$$\sigma_N(f) \xrightarrow[N\to+\infty]{\|\cdot\|_\infty} f.$$

— *Soit $p \in [1, +\infty[$ et $f \in L^p(\mathbb{T})$. Alors*

$$\sigma_N(f) \xrightarrow[N\to+\infty]{\|\cdot\|_{L^p(\mathbb{T})}} f.$$

Remarque 1.21. Dans la littérature, c'est plutôt ce corollaire qui est appelé théorème de Féjer ; j'ai préféré distinguer ainsi car ce corollaire est un résultat classique, valable pour toute identité approchée, et non seulement pour $(K_N)_{N\in\mathbb{N}}$, voir aussi le lemme 5.56 au Chapitre IV. La spécificité et la "nouveauté" du théorème de Féjer réside bien dans le fait de savoir que $(K_N)_{N\in\mathbb{N}}$ est une identité approchée. [27]

27. Le corollaire 1.20 (sous la dénomination "théorème de Féjer") est d'ailleurs souvent proposé en développement. C'est une bonne idée, il y a une partie calculatoire, une partie d'estimation, une partie plus abstraite sur des convergences de fonctions, donc on satisfait à plusieurs attentes du jury ; c'est aussi de longueur adaptable, car on peut choisir de ne développer qu'une partie du corollaire. Mais attention, comme à chaque fois, il faut bien y réfléchir pour justifier sa place dans la leçon dans laquelle vous envisagez de proposer ce développement. En effet, quel que soit le nom donné, il y a deux étapes bien distinctes :

1. montrer que $(K_N)_{N\in\mathbb{N}}$ est une identité approchée,
2. montrer que si $(\rho_N)_{N\in\mathbb{N}}$ est une identité approchée, alors sous les bonnes hypothèses sur f, $\rho_N * f$ converge vers f en un sens à préciser (vous pouvez très bien ne donner cet énoncé que pour $(K_N)_{N\in\mathbb{N}}$, il n'empêche que c'est vrai en général et que la démonstration est la même).

Si on souhaite proposer ce développement dans la leçon **246** sur les séries de Fourier, il me semble important d'avoir quand même conscience que la moitié de la preuve (je ne me lance pas dans une estimation affinée, à vous de chronométrer) concerne la convolution, et pas tellement, a priori, les séries de Fourier. Néanmoins :

— une moitié c'est déjà pas mal ! A condition bien sûr que vous la fassiez bien et qu'elle représente une partie substantielle de votre développement ; un candidat qui ne développerait que la seconde partie, et expédierait la première par manque de temps friserait le hors sujet !

— le résultat est tellement important dans la théorie des séries de Fourier, que ça justifie sa proposition en tant que développement (à condition que vous ayez bien-sûr conscience de cette importance, et que vous ayez une bonne vision des résultats qui découlent du théorème de Féjer),

— la convolution est en réalité très liée à l'analyse des séries de Fourier (pour le théorème de Féjer, mais pas seulement) ; si vous en avez conscience et que vous arrivez à le faire savoir au jury, alors une disgression sur la convolution et les identités approchées trouve finalement tout-à-fait sa place dans la leçon **246**.

Si vous proposez ce développement dans une autre leçon, il peut tout-à-fait arriver que la situation soit inversée, au sens où ce serait peut-être la seconde étape qui serait plus pertinente dans la leçon. Auquel cas vous pouvez commencer par développer celle-ci, et réserver la fin du développement au fait que $(K_N)_{N\in\mathbb{N}}$ est un exemple.

Démonstration. La preuve est la même que la preuve du lemme 5.56 au Chapitre IV. □

Remarque 1.22. Les conséquences de ce corollaire sont très importantes :

— le corollaire prouve que $\widehat{\cdot}$ est injective. En effet, si $f \in L^1(\mathbb{T})$ est tel que $\widehat{f} = 0$, alors pour tout $N \in \mathbb{N}, \sigma_N(f) = 0$ [28]. Et donc par le corollaire 1.20, $f = 0$ presque partout.

— une autre conséquence concerne la densité des polynomes trigonométriques : en effet, comme $\sigma_N(f)$ est lui-même un polynôme trigonométrique, on en déduit que l'ensemble des polynômes trigonométriques est dense dans $(C^0(\mathbb{T}), \|\cdot\|_\infty)$ (résultat qu'on appelle parfois "théorème de Weierstrass trigonométrique") et dans $(L^p(\mathbb{T}), \|\cdot\|_{L^p(\mathbb{T})})$ pour $p \in [1, +\infty[$. [29]

— si on combine avec le théorème de Césaro, on déduit la propriété intéressante suivante : si on sait que $(S_N(f))_{N\in\mathbb{N}}$ converge (disons dans $C^0(\mathbb{T})$ pour fixer les idées) et que $f \in C^0(\mathbb{T})$, alors la limite de $(S_N(f))_{N\in\mathbb{N}}$ ne peut être que f (car d'après le théorème de Césaro dans un cadre général d'espace vectoriel normé (dont la preuve est la même que dans le cas réel), si on note g la limite de $(S_N(f))_{N\in\mathbb{N}}$, alors $(\sigma_N(f))_{N\in\mathbb{N}}$ converge vers g, donc par unicité de la limite $f = g$). Cette remarque est valable si on remplace $(C^0(\mathbb{T}), \|\cdot\|_\infty)$ par $(L^p(\mathbb{T}), \|\cdot\|_{L^p(\mathbb{T})})$ pour $p \in [1, +\infty[$ et plus généralement par toute convergence qui fonctionne bien avec les identités approchées, voir la remarque 1.23

Remarque 1.23. Soit $f \in C^0_m(\mathbb{T})$. On peut montrer que

$$\sigma_N(f)(x_0) \xrightarrow[N\to+\infty]{} \frac{1}{2}\Big[f(x_0^+) + f(x_0^-)\Big].$$

En particulier si f est continue en x_0, cette limite vaut $f(x_0)$.

1.3 Convergence ponctuelle et uniforme

Le premier résultat s'intéresse à la convergence ponctuelle :

Théorème 1.24 (Théorème de Dirichlet)**.** *Soit $f \in C^1_m(\mathbb{T})$. Alors pour tout $x_0 \in \mathbb{T}$,*

$$S_N(f)(x_0) \xrightarrow[N\to+\infty]{} \frac{f(x_0^+) + f(x_0^-)}{2}.$$

Remarque 1.25. Ce résultat est sans doute le plus connu des étudiants. Il permet notamment de calculer des séries particulières, comme $\sum_{n\in\mathbb{N}^*} \frac{1}{n^2}$ et plein d'autres. Mais

— on ne doit pas penser que les séries de Fourier n'ont d'intérêt que pour calculer des sommes originales,

— la convergence ponctuelle ne tient pas vraiment compte de l'aspect fonctionnel des objets mis en jeu ;

— l'énoncé fait une hypothèse un peu surprenante sur le caractère C^1 par morceaux, alors que le résultat ne laisse pas présager de l'importance d'une dérivabilité.

La remarque à retenir de ce long discours, c'est que vous devez passer du temps à mesurer la pertinence de votre développement dans une leçon donnée (qui se mesure par les deux critères principaux :

— "le résultat a-t-il une importance particulière vis-à-vis de la leçon ?"

— "la démonstration du résultat utilise-t-elle de manière substantielle les outils du plan (à condition bien sûr que le plan ne soit pas lui-même hors sujet !) ?"

(notons qu'il n'est pas nécessaire de satisfaire les deux critères, mais indiscutable qu'il faut en remplir au moins un). Et que si vous souhaitez "recaser" un développement, il arrive souvent qu'on ait besoin d'un peu d'ajustements pour que ça colle vraiment bien à la leçon que vous traitez.

28. L'application $\widehat{\cdot}$ étant linéaire, il suffit d'étudier son noyau pour savoir qu'elle est injective.

29. Remarquons qu'on peut en déduire une preuve un peu plus expéditive du lemme de Riemann-Lebesgue ; en effet, au lieu de faire le lemme pour $f \in C^1_m(\mathbb{T})$, on peut le faire pour P un polynôme trigonométrique pour lequel le résultat est évident, et conclure en utilisant la densité des polynomes trigonométriques dans $L^1(\mathbb{T})$. Voir par exemple [Kat04].

Remarque 1.26. On peut généraliser le résultat précédent en supposant seulement f 2π-périodique, localement intégrable, et x_0 tel que f admet des limites à gauches et à droite en x_0 et il existe $\alpha > 0$ tel que

$$\int_0^\alpha \frac{|f(x_0+t)| - f(x_0^+)|}{t} dt < +\infty, \quad \text{et} \quad \int_0^\alpha \frac{|f(x_0-t)| - f(x_0^-)|}{t} dt < +\infty.$$

Remarque 1.27. Il faut attendre les années 60 pour clore la question de la convergence ponctuelle des séries de Fourier. En effet, on peut montrer (c'est très difficile) que si $f \in L^p(\mathbb{T})$ avec $p \in]1, +\infty[$, alors

$$S_N(f)(x) \xrightarrow[N\to+\infty]{} f(x) \text{ pour presque tout } x \in \mathbb{T}.$$

Ce résultat est dû à Carleson (1966) dans le cas $p = 2$ et a été généralisé à $p > 1$ par Hunt en 1968.

En 1926, Kolmogorov avait exhibé une fonction intégrable ($L^1(\mathbb{T})$) dont la série de Fourier diverge en tout point.

Démonstration. Pour $N \in \mathbb{N}$, par parité de D_N on a :

$$\begin{aligned} S_N(f)(x_0) &= D_N * f(x_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\sin((N+1/2)t)}{\sin(t/2)} f(x_0 - t) dt \\ &= \frac{1}{2\pi} \int_0^\pi \sin((N+1/2)t) \frac{f(x_0-t) + f(x_0+t)}{\sin(t/2)} dt \end{aligned}$$

et on utilise que $\frac{1}{2\pi} \int_0^\pi D_N(t)dt = 1/2$ pour avoir

$$S_N(f)(x_0) - \frac{f(x_0^+) + f(x_0^-)}{2} = \frac{1}{2\pi} \int_0^\pi \sin((N+1/2)t) \frac{\Big[f(x_0-t) - f(x_0^-)\Big] + \Big[f(x_0+t) - f(x_0^+)\Big]}{\sin(t/2)} dt$$

or par dérivabilité de f à droite en x_0, on a

$$\frac{f(x_0+t) - f(x_0^+)}{\sin(t/2)} \xrightarrow[t\to 0^+]{} 2f'(x_0^+), \qquad \text{et} \qquad \frac{f(x_0-t) - f(x_0^-)}{\sin(t/2)} \xrightarrow[t\to 0^+]{} -2f'(x_0^-)$$

donc la fonction $t \mapsto \dfrac{\Big[f(x_0-t) - f(x_0^-)\Big] + \Big[f(x_0+t) - f(x_0^+)\Big]}{\sin(t/2)}$ est intégrable sur $]0, \pi]$, donc par le lemme 1.8 de Riemann-Lebesgue, on obtient le résultat. □

On donne maintenant une condition de convergence uniforme de la série de Fourier :

Théorème 1.28. *Soit $f \in C^0(\mathbb{T})$ telle que $\widehat{f} \in \ell^1(\mathbb{Z})$. Alors*

$$S_N(f) \xrightarrow[N\to+\infty]{} f \text{ uniformément sur } \mathbb{T}.$$

Remarque 1.29. Comparez avec le théorème d'inversion de Fourier au paragraphe 2 !

Remarque 1.30. La continuité de f est en fait une condition nécessaire de convergence uniforme de la série de Fourier. Si on suppose seulement $f \in L^1(\mathbb{T})$ et $\widehat{f} \in \ell^1(\mathbb{Z})$, on peut en fait déduire de la même preuve que ci-dessous que f a un représentant $\widetilde{f} \in C^0(\mathbb{T})$ (à savoir la limite uniforme de $(S_N(f))_{N\in\mathbb{N}}$), et donc en particulier

$$S_N(f)(x) \xrightarrow[N\to+\infty]{} f(x) \text{ pour presque tout } x \in \mathbb{T}.$$

Remarquons aussi que si $f \in L^1(\mathbb{T})$ n'a pas de représentant continu, alors par contraposée, $\widehat{f} \notin \ell^1(\mathbb{Z})$. Ainsi la non-régularité de f se traduit par un comportement à l'infini de $\widehat{f}$ (qui ne peut pas tendre vers 0 trop vite ici).

Remarque 1.31. Si vous avez accès explicitement à $\widehat{f}$, il est en général facile de vérifier l'hypothèse $\widehat{f} \in \ell^1(\mathbb{Z})$, mais il est fréquent de ne pas connaître explicitement $\widehat{f}$. On peut alors utiliser le principe du lien entre la régularité de f et le comportement en l'infini de $\widehat{f}$:

— si $f \in C^2(\mathbb{T})$, alors par la proposition 1.10 et la remarque 1.12, on en déduit

$$\widehat{f}(n) = \underset{\infty}{o}\left(\frac{1}{n^2}\right)$$

donc $\widehat{f} \in \ell^1(\mathbb{Z})$,

— on peut largement améliorer le résultat précédent et montrer que

$$f \in C^0(\mathbb{T}) \cap C^1_m(\mathbb{T}) \quad \Rightarrow \quad \widehat{f} \in \ell^1(\mathbb{Z}),$$

voir le corollaire 1.36.

Démonstration. L'hypothèse $\widehat{f} \in \ell^1(\mathbb{Z})$ se traduit par

$$\sum_{n \in \mathbb{Z}} \|\widehat{f}(n) e_n\|_\infty < \infty$$

c'est-à-dire que la série de Fourier de f converge normalement, donc uniformément par complétude de $C^0(\mathbb{T})$ (voir la proposition 4.17 au Chapitre I).

Mais par une conséquence du théorème de Féjer (voir le dernier point de la remarque 1.22), la limite de $(S_N(f))_{N\in\mathbb{N}}$ au sens uniforme ne peut être que f, d'où le résultat. □

1.4 Cadre $L^2(\mathbb{T})$

Ce paragraphe nous permet de voir la particularité de l'espace $L^2(\mathbb{T})$ dans la famille des espaces L^p pour $p \in [1, +\infty]$, et qui permet de donner une interprétation géométrique à la question de la convergence de la série de Fourier.

Rappelons en effet que $L^2(\mathbb{T})$ muni de la norme $f \in L^2(\mathbb{T}) \mapsto \|f\|_{L^2(\mathbb{T})} = \sqrt{\frac{1}{2\pi}\int_{\mathbb{T}} |f(t)|^2 dt}$ est un espace de Banach et qu'on peut le munir d'un produit scalaire dont cette norme est issue :

$$\forall (f,g) \in L^2(\mathbb{T}),\ \langle f, g\rangle_{L^2(\mathbb{T})} := \frac{1}{2\pi}\int_{\mathbb{T}} f(t)\overline{g(t)}dt,$$

ce qui fait de $L^2(\mathbb{T})$ un espace de Hilbert : on renvoit au Chapitre IX pour une étude générale des espaces de Hilbert.

Remarquons également que la famille $(e_n)_{n\in\mathbb{Z}}$ est orthonormée pour ce produit scalaire, c'est-à-dire

$$\forall (i,j) \in \mathbb{Z}^2,\ \ \langle e_i, e_j\rangle_{L^2(\mathbb{T})} = \delta_{i,j}.$$

Enfin,

$$S_N(f) = \sum_{n=-N}^{N} \widehat{f}(n) e_n = \sum_{n=-N}^{N} \langle f, e_n\rangle_{L^2(\mathbb{T})} e_n \tag{1.8}$$

est la projection de f sur $\mathcal{P}_N = \mathrm{Vect}((e_n)_{n\in[\![-N,N]\!]})$ l'ensemble des polynomes trigonométriques de "degré" intérieur ou égal à N, c'est-à-dire que c'est le polynome trigonométrique le plus proche de f au sens de la norme $\|\cdot\|_{L^2(\mathbb{T})}$, comme on le verra dans la preuve du théorème 1.32

Théorème 1.32. *Soit $f \in L^2(\mathbb{T})$. Alors*

$$S_N(f) \xrightarrow[N\to+\infty]{\|\cdot\|_{L^2(\mathbb{T})}} f.$$

De plus,

$$\begin{array}{rccc} \widehat{\cdot}\ : & L^2(\mathbb{T}) & \longrightarrow & \ell^2(\mathbb{Z}) \\ & f & \longmapsto & \widehat{f} \end{array}$$

est une isométrie surjective[30]*, c'est-à-dire*

30. Rappelons qu'une isométrie linéaire est toujours injective; donc évidemment l'isométrie proposée ici est en fait bijective, mais on insiste sur la surjectivité qui est la vraie "nouveauté".

— *soit* $f \in L^1(\mathbb{T})$ *; alors* $f \in L^2(\mathbb{T})$ *si et seulement si* $\widehat{f} \in \ell^2(\mathbb{Z})$*, et*

$$\|f\|_{L^2(\mathbb{T})} = \|\widehat{f}\|_{\ell^2(\mathbb{Z})}$$

c'est-à-dire

$$\frac{1}{2\pi}\int_{\mathbb{T}} |f(t)|^2 dt = \sum_{n\in\mathbb{Z}} |\widehat{f}(n)|^2 \tag{1.9}$$

— *si* $(c_n) \in \ell^2(\mathbb{Z})$ *alors il existe une unique fonction* $f \in L^2(\mathbb{T})$ *telle que* $\forall n \in \mathbb{Z},\ c_n = \widehat{f}(n)$*.*

Remarque 1.33. En rapport à la remarque 1.9, on voit que l'espace $L^2(\mathbb{T})$ fait partie des espaces pour lesquels on sait identifier explicitement "l'espace de Fourier" associé.

Remarque 1.34. L'égalité (1.18) s'appelle l'égalité de Parseval. Il arrive qu'elle permette de calculer des sommes originales, mais il serait dommage de résumer ce théorème à cette égalité et à ces applications.

Remarque 1.35. Comparez avec le théorème de Plancherel au paragraphe 2. Ici c'est un peu plus simple car $L^2(\mathbb{T}) \subset L^1(\mathbb{T})$, donc le cadre $L^2(\mathbb{T})$ est une restriction du cadre $L^1(\mathbb{T})$, ce qui n'est plus le cas si on remplace $\mathbb{T}$ par $\mathbb{R}$.

Démonstration. **Étape 1 (rappel de géométrie) :** la formule (1.8) donne le projeté orthogonal de f sur $\mathcal{P}_N$ pour $N \in \mathbb{N}$, au sens où $S_N(f)$ est l'élément de $\mathcal{P}_N$ le plus proche de f au sens de la distance associée à $\|\cdot\|_{L^2(\mathbb{T})}$. En effet, pour $N \in \mathbb{N}$ et $Q \in \mathcal{P}_N$, on a [31]

$$\|f - Q\|^2_{L^2(\mathbb{T})} = \|f - S_N(f)\|^2_{L^2(\mathbb{T})} + \|S_N(f) - Q\|^2_{L^2(\mathbb{T})} \tag{1.10}$$

car $f - S_N(f) \in \mathcal{P}_N^{\perp}$ et $S_N(f) - Q \in \mathcal{P}_N$. D'où

$$\forall N \in \mathbb{N},\ \ \forall Q \in \mathcal{P}_N,\ \ \|f - S_N(f)\|_{L^2(\mathbb{T})} \leq \|f - Q\|_{L^2(\mathbb{T})}. \tag{1.11}$$

Étape 2 : convergence $L^2(\mathbb{T})$. par le théorème de Féjer, ceci implique bien la convergence $L^2(\mathbb{T})$ de $(S_N(f))_{N\in\mathbb{N}}$. En effet, par (1.11) et par le corollaire 1.20, on a

$$\|f - S_N(f)\|_{L^2(\mathbb{T})} \leq \|f - \sigma_N(f)\|_{L^2(\mathbb{T})} \xrightarrow[N\to+\infty]{} 0.$$

Étape 3 : isométrie. En appliquant (1.10) avec $Q = 0$, on a en particulier

$$\|f\|^2_{L^2(\mathbb{T})} = \|f - S_N(f)\|^2_{L^2(\mathbb{T})} + \|S_N(f)\|^2_{L^2(\mathbb{T})}$$

or un calcul élémentaire [32] donne

$$\|S_N(f)\|^2_{L^2(\mathbb{T})} = \sum_{n=-N}^{N} |\widehat{f}(n)|^2.$$

En faisant tendre N vers $+\infty$ et en appliquant l'étape précédente [33], on obtient à la limite

$$\|f\|^2_{L^2(\mathbb{T})} = \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2.$$

31. C'est en quelque sorte le théorème de Pythagore !
32. Qui n'est autre que le théorème de Pythagore avec $N+1$ vecteurs orthogonaux 2 à 2.
33. Remarquez que si on ignore l'étape 2 de convergence $L^2(\mathbb{T})$, on récupère quand même l'inégalité

$$\|f\|^2_{L^2(\mathbb{T})} \geq \sum_{n=-\infty}^{\infty} |\widehat{f}(n)|^2,$$

qu'on appelle souvent inégalité de Bessel.

Étape 4 : surjectivité. Soit $(c_n)_{n\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$. On peut voir que $\left(S_N := \sum_{n=-N}^{N} c_n e_n\right)_{N\in\mathbb{N}}$ est une suite de Cauchy dans $L^2(\mathbb{T})$: en effet

$$\forall N \in \mathbb{N}, \forall p \in \mathbb{N}, \quad \|S_{N+p} - S_N\|^2_{L^2(\mathbb{T})} = \sum_{N+1\leq |n| \leq N+p} |c_n|^2 \leq \sum_{|n|\geq N+1} |c_n|^2 \xrightarrow[N\to+\infty]{} 0,$$

et par complétude de $L^2(\mathbb{T})$, S_N a une limite pour la norme $L^2(\mathbb{T})$, qu'on appelle f. De plus, pour $k \in \mathbb{Z}$ on a d'une part

$$\forall N > |k|, \quad \langle S_N, e_k\rangle_{L^2(\mathbb{T})} = c_k,$$

et d'autre part par continuité du produit scalaire,

$$\langle S_N, e_k\rangle_{L^2(\mathbb{T})} \xrightarrow[N\to+\infty]{} \langle f, e_k\rangle_{L^2(\mathbb{T})} = \widehat{f}(k)$$

d'où $\widehat{f}(k) = c_k$ pour tout $k \in \mathbb{Z}$. □

Corollaire 1.36. *Soit $f \in C^0(\mathbb{T}) \cap C^1_m(\mathbb{T})$. Alors*

$$\widehat{f} \in \ell^1(\mathbb{Z}).$$

Démonstration. D'après le point 3 de la proposition 1.10 et la note 19, on peut écrire

$$\forall n \in \mathbb{Z}^*, \quad \widehat{f}(n) = \frac{1}{in}\widehat{f'}(n),$$

et donc par l'inégalité de Cauchy-Schwartz dans $\ell^2(\mathbb{Z})$, on obtient

$$\sum_{n\in\mathbb{Z}} |\widehat{f}(n)| \leq |\widehat{f}(0)| + \sqrt{\sum_{n\in\mathbb{Z}^*} \frac{1}{n^2}}\sqrt{\sum_{n\in\mathbb{Z}^*} |\widehat{f'}(n)|^2}$$

mais comme $f' \in C^0_m(\mathbb{T}) \subset L^2(\mathbb{T})$, par le théorème 1.32, $\widehat{f'} \in \ell^2(\mathbb{Z})$, d'où le résultat. □

1.5 Compléments

- Défauts de convergence :
  - On peut montrer qu'il existe des fonctions continues dont la série de Fourier diverge en un point, soit par des méthodes explicites[34] ou par des constructions d'analyse fonctionnelle avec le théorème de Baire (voir le Chapitre IX).
  - Si on prend une fonction régulière par morceaux, telle que $f(x) = \frac{f(x^+)+f(x^-)}{2}$ en tout point $x \in \mathbb{T}$, le théorème de Dirichlet affirme la convergence simple de $(S_N(f))_{N\in\mathbb{N}}$ vers f. Si f n'est pas continue, la convergence ne peut pas être uniforme. On peut même quantifier ce fait dans des exemples explicites, comme

    $$\forall x \in]-\pi,\pi], \quad f(x) = \begin{cases} 1 \text{ si } x \in]0,\pi[\\ 1/2 \text{ si } x \in \{0,\pi\} \\ 0 \text{ si } x \in]-\pi, 0[\end{cases}.$$

    C'est ce qu'on appelle le phénomène de Gibbs, voir par exemple [Gou08b, Exercice 6 page 266].
- on peut améliorer les estimations sur le comportement des coefficients de Fourier en fonction de la régularité de f :

34. Un exemple fourni par du Bois Reymond en 1876 : la série

$$\forall x \in [-\pi,\pi], \quad \sum_{n\in\mathbb{N}^*} \frac{1}{n^2} \sin\left(\frac{2^{n^3}+1}{2}|x|\right)$$

définit une fonction continue sur $\mathbb{T}$ dont la série de Fourier diverge en 0 : voir [Gou08b, Exercice 4 page 264].

  - on peut améliorer le corollaire 1.36 et montrer que si $f \in C^{0,\alpha}(\mathbb{T})$ avec $\alpha > 1/2$, alors $\widehat{f} \in \ell^1(\mathbb{Z})$ (et donc le théorème 1.28 s'applique).
  - on peut montrer que si f est à variations bornées, alors $\widehat{f}(n) = O_{\pm\infty}\left(\frac{1}{|n|}\right)$.
- on peut obtenir d'autres conditions pour la convergence de $S_N(f)$, notamment un résultat dû à Hardy : si $\widehat{f}(n) = O_{\pm\infty}\left(\frac{1}{|n|}\right)$, alors $S_N(f)(x)$ et $\sigma_N(f)(x)$ convergent simplement pour les mêmes valeurs de $x \in \mathbb{T}$ et uniformément sur les mêmes ensembles (voir [Kat04, Paragraphe 2.2]).
- On peut trouver des applications intéressantes des séries de Fourier :
  - la formule sommatoire de Poisson, voir le paragraphe 2.
  - la résolution d'équations aux dérivées partielles, comme l'équation de la chaleur sur $\mathbb{T}$ ou sur un segment (voir [DM72, page 63-68])
  - des inégalités fonctionnelles, comme l'inégalité de Wirtinger (voir [Gou08b, page 264])

$$\forall f \in C^1(\mathbb{T}) \text{ telle que } \int_{\mathbb{T}} f(t)dt = 0, \quad \|f\|_{L^2(\mathbb{T})} \leq \|f'\|_{L^2(\mathbb{T})}.$$

2 Transformation de Fourier

Cette fois-ci, les fonctions sont définies sur $\mathbb{R}$ [35]. Le fait que cet ensemble n'est pas compact sera une différence importante avec le paragraphe précédent. En effet, on a les inclusions $C^0(\mathbb{T}) \subset L^2(\mathbb{T}) \subset L^1(\mathbb{T})$, mais aucune inclusion entre les espaces $C_b^0(\mathbb{R}), L^2(\mathbb{R}), L^1(\mathbb{R})$. Précisons d'ailleurs que l'on travaille avec l'espace mesuré $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda)$.

Il ne coûte pas grand chose de généraliser à des fonctions définies sur $\mathbb{R}^N$ avec $N \geq 2$. Par simplicité, on préfère présenter les choses sur $\mathbb{R}$ en premier lieu, et on fera quelques remarques sur le cas de la dimension supérieure. Mais outre quelques adaptations mineures via des notions élémentaires de fonctions de plusieurs variables, tout se passe de façon similaire.

2.1 Cadre $L^1(\mathbb{R})$

2.1.1 Définition

Définition-Proposition 2.1. Soit $f \in L^1(\mathbb{R})$. On peut poser [36]

$$\forall \xi \in \mathbb{R}, \quad \widehat{f}(\xi) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\xi x} dx. \tag{2.1}$$

De plus, $\widehat{f}$ est une fonction continue, bornée, et qui tend vers 0 [37] quand ξ tend vers $\pm\infty$, et

$$\|\widehat{f}\|_\infty \leq \frac{1}{\sqrt{2\pi}} \|f\|_1. \tag{2.2}$$

On appelle $\widehat{f}$ la transformée de Fourier de f.
Ainsi l'opérateur [38]

$$\begin{array}{rccc} \mathscr{F} : & L^1(\mathbb{R}) & \longrightarrow & C^0_{\to 0}(\mathbb{R}) \\ & f & \longmapsto & \widehat{f} \end{array}$$

35. Et toujours à valeurs dans $\mathbb{C}$.

36. Remarquons qu'on a choisi pour variable de $\widehat{f}$ la lettre "xi" ξ, un peu inhabituelle (on conseille le lecteur de s'entraîner à l'écrire suffisamment de fois pour que ça devienne un automatisme), mais néanmoins répendue dans ce cadre. Bien sûr, du point de vue de la rigueur ce choix n'a aucune importance. Mais on conseille quand même le lecteur de choisir une lettre atypique comme variable "de Fourier", surtout parce qu'il ne faudra pas la confondre avec la variable initiale (qu'on a notée x ici, et qui figure comme variable d'intégration). Dans le cadre des séries de Fourier, il n'y avait pas vraiment de confusion possible car le rôle de ξ était joué par $n \in \mathbb{Z}$, qui pouvait difficilement être confondu avec la variable $x \in \mathbb{T}$. Mais ici l'espace des x et des ξ (des variables initiales et de Fourier) étant les mêmes, il faut redoubler de vigilance.

37. Il y a une redondance volontaire dans cet énoncé, puisqu'une fonction continue qui tend vers 0 est nécessairement bornée. On a gardé cette formulation pour distinguer en premier lieu que $\widehat{f}$ est bornée, ce qui est élémentaire, et en second lieu qu'elle tend vers 0, ce qui n'est pas élémentaire.

38. Qu'on appelle transformation de Fourier.

est bien défini, linéaire et continu [39], où on a muni $L^1(\mathbb{R})$ de la norme $\|\cdot\|_1$ et

$$C^0_{\to 0}(\mathbb{R}) = \{g \in C^0(\mathbb{R}),\ \lim_{\xi\to\pm\infty} g(\xi) = 0\}$$

de la norme $\|\cdot\|_\infty$.

Remarque 2.2. L'espace $C^0_{\to 0}(\mathbb{R})$ est un sous-espace vectoriel fermé de $(C^0_b(\mathbb{R}), \|\cdot\|_\infty)$ et est donc un espace de Banach.

Remarque 2.3. Dans l'espace de départ de $\mathscr{F}$ on travaille avec des classes de fonctions, alors que dans l'espace d'arrivée on a des "vraies" fonctions. Notons que ceci est légitime car si f et g sont égales presque partout, alors $\widehat{f}$ et $\widehat{g}$ sont égales partout.

Remarque 2.4. Il existe différentes conventions pour la définition de la transformation de Fourier ; on rencontre par exemple

$$\xi \longmapsto \int_{\mathbb{R}} f(x)e^{-i\xi x}dx, \qquad \text{ou} \qquad \xi \longmapsto \int_{\mathbb{R}} f(x)e^{-2i\pi\xi x}dx,$$

qui ont leurs avantages et leurs inconvénients par rapport à (2.1).

Notons aussi qu'en probabilités, on utilisera plutôt la convention [40] :

$$t \longmapsto \int_{\mathbb{R}} f(x)e^{itx}dx$$

qu'on généralisera au cadre des mesures.

Démonstration. Soit $f \in L^1(\mathbb{R})$. La fonction $\widehat{f}$ est bien définie car pour tout $\xi \in \mathbb{R}$,

$$\int_{\mathbb{R}} |f(x)e^{-i\xi x}|dx = \|f\|_1 < +\infty.$$

Pour montrer la continuité de $\widehat{f}$, on applique le théorème de régularité des intégrales à paramètre (théorème 3.16 au Chapitre IV) : en effet on vérifie bien les hypothèses :

— à $\xi \in \mathbb{R}$ fixé, la fonction $x \mapsto f(x)e^{-i\xi x}$ est mesurable car f est mesurable et $x \mapsto e^{-i\xi x}$ est continue, [41],

— à $x \in \mathbb{R}$ fixé, la fonction $\xi \mapsto f(x)e^{-i\xi x}$ est continue,

— on a la domination

$$\forall \xi \in \mathbb{R},\quad |f(x)e^{-i\xi x}| \leq |f(x)| \quad \text{pour presque tout } x \in \mathbb{R}$$

avec $f \in L^1(\mathbb{R})$ indépendante de ξ.

On peut maintenant constater que $\widehat{f}$ est bornée du fait de l'inégalité

$$\forall \xi \in \mathbb{R},\quad \left|\widehat{f}(\xi)\right| \leq \frac{1}{\sqrt{2\pi}}\left|\int_{\mathbb{R}} f(x)e^{-i\xi x}dx\right| \leq \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\left|f(x)e^{-i\xi x}\right|dx = \frac{1}{\sqrt{2\pi}}\|f\|_1,$$

d'où l'inégalité (2.2).

Du fait de la linéarité de l'intégrale de Lebesgue, on obtient la linéarité de $\mathscr{F}$, et donc l'inégalité (2.2) montre la continuité de $\mathscr{F}$.

Il reste donc à montrer que $\widehat{f}(\xi) \xrightarrow[\xi\to\pm\infty]{} 0$ qui est une variante du lemme de Riemann-Lebesgue et dont la preuve est similaire, voir le lemme 1.8. $\square$

39. Rappelons-nous donc que si $(f_n)_{n\in\mathbb{N}}$ est une suite de fonctions de $L^1(\mathbb{R})$ qui converge au sens $L^1(\mathbb{R})$ vers f, alors $(\widehat{f_n})_{n\in\mathbb{N}}$ converge uniformément vers $\widehat{f}$.

40. On respecte aussi l'usage habituel en probabilités qui est de noter t la variable de Fourier, malgré le conseil de la note 36.

41. Cette ligne pourrait être retirée ; en vérité, pour être rigoureux elle aurait dû être précisée plus tôt, quand on a montré que $\widehat{f}$ était bien définie.

2.1.2 Théorème d'inversion de Fourier

La première question naturelle est de savoir s'il est possible de reconstruire f à partir de la connaissance de $\widehat{f}$ [42], et si cela est possible, de savoir si on a un moyen "explicite" de le faire. Le résultat suivant fournit une réponse aux deux questions :

Théorème 2.5 (Théorème d'inversion de Fourier)**.** *Soit $f \in L^1(\mathbb{R})$ telle que $\widehat{f} \in L^1(\mathbb{R})$ [43], c'est-à-dire*

$$\int_{\mathbb{R}} |\widehat{f}(\xi)| d\xi < +\infty.$$

Alors on a

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \widehat{f}(\xi) e^{+i\xi x} d\xi = \widehat{\widehat{f}}(-x) \qquad \textit{pour presque tout } x \in \mathbb{R}. \tag{2.3}$$

Autrement dit, on a

$$\forall f \in L^1(\mathbb{R}) \textit{ tel que } \mathscr{F}(f) \in L^1(\mathbb{R}), \quad \mathscr{F} \circ \mathscr{F}(f) = \sigma(f)$$

où $\sigma(f)(x) = f(-x)$ pour $x \in \mathbb{R}$ et $f \in L^1(\mathbb{R})$.

On peut aussi définir

$$\begin{array}{rccc} \mathscr{G} : & L^1(\mathbb{R}) & \longrightarrow & C^0_{\to 0}(\mathbb{R}) \\ & g & \longmapsto & \left[x \in \mathbb{R} \mapsto \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(\xi) e^{+i\xi x} d\xi \right] \end{array}$$

qu'on appelle transformation de Fourier inverse, et alors

$$\forall f \in L^1(\mathbb{R}) \textit{ tel que } \mathscr{F}(f) \in L^1(\mathbb{R}), \quad \mathscr{G} \circ \mathscr{F}(f) = f, \tag{2.4}$$

$$\forall f \in L^1(\mathbb{R}) \textit{ tel que } \mathscr{G}(f) \in L^1(\mathbb{R}), \quad \mathscr{F} \circ \mathscr{G}(f) = f. \tag{2.5}$$

Avant de montrer ce résultat, on donne un corollaire important :

Corollaire 2.6. *$\mathscr{F}$ est un opérateur injectif.*

Démonstration. En effet, si $f \in L^1(\mathbb{R})$ est tel que $\widehat{f} = 0$, alors $\widehat{f} \in L^1(\mathbb{R})$ donc on peut appliquer le théorème 2.5, et l'équation (2.3) montrer que f est nulle presque partout. □

Remarque 2.7. On aurait envie de penser à $\mathscr{G}$ comme à l'inverse de $\mathscr{F}$; d'ailleurs, on croise parfois la notation $\mathscr{G} = \mathscr{F}^{-1}$. Mais attention à cette notation :

1. remarquons d'abord que l'opérateur $\mathscr{F}$ n'est pas bijectif car il n'est pas surjectif, plus précisément son image n'est pas $C^0_{\to 0}(\mathbb{R})$ [44],
2. mais le problème est plus sérieux : on pourrait en effet "rendre" l'opérateur bijectif en remplaçant son image par $\mathrm{Im}(\mathscr{F})$, quand bien même on ne connait pas cet espace. Mais dans ce cas il y a un problème de compatibilité des espaces : en effet $\mathscr{F}$ et $\mathscr{G}$ sont définies sur $L^1(\mathbb{R})$, mais ce dernier n'est pas stable. On ne peut donc pas définir $\mathscr{F} \circ \mathscr{G}$ ou $\mathscr{G} \circ \mathscr{F}$ sans hypothèse sur les variables comme on l'a fait dans (2.4),(2.5).

Il faut donc retenir que $\mathscr{G}$ est un bon candidat formel à être l'inverse de $\mathscr{F}$, mais qu'on ne peut écrire $\mathscr{F} \circ \mathscr{G} = \mathscr{G} \circ \mathscr{F} = Id$ que quand la formule a effectivement un sens. Cela motivera le fait de choisir des espaces plus propices [45].

Remarquons enfin que $\mathscr{G}$ ressemble comme deux gouttes d'eau à $\mathscr{F}$. Toutes les propriétés valables sur $\mathscr{F}$ le seront aussi sur $\mathscr{G}$, au signe près. D'ailleurs, $\mathscr{F}$ est presque une involution, c'est-à-dire que son inverse est presque égal à $\mathscr{F}$. En fait, si on itère la formule, on a $\mathscr{F}^4(f) = f$ à condition de faire les hypothèses pour que cette formule ait un sens [46].

42. Ce qui se traduit par : $\mathscr{F}$ est-il injectif ?
43. En fait, on devrait plutôt écrire $\widehat{f} \in \mathcal{L}^1(\mathbb{R})$, puisqu'on a dit que $\widehat{f}$ était une "vraie" fonction ; en pratique on ne fait pas cette distinction.
44. Comme pour les séries de Fourier, ce résultat n'est pas simple. On peut trouver une preuve élémentaire dans [Laa07, Exo 8.14] qui consiste à affiner le fait que $\lim_{\pm\infty} \widehat{f} = 0$; une autre preuve utilise le théorème d'isomorphisme de Banach (voir Chapitre IX), voir par exemple [Wag12a, Exo 2.47.1].
45. Le lecteur aura compris qu'un espace est propice s'il est stable par $\mathscr{F}$.
46. Il faut donc a priori supposer $f, \mathscr{F}(f), \mathscr{F}^2(f), \mathscr{F}^3(f) \in L^1(\mathbb{R})$.

Remarque 2.8. Comparez avec le théorème 1.28 pour les séries de Fourier [47]. Des remarques similaires sont d'ailleurs à faire :

- l'hypothèse $\widehat{f} \in L^1(\mathbb{R})$ n'est pas toujours commode à vérifier, car $\widehat{f}$ n'est pas toujours explicite,
- le théorème 2.5 donne en fait une condition nécessaire pour que $\widehat{f} \in L^1(\mathbb{R})$, à savoir que f ait un représentant qui soit une fonction continue sur $\mathbb{R}$. En effet, si $\widehat{f} \in L^1(\mathbb{R})$, alors $x \mapsto \widehat{\widehat{f}}(x)$ est une fonction continue car au signe près c'est la transformée de Fourier d'une fonction de $L^1(\mathbb{R})$ (la continuité découle donc du théorème de continuité des intégrales à paramètre, comme on l'a justifié dans la preuve de la définition/proposition 2.1). Or f est égale presque partout à cette fonction continue. Ainsi, par contraposée, on conclut que si f n'a pas de représentant continu [48], alors $\widehat{f} \notin L^1(\mathbb{R})$.
- on verra une condition suffisante en conséquence de la proposition 2.15 :
$$\text{Si } f \in C^2(\mathbb{R}) \text{ est telle que } (f, f', f'') \in L^1(\mathbb{R}), \quad \text{alors} \quad \widehat{f}(\xi) \underset{\xi\to\pm\infty}{=} o\left(\frac{1}{\xi^2}\right)$$
et donc $\widehat{f} \in L^1(\mathbb{R})$ [49]

Pour la démonstration, on aura besoin de connaître explicitement la transformée de Fourier d'une fonction classique ; nous choisissons la Gaussienne :

Exercice 2.9 (Transformée de Fourier de la Gaussienne)**.** On pose

$$\forall x \in \mathbb{R}, \quad \varphi(x) = e^{-\frac{x^2}{2}}.$$

L'objectif de l'exercice est de montrer $\widehat{\varphi} = \varphi$ en justifiant le calcul formel [50] :

$$\begin{aligned}
\forall \xi \in \mathbb{R}, \ \widehat{\varphi}(\xi) &= \int_{\mathbb{R}} e^{-\frac{x^2}{2}} e^{-i\xi x} \frac{dx}{\sqrt{2\pi}} = \int_{\mathbb{R}} e^{-\frac{(x+i\xi)^2}{2}} e^{-\frac{\xi^2}{2}} \frac{dx}{\sqrt{2\pi}} \\
&= \int_{\mathbb{R}} e^{-\frac{x^2}{2}} \frac{dx}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} = e^{-\frac{\xi^2}{2}}.
\end{aligned}$$

1. Montrer qu'on peut définir
$$\forall z \in \mathbb{C}, \ \ F(z) = \int_{\mathbb{R}} e^{-\frac{x^2}{2}} e^{-zx} \frac{dx}{\sqrt{2\pi}}.$$
2. Montrer que F est holomorphe sur $\mathbb{C}$.
3. Montrer que
$$\forall z \in \mathbb{R}, \ \ F(z) = e^{\frac{z^2}{2}}.$$
4. En déduire $\widehat{\varphi} = \varphi$.

Démonstration du théorème 2.5 :
Idée : on aimerait partir du terme compliqué dans (2.3) et appliquer le théorème de Fubini :

$$\begin{aligned}
\int_{\mathbb{R}} \widehat{f}(\xi) e^{i\xi x} \frac{d\xi}{\sqrt{2\pi}} &= \int_{\mathbb{R}} \int_{\mathbb{R}} f(y) e^{-i\xi y} e^{+i\xi x} \frac{dy}{\sqrt{2\pi}} \frac{d\xi}{\sqrt{2\pi}} \\
&= \int_{\mathbb{R}} f(y) \int_{\mathbb{R}} e^{-i\xi(y-x)} \frac{d\xi}{\sqrt{2\pi}} \frac{dy}{\sqrt{2\pi}}
\end{aligned}$$

47. Les hypothèses sont semblables (voir la remarque 1.30 à propos de l'hypothèse de continuité du théorème 1.28.), et la conclusion peut aussi être considérée comparable, si on pense à $\sum_{n\in\mathbb{Z}} \widehat{f}(n) e^{inx}$ comme à une intégrable en la variable $n \in \mathbb{Z}$.
48. C'est le cas par exemple des fonctions $\mathbb{1}_{[-a,a]}$, dont on peut calculer la transformée de Fourier explicitement : on conseille le lecteur de constater que $\widehat{\mathbb{1}_{[-a,a]}}$ n'est effectivement pas dans $L^1(\mathbb{R})$, voir l'exercice 2.10.
49. Précisons que $\widehat{f}$ étant une fonction continue, la question de l'intégrabilité de $\widehat{f}$ ne se pose qu'aux voisinages de $\pm\infty$.
50. Ce calcul n'est pas correct car on fait un changement de variable sur $\mathbb{C}$ et donc on devrait écrire $\mathbb{R} - i\xi$ dans la dernière intégrale.

mais cela n'est pas possible car les hypothèses du théorème de Fubini ne s'appliquent pas, et pour cause on voit que l'écriture obtenue en fin n'a pas de sens [51].

On utilise donc la fonction $\varphi : x \in \mathbb{R} \mapsto e^{-\frac{x^2}{2}}$, et on glisse $\varphi(\varepsilon\xi)$ dans le calcul précédent, qui approche 1 de façon adéquate. Plus précisémment, on pose

$$\forall\varepsilon > 0,\ \forall x \in \mathbb{R},\ I_\varepsilon(f)(x) = \int_{\mathbb{R}} \widehat{f}(\xi)e^{i\xi x}\varphi(\varepsilon\xi)\frac{d\xi}{\sqrt{2\pi}}.$$

Le théorème de Fubini s'applique cette fois-ci car

$$\int_{\mathbb{R}}\int_{\mathbb{R}} |f(y)\varphi(\varepsilon\xi)|dyd\xi = \|f\|_1 \int_{\mathbb{R}} \varphi(\varepsilon\xi)d\xi < +\infty.$$

Ainsi en reprenant le calcul précédent on obtient

$$\begin{aligned}\forall\varepsilon > 0, \forall x \in \mathbb{R},\ \ I_\varepsilon(f)(x) &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(y)e^{-i\xi y}e^{+i\xi x}\varphi(\varepsilon\xi)\frac{dy}{\sqrt{2\pi}}\frac{d\xi}{\sqrt{2\pi}} \\ &= \int_{\mathbb{R}} f(y)\int_{\mathbb{R}} e^{-i\xi(y-x)}\varphi(\varepsilon\xi)\frac{d\xi}{\sqrt{2\pi}}\frac{dy}{\sqrt{2\pi}} \\ &= \int_{\mathbb{R}} f(y)\underbrace{\int_{\mathbb{R}} e^{-i\xi\frac{(y-x)}{\varepsilon}}\varphi(\xi)\frac{d\xi}{\varepsilon\sqrt{2\pi}}}_{=\frac{1}{\varepsilon}\widehat{\varphi}\left(\frac{y-x}{\varepsilon}\right)}\frac{dy}{\sqrt{2\pi}}\end{aligned}$$

On est amené à poser

$$\forall\varepsilon > 0,\ \forall x \in \mathbb{R},\ \ \rho_\varepsilon(x) = \frac{1}{\varepsilon\sqrt{2\pi}}\widehat{\varphi}\left(\frac{x}{\varepsilon}\right) = \frac{1}{\varepsilon\sqrt{2\pi}}\varphi\left(\frac{x}{\varepsilon}\right).$$

Comme la fonction $\frac{\varphi}{\sqrt{2\pi}}$ est positive et d'intégrale 1, $(\rho_\varepsilon)_{\varepsilon>0}$ est une identité approchée, voir l'exemple 5.50 au Chapitre IV. Au final on a montré

$$\forall\varepsilon > 0,\ \ \forall x \in \mathbb{R},\ \ \ I_\varepsilon(f)(x) = f * \rho_\varepsilon(x).$$

Par le lemme 5.56, on a d'une part

$$f * \rho_\varepsilon \xrightarrow[\varepsilon\to 0]{L^1} f,$$

et donc la convergence a aussi lieu presque partout pour une certaine suite $\varepsilon_n \to 0$, d'après le corollaire 5.24 du Chapitre IV. D'autre part, le théorème de convergence dominée donne pour un $x \in \mathbb{R}$ fixé

$$I_\varepsilon(f)(x) \xrightarrow[\varepsilon\to 0]{} \int_{\mathbb{R}} \widehat{f}(\xi)e^{i\xi x}\frac{d\xi}{\sqrt{2\pi}}$$

car on a la convergence simple de $\varphi(\varepsilon x)$ vers $\varphi(0) = 1$, et la domination

$$\forall\varepsilon > 0, \forall\xi \in \mathbb{R},\ \ \left|\widehat{f}(\xi)e^{i\xi x}\varphi(\varepsilon\xi)\right| \leq \left|\widehat{f}(\xi)\right|$$

indépendante de ε et telle que $\widehat{f} \in L^1(\mathbb{R})$ (par hypothèse).

Par unicité de la limite, on conclut que

$$\int_{\mathbb{R}} \widehat{f}(\xi)e^{i\xi x}\frac{d\xi}{\sqrt{2\pi}} = f(x)$$

pour presque tout $x \in \mathbb{R}$. □

51. L'intégrale $\int_{\mathbb{R}} e^{i\omega x}dx$ n'a pas de sens (sauf éventuellement si $\omega = 0$ car alors la fonction est réelle positive, et on peut dire que l'intégrale vaut $+\infty$).

2.1.3 Exemples

Exercice 2.10. On pose pour $a \in]0,+\infty[$, $f_a = \mathbb{1}_{[-a,a]}$.

1. Calculer $\widehat{f_a}$.
2. Justifier $\widehat{f_a} \notin L^1(\mathbb{R})$ par le calcul d'une part, et en utilisant le théorème 2.5 d'autre part (voir la remarque 2.8).

Exercice 2.11. 1. On pose pour $a \in]0,+\infty[$,

$$\forall x \in \mathbb{R}, \quad f_a(x) = e^{-a|x|}.$$

Calculer $\widehat{f_a}$.

2. On pose pour $a \in]0,+\infty[$,

$$\forall x \in \mathbb{R}, \quad g_a(x) = \frac{1}{x^2+a^2}.$$

   (a) Calculer $\widehat{g_a}$ par la formule des résidus.
   (b) Calculer $\widehat{g_a}$ à partir du théorème 2.5.

2.1.4 Opérations

Comme pour les coefficients de Fourier, on souhaite savoir comment se comporte $\mathscr{F}$ vis-à-vis des opérations usuelles sur les fonctions :

Proposition 2.12. *1. Soit $f \in L^1(\mathbb{R})$ et $\alpha \in \mathbb{R}$. On pose $\tau_\alpha f : x \mapsto f(x-\alpha)$. Alors*

$$\forall \xi \in \mathbb{R}, \quad \widehat{\tau_\alpha f}(\xi) = e^{-i\alpha\xi}\widehat{f}(\xi).$$

En posant $e_\alpha : x \mapsto e^{i\alpha x}$, on a aussi

$$\forall \xi \in \mathbb{R}, \quad \widehat{e_\alpha f}(\xi) = \tau_\alpha \widehat{f}(\xi).$$

2. Soit $f \in L^1(\mathbb{R})$ et $\lambda \in]0,+\infty[$. On pose $\mu_\lambda f : x \mapsto f(\lambda x)$. Alors

$$\forall \xi \in \mathbb{R}, \quad \widehat{\mu_\lambda f}(\xi) = \frac{1}{\lambda}\widehat{f}\left(\frac{\xi}{\lambda}\right) = \frac{1}{\lambda}\mu_{1/\lambda}\widehat{f}(\xi),$$

et

$$\forall \xi \in \mathbb{R}, \quad \mu_\lambda \widehat{f}(\xi) = \frac{1}{\lambda}\widehat{\mu_{1/\lambda} f}(\xi).$$

3. Soit $(f,g) \in L^1(\mathbb{R})$. Alors $f * g \in L^1(\mathbb{R})$ et

$$\widehat{f * g} = \sqrt{2\pi}\widehat{f}.\widehat{g}.$$

Remarque 2.13. En général on connaît l'existence des deux premiers points, mais pas les formules exactes qui se retrouvent très facilement par changement de variable. En notant E_α l'opérateur de $L^1(\mathbb{R})$ dans lui-même qui envoie f sur le produit $e_\alpha f$, on peut en fait écrire :

$$\mathscr{F} \circ \tau_\alpha = E_{-\alpha} \circ \mathscr{F}, \qquad \tau_\alpha \circ \mathscr{F} = \mathscr{F} \circ E_\alpha,$$

$$\mathscr{F} \circ \mu_\lambda = \frac{1}{\lambda}\mu_{1/\lambda} \circ \mathscr{F}, \qquad \mu_\lambda \circ \mathscr{F} = \frac{1}{\lambda}\mathscr{F} \circ \mu_{1/\lambda}.$$

Démonstration. La preuve est élémentaire. Le dernier point utilise le théorème de Fubini. □

Exercice 2.14. 1. Exprimer $\overline{\widehat{f}}$ en fonction de $\widehat{\overline{f}}$.

2. Montrer que si f est paire, alors $\widehat{f}$ également.
3. Montrer que si f est réelle et paire, alors $\widehat{f}$ également.

On termine par la façon dont $\mathscr{F}$ se comporte vis-à-vis de la dérivation. Si on pose D l'opérateur de dérivation (disons défini de $C^1(\mathbb{R})$ dans $C^0(\mathbb{R})$), on est intéressé par

$$\mathscr{F} \circ D \text{ et } D \circ \mathscr{F}$$

(attention, cette composition n'a pas de sens si on ne fait pas attention aux espaces choisis ; on fera attention dans l'énoncé suivant à ce que toutes les transformées de Fourier qui apparaissent ont un sens).

Proposition 2.15. *1. Soit $f \in C^1(\mathbb{R})$ telle que $f \in L^1(\mathbb{R})$ et $f' \in L^1(\mathbb{R})$. Alors*

$$\forall \xi \in \mathbb{R}, \quad \widehat{f'}(\xi) = i\xi \widehat{f}(\xi)$$

et en particulier

$$\forall \xi \in \mathbb{R}^*, \quad |\widehat{f}(\xi)| \leq \frac{\|f\|_1}{\sqrt{2\pi}|\xi|}, \qquad \text{et} \quad \widehat{f}(\xi) = \underset{\pm\infty}{o}\left(\frac{1}{|\xi|}\right).$$

2. On pose $m(x) = x$. Soit $f \in L^1(\mathbb{R})$ telle que $mf \in L^1(\mathbb{R})$. Alors $\widehat{f}$ est de classe C^1 sur $\mathbb{R}$ et[52]

$$\forall \xi \in \mathbb{R}, \quad (\widehat{f})'(\xi) = -i\widehat{mf}(\xi).$$

Démonstration. 1. On voudrait justifier l'intégration par parties suivante :

$$\begin{aligned} \forall \xi \in \mathbb{R}, \quad \widehat{f'}(\xi) &= \int_{\mathbb{R}} f'(x) e^{-i\xi x} \frac{dx}{\sqrt{2\pi}} \\ &= \left[f(x) \frac{e^{-i\xi x}}{\sqrt{2\pi}} \right]_{-\infty}^{+\infty} - \int_{\mathbb{R}} f(x)(-i\xi) e^{-i\xi x} \frac{dx}{\sqrt{2\pi}} \end{aligned}$$

mais pour cela il faut justifier qu'on a

$$f(x) \xrightarrow[x \to \pm\infty]{} 0$$

(rappelons que les hypothèses $f \in C^1(\mathbb{R}) \cap L^1(\mathbb{R})$ ne suffisent pas pour avoir cela). Le raisonnement est le suivant :

— comme f' est intégrable et $f(x) = f(0) + \int_0^x f'(t)dt$, la fonction f a bien des limites en $\pm\infty$,

— ces limites ne peuvent être que 0 puisque f est intégrable[53].

2. Il s'agit d'appliquer le théorème de dérivation des intégrales à paramètre (théorème 3.21 au Chapitre IV).

□

Remarque 2.16. On peut donner des versions itérées : soit $p \in \mathbb{N}^*$.

1. Si $f \in C^p(\mathbb{R})$, et $(f, f', \ldots, f^{(p)}) \in L^1(\mathbb{R})$, alors

$$\forall k \in [\![0, p]\!], \ \forall \xi \in \mathbb{R}, \quad \widehat{f^{(k)}}(\xi) = (i\xi)^k \widehat{f}(\xi), \qquad \text{et} \qquad \widehat{f}(\xi) = \underset{\pm\infty}{o}\left(\frac{1}{|\xi|^k}\right).$$

2. Si $f \in L^1(\mathbb{R})$ et $m^p f \in L^1(\mathbb{R})$, alors on voit facilement que $m^k f \in L^1(\mathbb{R})$ pour tout $k \in [\![0, p]\!]$, et donc $\widehat{f}$ est de classe C^k et

$$\forall k \in [\![0, p]\!], \forall \xi \in \mathbb{R}, \quad \widehat{f}^{(k)}(\xi) = (-i)^k \widehat{m^k f}(\xi).$$

On peut retenir les deux principes suivants :

52. On lit souvent $\widehat{xf}$ à la place de $\widehat{mf}$, mais cette version n'est pas rigoureuse, et peut amener des confusions. On incite donc à la prudence en utilisant des notations rigoureuses.

53. Vous devez savoir détailler ce point si le jury le demande.

« la régularité de f se traduit par un comportement en l'infini de $\widehat{f}$ »
« le comportement en l'infini de f se traduit par de la régularité pour $\widehat{f}$ ».

Le premier principe est similaire à ce qu'on a pu voir pour les séries de Fourier, le second est nouveau car ça n'avait pas vraiment de sens de se poser la question de la régularité de $\widehat{f}$ en la variable $n \in \mathbb{Z}$. À partir du théorème d'inversion 2.5, qui permet de voir f comme Fourier de lui-même (au signe près et si cela a un sens), les réciproques des deux principes précédents sont possibles, mais comme pour les séries de Fourier, on récupère des réciproques partielles. On a par exemple

$$\widehat{f}(\xi) = \underset{\pm\infty}{o}\left(\frac{1}{|\xi|^2}\right) \quad \Rightarrow \quad \widehat{f} \in L^1(\mathbb{R}) \quad \Rightarrow \quad f \in C^0(\mathbb{R}),$$

(quand on écrit $f \in C^0(\mathbb{R})$, il faut comprendre "f admet un représentant continu"), et sa version itérée :

$$\widehat{f}(\xi) = \underset{\pm\infty}{o}\left(\frac{1}{|\xi|^{k+2}}\right) \quad \Rightarrow \quad f \in C^k(\mathbb{R}).$$

Remarque 2.17. Si on note M l'opérateur qui a une fonction f associe la fonction mf, alors on a formellement

$$\mathscr{F} \circ D = iM \circ \mathscr{F}, \qquad D \circ \mathscr{F} = -i\mathscr{F} \circ M.$$

Exercice 2.18. Voici un autre résultat classique sur le lien entre le comportement en l'infini d'une fonction et la régularité de sa transformée de Fourier : montrer que si $f \in L^1(\mathbb{R})$ est à support compact sur $\mathbb{R}$, alors $\widehat{f}$ est analytique sur $\mathbb{R}$.

2.1.5 Quelques exemples d'utilisation de la transformation de Fourier

Idée générale :

on part d'un problème (P) (disons une équation fonctionnelle pour fixer les idées) dont l'inconnue est une fonction $f : \mathbb{R} \to \mathbb{C}$, que l'on cherche dans un espace fonctionnel qu'on note X.

On fait opérer la tranformation de Fourier sur le problème (P), et on obtient un autre problème $(\widehat{P})$ dont l'inconnue est $\widehat{f} \in \widehat{X}$ où $\widehat{X} = \{\widehat{g},\ g \in X\}$. On utilise à cette étape les règles du paragraphe 2.1.4 pour voir le comportement de $\mathscr{F}$ vis-à-vis des opérations qui interviennent dans le problème (P).

On espère bien sûr que ce nouveau problème est plus simple à résoudre que le précédent, et on le résout quand cela est possible.

Comme pour l'utilisation des séries de Fourier, une difficulté non négligeable est que l'espace $\widehat{X}$ est difficile à identifier en général. Cela n'empêche pas de considérer le problème $(\widehat{P})$, ni de le résoudre. Mais on devra se demander si les solutions trouvées correspondent bien à des solutions du problème (P). En général, on pourra résoudre le problème $(\widehat{P})$ dans un espace Y dont on sait qu'il est plus gros que $\widehat{X}$: alors il se pourrait qu'une solution trouvée de $(\widehat{P})$ ne soit pas dans $\widehat{X}$.

Du fait de ces difficultés, on raisonne rarement par équivalence dans l'utilisation de la transformation de Fourier.

Aussi, pour le moment on suppose que $X \subset L^1(\mathbb{R})$ puisqu'il s'agit du cadre dans lequel on a défini la transformation de Fourier. On pourra revenir sur cette hypothèse quand on aura étudié d'autres cadres.

Exercice 2.19 (Non existence d'un élément neutre pour la convolution dans $L^1(\mathbb{R})$)**.** Rappelons que la convolution est une loi interne à $L^1(\mathbb{R})$ (voir la proposition 5.30 au Chapitre IV). On se pose la question de l'existence d'un élément neutre pour cette loi. Le problème à résoudre est donc de trouver e tel que

$$e \in L^1(\mathbb{R}), \qquad \text{et} \qquad \forall f \in L^1(\mathbb{R}),\ f * e = f.$$

1. Ecrire la version Fourier du problème.
2. Résoudre la version Fourier du problème dans l'espace $Y = C^0_b(\mathbb{R})$ [54].

54. On choisit volontairement un espace trop gros ; l'objectif est de réaliser que l'ensemble des solutions au problème $(\widehat{P})$ change en fonction de Y, et que le lecteur puisse imaginer des situations où les solutions de $(\widehat{P})$ ne correspondent pas toutes à des solutions de (P) dans l'espace X. On aurait aussi pu prendre directement $Y = C^0_{\to 0}(\mathbb{R})$, auquel cas $(\widehat{P})$ n'a aucune solution.

3. Montrer qu'il n'existe pas d'élément neutre pour la convolution dans $L^1(\mathbb{R})$.

Exemple 2.20. Présentons formellement l'exemple célèbre de la résolution de l'équation

$$-y'' + y = g,$$

où g est une donnée, et y est l'inconnue. Pour que l'équation ait un sens, on demande a priori que g soit continue, et on cherche y de classe C^2 sur $\mathbb{R}$. Afin d'utiliser la transformation de Fourier, on demande aussi $g \in L^1(\mathbb{R})$, et on demande $y \in L^1(\mathbb{R})$, ce qui peut paraître peu habituel ; c'est le prix à payer pour utiliser l'approche de Fourier pour le moment [55]. Formellement (on précise des énoncés plus correct ci-dessou), on a la suite d'équivalences :

$$\begin{aligned}
-y'' + y = g \quad &\Leftrightarrow \quad \widehat{-y'' + y} = \widehat{g} \Leftrightarrow \forall \xi \in \mathbb{R}, -(i\xi)^2 \widehat{y}(\xi) + \widehat{y}(\xi) = \widehat{g}(\xi) \\
&\Leftrightarrow \quad \forall \xi \in \mathbb{R}, (\xi^2 + 1)\widehat{y} = \widehat{g} \Leftrightarrow \forall \xi \in \mathbb{R}, \widehat{y}(\xi) = \frac{\widehat{g}(\xi)}{1 + \xi^2} \\
&\Leftrightarrow \quad y = \mathscr{G}\left(\frac{\widehat{g}}{1 + m^2}\right)
\end{aligned}$$

où on a utilisé la notation $m : \xi \in \mathbb{R} \mapsto \xi$, et $\mathscr{G}$ est la transformation de Fourier inverse. Remarquons aussi qu'avec l'exercice 2.11, on a

$$\forall \xi \in \mathbb{R}, \quad \frac{1}{1 + \xi^2} = \sqrt{\frac{\pi}{2}} \widehat{h}(\xi), \qquad \text{où } h : x \in \mathbb{R} \mapsto e^{-|x|}$$

et donc on peut reprendre les équivalences précédentes, en utilisant la proposition 2.12

$$\begin{aligned}
-y'' + y = g \quad &\Leftrightarrow \quad \forall \xi \in \mathbb{R}, \widehat{y}(\xi) = \sqrt{\frac{\pi}{2}} \frac{1}{\sqrt{2\pi}} \widehat{f * h} \\
&\Leftrightarrow \quad \forall x \in \mathbb{R}, \; y(x) = \frac{1}{2} g * h(x) = \frac{1}{2} \int_{\mathbb{R}} f(x - t) e^{-|t|} dt.
\end{aligned}$$

Pour que la démonstration précédente soit correcte et non seulement formelle, il faut

— que toutes les transformées de Fourier qui interviennent soient bien définies,

— que la solution trouvée soit effectivement de classe C^2.

On peut par exemple supposer $g \in C^2(\mathbb{R})$ telle que $(g, g', g'') \in L^1(\mathbb{R})$ [56]. Alors le calcul précédent montre qu'il existe une unique solution $y \in C^2(\mathbb{R})$ telle que $(y, y', y'') \in L^1(\mathbb{R})$.

Exercice 2.21. Trouver l'ensemble des fonctions $f \in L^1(\mathbb{R})$ telles que

$$f * f = f.$$

Exercice 2.22. Soit $c \in \mathbb{R}$. Trouver l'ensemble des fonctions $f \in C^1(\mathbb{R}) \cap L^1(\mathbb{R})$ telles que

$$\forall x \in \mathbb{R}, \quad f'(x) = f(x + c).$$

Exercice 2.23. Trouver l'ensemble des fonctions de $L^1(\mathbb{R})$ telles que

$$f(2x) + f(2x - 1) = f(x), \qquad \text{pour presque tout } x \in \mathbb{R}.$$

55. L'intérêt des distributions tempérées étant entre autres de diminuer drastiquement les hypothèses à faire pour avoir le droit de faire agir la transformation de Fourier.

56. Ces hypothèses ne sont certainement pas optimales. Elles impliquent que $g * h$ est de classe C^2, et que $(g * h)' = g' * h$, $(g * h)'' = g'' * h$ de sorte que ces fonctions soient bien dans $L^1(\mathbb{R})$.

2.1.6 Généralisation à $\mathbb{R}^N$

On peut généraliser ce qui précède à des fonctions $f : \mathbb{R}^N \to \mathbb{C}$ avec $N \in \mathbb{N}^*$, en considérant :

$$\forall \xi \in \mathbb{R}^N, \quad \widehat{f}(\xi) := \frac{1}{(2\pi)^{N/2}} \int_{\mathbb{R}^N} f(x) e^{-i\xi\cdot x} dx$$

où $\xi \cdot x = \sum_{k=0}^{N} \xi_k x_k$ est le produit scalaire euclidien sur $\mathbb{R}^N$.

Exercice 2.24 (Généralisation de la proposition 2.15 en dimension supérieure)**.** Soit $f \in L^1(\mathbb{R}^N)$.

1. (a) On suppose $f \in C^1(\mathbb{R}^N)$, $k \in [\![1, N]\!]$ et $\partial_k f \in L^1(\mathbb{R})$. Montrer que

$$\forall \xi \in \mathbb{R}^N, \quad \widehat{\partial_k f}(\xi) = i\xi_k \widehat{f}(\xi).$$

   (b) On suppose $f \in C^2(\mathbb{R}^N)$. Calculer $\widehat{\Delta f}$.

2. (a) Soit $k \in [\![1, N]\!]$. On suppose $m_k f \in L^1(\mathbb{R}^N)$ où $m_k : x \mapsto x_k$. Montrer que $\widehat{f}$ est dérivable en la $k$ième variable et que

$$\forall \xi \in \mathbb{R}^N, \quad \partial_k \widehat{f}(\xi) = -i\widehat{m_k f}(\xi).$$

   (b) On suppose $m_k f \in L^1(\mathbb{R}^N)$ pour tout $k \in [\![1, N]\!]$. Montrer que $\widehat{f}$ est de classe C^1 sur $\mathbb{R}^N$.

Exercice 2.25 (Transformée de Fourier d'une Gaussienne en dimension supérieure)**.** Soit $A \in S_N^{++}(\mathbb{R})$. On pose

$$\forall x \in \mathbb{R}^N, \quad v_A(x) = e^{-\frac{\langle Ax, x\rangle}{2}}.$$

Montrer que [57]

$$\forall \xi \in \mathbb{R}^N, \quad \widehat{v_A}(\xi) = \frac{1}{\sqrt{\det(A)}} e^{-\frac{\langle A^{-1}x, x\rangle}{2}}.$$

2.2 Cadre $L^2(\mathbb{R})$

Commençons par un calcul élémentaire via la formule d'inversion :

Lemme 2.26. *Soit $f \in L^1(\mathbb{R})$ telle que $\widehat{f} \in L^1(\mathbb{R})$. Alors*

$$\|\widehat{f}\|_2 = \|f\|_2. \tag{2.6}$$

Remarque 2.27. Les hypothèses ici sont faites pour des raisons techniques ; nous allons voir que l'égalité (2.6) est valable dès que $\widehat{f}$ est définie.

Remarquons également que ces hypothèses impliquent $(f, \widehat{f}) \in L^2(\mathbb{R})$: en effet $\widehat{f} \in C^0_{\to 0}(\mathbb{R})$, donc il existe $R_0 \in \mathbb{R}$ tel que

$$\forall \xi \in \mathbb{R} \text{ tel que } |\xi| \geq R_0, \quad |\widehat{f}(\xi)| \leq 1, \qquad \text{et donc } |\widehat{f}(x)|^2 \leq |\widehat{f}(x)|$$

et comme $\widehat{f} \in L^1(\mathbb{R})$, on obtient finalement $\widehat{f} \in L^2(\mathbb{R})$ [58]. La même preuve fonctionne pour f et donne $f \in L^2(\mathbb{R})$; on peut aussi bien sûr appliquer (2.6) qui montre que $f \in L^2(\mathbb{R})$ puisque $\widehat{f} \in L^2(\mathbb{R})$.

Démonstration. En utilisant le théorème de Fubini et le théorème 2.5, on obtient

$$\begin{aligned}
\int_{\mathbb{R}} |\widehat{f}(\xi)|^2 d\xi &= \int_{\mathbb{R}} \widehat{f}(\xi)\overline{\widehat{f}(\xi)} d\xi \\
&= \int_{\mathbb{R}} \widehat{f}(\xi) \int_{\mathbb{R}} \overline{f(x)} e^{+i\xi x} \frac{dx}{\sqrt{2\pi}} d\xi \\
&= \int_{\mathbb{R}} \overline{f(x)} \int_{\mathbb{R}} \widehat{f}(\xi) e^{+i\xi x} \frac{d\xi}{\sqrt{2\pi}} dx \\
&= \int_{\mathbb{R}} \overline{f(x)} f(x) dx = \int_{\mathbb{R}} |f(x)|^2 dx
\end{aligned}$$

57. On pourra diagonaliser A et utiliser le résultat de l'exercice 2.9.

58. En tant que fonction continue, elle est intégrable sur $[-R_0, R_0]$, et la majoration de $|\widehat{f}|^2$ par $|\widehat{f}|$ sur $\mathbb{R} \setminus [-R_0, R_0]$ donne l'intégrabilité sur $\mathbb{R} \setminus [-R_0, R_0]$.

Remarquons que l'on pouvait bien appliquer la formule d'inversion car $\widehat{f} \in L^1(\mathbb{R})$, de même qu'on pouvait appliquer le théorème de Fubini car

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |\widehat{f}(\xi)\overline{f(x)}e^{+i\xi x}|dxd\xi = \|f\|_1 \|\widehat{f}\|_1 < +\infty.$$

□

On constate en particulier que sous les hypothèses du théorème, $f \in L^2(\mathbb{R}) \Leftrightarrow \widehat{f} \in L^2(\mathbb{R})$. Autrement dit, $L^2(\mathbb{R})$ est un espace intéressant car il est stable par $\mathscr{F}$. Le problème étant, outre les hypothèses du lemme (qui ne sont pas optimales) que $L^2(\mathbb{R})$ n'est pas un sous-espace de $L^1(\mathbb{R})$, donc on ne peut pas considérer la restriction de $\mathscr{F}$ à $L^2(\mathbb{R})$ comme on l'a fait pour le théorème de Parseval 1.32 pour les séries de Fourier. Néanmoins on a le résultat suivant :

Théorème 2.28 (Théorème de Plancherel)**.** *Il existe un unique opérateur*

$$\widetilde{\mathscr{F}} : L^2(\mathbb{R}) \longrightarrow L^2(\mathbb{R})$$

tel que

— $\widetilde{\mathscr{F}}$ *est un prolongement de la restriction de* $\mathscr{F}$ *à* $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$*, c'est-à-dire*

$$\forall f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}), \quad \widetilde{\mathscr{F}}(f) = \mathscr{F}(f) = \widehat{f},$$

— $\widetilde{\mathscr{F}}$ *est une isométrie*[59]*, c'est-à-dire*

$$\forall (f,g) \in L^2(\mathbb{R}) \times L^2(\mathbb{R}), \quad \int_{\mathbb{R}} f(x)\overline{g(x)}dx = \int_{\mathbb{R}} \widetilde{\mathscr{F}}(f)(\xi)\overline{\widetilde{\mathscr{F}}(g)(\xi)}d\xi, \tag{2.7}$$

et en particulier avec $g = f$*,*

$$\forall f \in L^2(\mathbb{R}), \quad \left\|\widetilde{\mathscr{F}}(f)\right\|_2 = \|f\|_2,$$

— $\widetilde{\mathscr{F}}$ *est surjective*[60]*.*

De plus, on a

$$\widetilde{\mathscr{F}} \circ \widetilde{\mathscr{F}} = \sigma \qquad \text{sur } L^2(\mathbb{R})$$

où $\sigma(f)(x) = f(-x)$ *pour* $x \in \mathbb{R}$ *et* $f \in L^2(\mathbb{R})$*.*
Autrement dit, on peut définir $\widetilde{\mathscr{G}} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ *qui est linéaire, isométrique, égal à* $\mathscr{G}$ *sur* $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$*, et on a alors*

$$\widetilde{\mathscr{G}} \circ \widetilde{\mathscr{F}} = \widetilde{\mathscr{F}} \circ \widetilde{\mathscr{G}} = Id_{L^2(\mathbb{R})}.$$

Remarque 2.29. Dit autrement, l'opérateur $\mathscr{F}$ restreint à $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ se prolonge en une isométrie surjective de $L^2(\mathbb{R})$. En pratique, comme souvent pour des prolongements, on confondra $\mathscr{F}$ et $\widetilde{\mathscr{F}}$, et on pourra noter $\widehat{f}$ que f soit dans $L^1(\mathbb{R})$ ou dans $L^2(\mathbb{R})$. Dans ce cas on a

$$f \in L^2(\mathbb{R}) \Longleftrightarrow \widehat{f} \in L^2(\mathbb{R})$$

sans avoir besoin des conditions du lemme 2.26.

Mais attention, si $f \in L^2(\mathbb{R}) \setminus L^1(\mathbb{R})$, on peut écrire $\widehat{f}$, mais

$$\text{on } \textit{ne peut pas} \text{ écrire } \widehat{f}(\xi) = \int_{\mathbb{R}} f(x)e^{-i\xi x}\frac{dx}{\sqrt{2\pi}}$$

59. Et donc en particulier $\widetilde{\mathscr{F}}$ est continue.
60. Et donc bijective, puisque le fait que c'est une isométrie (linéaire) implique qu'elle est injective.

car cette intégrale n'a pas de sens. Si on veut faire des calculs sur $\widehat{f}$ lorsque $f \in L^2(\mathbb{R})$, on doit raisonner par densité, c'est-à-dire écrire les formules pour $(f_n)_{n\in\mathbb{N}}$ une suite d'éléments de $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ qui converge vers f en norme $L^2(\mathbb{R})$, puis voir si les formules passent à la limite. Eventuellement, on peut souhaiter avoir une suite (f_n) explicite, ce qui n'est pas difficile : on peut en effet considérer $(f_n = f\mathbb{1}_{[-n,n]})_{n\in\mathbb{N}}$, qui est bien telle que

$$\forall n \in \mathbb{N}, f_n \in L^2(\mathbb{R}) \text{ et } f_n \xrightarrow[n\to+\infty]{L^2(\mathbb{R})} f.$$

Notons d'ailleurs que cela implique qu'il existe une extraction φ telle que $f_{\varphi(n)} \xrightarrow{\quad} {}_{n\to+\infty} f$ presque partout, et donc qu'on peut écrire [61]

$$\widehat{f}(\xi) = \lim_{n\to+\infty} \int_{-\varphi(n)}^{\varphi(n)} f(x)e^{-i\xi x}\frac{dx}{\sqrt{2\pi}} \quad \text{pour presque tout } \xi \in \mathbb{R}.$$

Exercice 2.30 (Condition suffisante pour le théorème d'inversion)**.** De la même manière que pour le corollaire 1.36, montrez que si $f \in C^1(\mathbb{R}) \cap L^1(\mathbb{R})$ et $f' \in L^2(\mathbb{R})$, alors $\widehat{f} \in L^1(\mathbb{R})$.

Démonstration. — Posons $D = \{f \in L^1(\mathbb{R}), \widehat{f} \in L^1(\mathbb{R})\}$. C'est un sous-ensemble de $L^2(\mathbb{R})$ d'après la remarque 2.27. On a

$$\mathscr{F} : (D, \|\cdot\|_2) \longrightarrow (L^2(\mathbb{R}), \|\cdot\|_2)$$

qui est linéaire continue d'après le lemme 2.26, et D est dense dans $L^2(\mathbb{R})$ car [62] D contient $C_c^2(\mathbb{R})$ d'après la fin de la remarque 2.8, ce dernier étant dense par le théorème 5.45 au Chapitre IV. Comme $L^2(\mathbb{R})$ est complet, le corollaire 2.51 au Chapitre III montre qu'il existe une unique application $\widetilde{\mathscr{F}} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ qui est linéaire continue et égale à $\mathscr{F}$ sur D.

— L'opérateur $\widetilde{\mathscr{F}}$ est une isométrie de $L^2(\mathbb{R})$ car l'égalité

$$\forall f \in D, \quad \|f\|_{L^2(\mathbb{R})} = \|\widetilde{\mathscr{F}}(f)\|_{L^2(\mathbb{R})}$$

s'étend à $f \in L^2(\mathbb{R})$ par densité. Par formule de polarisation [63], on en déduit (2.7).

— On a $\mathscr{F} = \widetilde{\mathscr{F}}$ sur D par construction, mais on aimerait étendre cette égalité à $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Pour cela, constatons en reprenant la preuve du théorème 5.45 du Chapitre IV sur la densité de $C_c^\infty(\mathbb{R})$ dans $L^p(\mathbb{R})$ si $p \in [1, +\infty[$, que la construction d'une $f_n \in C_c^\infty(\mathbb{R})$ qui converge vers f en norme $L^p(\mathbb{R})$ ne dépend pas de la valeur de p ; ainsi si $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, il existe une suite $(f_n)_{n\in\mathbb{N}}$ de fonctions de D telle que

$$f_n \xrightarrow[n\to+\infty]{L^1(\mathbb{R})} f, \quad \text{et} \quad f_n \xrightarrow[n\to+\infty]{L^2(\mathbb{R})} f$$

et donc par les propriétés de continuité de $\mathscr{F}$ et $\widetilde{\mathscr{F}}$ respectivement, on a

$$\widetilde{f_n} \xrightarrow[n\to+\infty]{\|\cdot\|_\infty} \mathscr{F}(f), \quad \text{et} \quad \widetilde{f_n} \xrightarrow[n\to+\infty]{L^2(\mathbb{R})} \widetilde{\mathscr{F}}(f).$$

Comme la convergence L^2 implique la convergence presque partout à extraction près, on en déduit $\mathscr{F}(f) = \widetilde{\mathscr{F}}(f)$ presque partout.

— On peut appliquer les étapes précédentes à $\mathscr{G}$ sans différence, et ainsi on construit $\widetilde{\mathscr{G}}$ qui est une isométrie linéaire, et coïncide avec $\mathscr{F}$ sur $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. L'égalité

$$\forall f \in D, \ \widetilde{\mathscr{F}} \circ \widetilde{\mathscr{G}}(f) = \widetilde{\mathscr{G}} \circ \widetilde{\mathscr{F}}(f) = f$$

s'étend donc par densité à $L^2(\mathbb{R})$, ce qui montre que $\widetilde{\mathscr{G}}$ est l'inverse de $\widetilde{\mathscr{F}}$, et en particulier que $\widetilde{\mathscr{F}}$ est surjective.

□

Remarque 2.31. Maintenant qu'on a défini $\mathscr{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ [64], on peut reprendre toutes les

61. On a donc une version faible d'intégrale semi-convergente (version faible car on sait qu'il y a convergence seulement si les bornes sont $[-\varphi(n), \varphi(n)]$).

62. Il existe plusieurs justifications de la densité de D dans $L^2(\mathbb{R})$; certains disent $D \supset \mathscr{S}(\mathbb{R})$, mais alors il faut savoir montrer que $\mathscr{S}(\mathbb{R})$ est dense, ce qu'on fait en général en utilisant également le théorème 5.45 au Chapitre IV.

63. Les formules de polarisation permettent d'exprimer le produit scalaire à partir de la norme, dans un espace préhilbertien, voir le Chapitre IX.

64. Rappelons que par abus, on note $\mathscr{F}$ à la place de $\widetilde{\mathscr{F}}$.

propriétés du paragraphe 2.1 et se demander si elles sont encore valables pour $f \in L^2(\mathbb{R})$. Il y a des propriétés qui fonctionnent bien (voir l'exercice 2.32), mais certaines propriétés ne sont plus valable : à titre d'exemple, il n'est plus vrai que $\widehat{f} \in C^0_{\to 0}(\mathbb{R})$ puisque $\mathscr{F}$ étant bijective, toute fonction de $L^2(\mathbb{R})$ est l'image par $\mathscr{F}$ d'une fonction de $L^2(\mathbb{R})$. Donc si on choisit $g \in L^2(\mathbb{R}) \setminus C^0_{\to 0}(\mathbb{R})$, on sait qu'il existe $f \in L^2(\mathbb{R})$ tel que $g = \widehat{f}$, ce qui montre que $\widehat{f}$ n'est pas continu en général, si on a $f \in L^2(\mathbb{R}) \setminus L^1(\mathbb{R})$.

Exercice 2.32. *[Extension à $L^2(\mathbb{R})$]* Dans cet exercice, on généralise à $L^2(\mathbb{R})$ des propriétés connues pour des fonctions de $L^1(\mathbb{R})$.

1. Montrer que, si $f \in L^2(\mathbb{R})$ et $\alpha \in \mathbb{R}$, alors $\tau_\alpha f \in L^2(\mathbb{R})$ [65] et $\widehat{\tau_\alpha f}(\xi) = e^{-i\xi\alpha}\widehat{f}(\xi)$ pour presque tout $\xi \in \mathbb{R}$.
2. Montrer que pour tout $f \in L^1(\mathbb{R})$, $g \in L^2(\mathbb{R})$, la fonction $t \mapsto f(t)g(x-t)$ est intégrable pour presque tout $x \in \mathbb{R}$, que la fonction $f * g$ est bien définie presque partout, et que $f * g \in L^2(\mathbb{R})$ avec $\|f * g\|_2 \leq \|f\|_1 \|g\|_2$.
3. Montrer que pour tout $f \in L^1(\mathbb{R})$, $g \in L^2(\mathbb{R})$,
$$\widehat{f * g} = \sqrt{2\pi}\widehat{f}\,\widehat{g} \ \ p.p.$$
Cette formule a-t-elle un sens si $f, g \in L^2(\mathbb{R})$?

Exercice 2.33. Pour $a \in \mathbb{R}_+^*$, on pose $f_a = \mathbb{1}_{[-a,a]}$.

1. Justifier sans calcul que $\widehat{f_a}$, la transformée de Fourier de f_a, appartient à $L^2(\mathbb{R})$ mais pas à $L^1(\mathbb{R})$.
2. Calculer la valeur de
$$\int_{\mathbb{R}} \left(\frac{\sin(x)}{x}\right)^2 dx.$$

Remarque 2.34 (Espaces de Sobolev)**.** L'espace $L^2(\mathbb{R})$ est donc très commode du point de vue de la transformation de Fourier. Afin de résoudre les difficultés évoquées à la remarque 2.16 où on voit qu'on n'a pas de bonnes caractérisations via la transformation de Fourier des espaces $C^k(\mathbb{R})$ ou plus précisément des espaces $\{f \in C^k(\mathbb{R}), (f, f', \ldots, f^{(k)}) \in L^1(\mathbb{R})^k\}$, on peut construire des espaces de régularité adaptés au cadre $L^2(\mathbb{R})$: on pose pour $s \in \mathbb{R}_+$:

$$H^s(\mathbb{R}) = \left\{ f \in L^2(\mathbb{R}),\ \int_{\mathbb{R}} (1 + |\xi|^2)^s |\widehat{f}(\xi)|^2 d\xi < +\infty \right\} \tag{2.8}$$

que l'on peut munir de la norme

$$\forall f \in H^s(\mathbb{R}),\ \ \|f\|_{H^s(\mathbb{R})} := \sqrt{\int_{\mathbb{R}} (1 + |\xi|^2)^s |\widehat{f}(\xi)|^2 d\xi},$$

et qu'on appelle espace de Sobolev d'ordre s. On peut montrer :

— pour tout $s \in \mathbb{R}_+$, $(H^s(\mathbb{R}), \|\cdot\|_s)$ est un espace de Banach, et même un espace de Hilbert car sa norme est issue du produit scalaire :
$$\forall (f, g) \in H^s(\mathbb{R}),\ \ \langle f, g\rangle_{H^s(\mathbb{R})} := \int_{\mathbb{R}} (1 + |\xi|^2)^s \widehat{f}(\xi)\overline{\widehat{g}(\xi)} d\xi.$$

— $(H^s(\mathbb{R}))_{s \in \mathbb{R}_+}$ est une famille décroissante d'espaces : plus précisément, pour tout $0 \leq s \leq s'$,
$$L^2(\mathbb{R}) = H^0(\mathbb{R}) \supset H^s(\mathbb{R}) \supset H^{s'}(\mathbb{R}) \supset C_c^\infty(\mathbb{R}).$$

— Si $s > \frac{1}{2}$ et $f \in H^s(\mathbb{R})$, alors $\widehat{f} \in L^1(\mathbb{R})$, $f \in C^0_{\to 0}(\mathbb{R})$ [66] et
$$\|f\|_\infty \leq C_s \|f\|_{H^s(\mathbb{R})}$$
où C_s est une constante qui ne dépend pas de f.

65. Rappelons que $\tau_\alpha f : x \mapsto f(x - \alpha)$.
66. Un tel résultat où on voit que les fonctions de $H^s(\mathbb{R})$ ont des propriétés de régularité (ici la continuité) fait partie d'une famille de résultats appelés "théorèmes d'injection de Sobolev".

Dit autrement, $(H^s(\mathbb{R}))_{s\in\mathbb{R}_+}$ est une famille d'espaces de régularité d'ordre $s \in \mathbb{R}_+$ construits sur la propriété exacte que sa transformation de Fourier doit satisfaire. Observons la justification suivante : pour $s = 1$, on a si $f \in \{g \in C^1(\mathbb{R}), (f, f') \in L^1(\mathbb{R})\}$:

$$\|f\|^2_{H^1(\mathbb{R})} = \|(1+|\xi|^2)^{1/2}\widehat{f}\|^2_{L^2(\mathbb{R})} \simeq \|\widehat{f}\|^2_{L^2(\mathbb{R})} + \|\xi\widehat{f}\|^2_{L^2(\mathbb{R})} = \|f\|^2_{L^2(\mathbb{R})} + \|f'\|^2_{L^2(\mathbb{R})}$$

où $\simeq$ signifie que les normes sont équivalentes.

Remarquons enfin qu'on peut montrer

$$H^1(\mathbb{R}) = \{f \in L^2(\mathbb{R}),\ f' \in L^2(\mathbb{R})\}$$

où f' est compris au sens des distributions (et le lecteur devinera la formule ad hoc pour $H^s(\mathbb{R})$ lorsque $s \in \mathbb{N}$).

2.3 L'espace de Schwartz

Comme on l'a expliqué à la remarque 2.7, avoir un espace stable par $\mathscr{F}$ est une bonne nouvelle ; en approfondissant la proposition 2.15 et les commentaires de la remarque 2.16, on peut définir un espace qui allie régularité et comportement à l'infini :

Définition 2.35. Soit $\varphi : \mathbb{R} \to \mathbb{C}$. On dit que $\varphi \in \mathscr{S}(\mathbb{R})$ si $\varphi \in C^\infty(\mathbb{R})$ et

$$\forall k \in \mathbb{N},\ \ \forall p \in \mathbb{N},\quad x \mapsto x^k\varphi^{(p)}(x) \text{ est bornée.}$$

On pose alors

$$\forall (k,p) \in \mathbb{N}^2,\ \ \forall \varphi \in C^\infty(\mathbb{R}),\quad N_{k,p}(\varphi) = \sup_{x\in\mathbb{R}} |x|^k|\varphi^{(p)}(x)|,$$

et

$$\mathscr{S}(\mathbb{R}) = \{f \in C^\infty(\mathbb{R}),\ \forall (k,p) \in \mathbb{N}^2,\ N_{k,p}(\varphi) < +\infty\}.$$

Remarque 2.36 (*)**.** Les fonctions $(N_{k,p})_{(k,p)\in\mathbb{N}^2}$ sont des semi-normes[67] sur $\mathscr{S}(\mathbb{R})$, et donc en particulier $\mathscr{S}(\mathbb{R})$ est un espace vectoriel.

Comme au paragraphe 4.4 au Chapitre III, on peut construire une métrique d par

$$\forall \varphi_1, \varphi_2 \in \mathscr{S}(\mathbb{R}),\ \ d(\varphi_1, \varphi_2) = \sum_{(k,p)\in\mathbb{N}^2} \frac{1}{2^{k+p}} \min\left\{1, N_{k,p}(\varphi_1 - \varphi_2)\right\}$$

qui est telle que pour toute suite $(\varphi_n)_{n\in\mathbb{N}}$ d'éléments de $\mathscr{S}(\mathbb{R})$,

$$\varphi_n \xrightarrow[n\to+\infty]{d} \varphi \iff \forall (k,p) \in \mathbb{N}^2,\ N_{k,p}(\varphi_n - \varphi) \xrightarrow[n\to+\infty]{} 0.$$

Notons que $(\mathscr{S}(\mathbb{R}), d)$ est un espace métrique complet.

Regardons les premières propriétés indépendantes de la transformation de Fourier :

Proposition 2.37. — *On a les inclusions*

$$C_c^\infty(\mathbb{R}) \subset \mathscr{S}(\mathbb{R}) \subset L^p(\mathbb{R})$$

pour tout $p \in [1, \infty]$. En conséquence $\mathscr{S}(\mathbb{R})$ est dense dans $L^p(\mathbb{R})$ si $p \in [1, +\infty[$.

— *L'espace $\mathscr{S}(\mathbb{R})$ est stable par dérivation et par multiplication par un polynôme, c'est-à-dire*

$$\forall \varphi \in \mathscr{S}(\mathbb{R}), \forall P \in \mathbb{C}[X],\ \ \varphi' \in \mathscr{S}(\mathbb{R}) \quad \textit{et} \quad P\varphi \in \mathscr{S}(\mathbb{R}).$$

67. Voir la note 52 au Chapitre III pour une définition des semi-normes. Remarquons que si $p = 0$, $N_{k,0}$ est même une norme pour $k \in \mathbb{N}$; mais on ne munit pas $\mathscr{S}(\mathbb{R})$ d'une telle norme car elle ne rend pas l'espace complet.

Exemple 2.38. Les fonctions de $\mathscr{S}(\mathbb{R})$ doivent tendre vers 0 plus vite que tout polynôme[68] ainsi que toutes ses dérivées ; les fonctions régulières à support compact sont donc dans $\mathscr{S}(\mathbb{R})$. L'exemple typique d'une fonction de $\mathscr{S}(\mathbb{R})$ qui n'est pas à support compact est la fonction

$$x \mapsto e^{-x^2}.$$

Démonstration. — Si $\varphi \in C_c^\infty(\mathbb{R})$, alors pour tout $(k,p) \in \mathbb{N}^2$, la fonction $x \mapsto x^k \varphi^{(p)}(x)$ est continue à support compact, et est donc bornée ; ainsi $\varphi \in \mathscr{S}(\mathbb{R})$.

— Si $\varphi \in \mathscr{S}(\mathbb{R})$, alors $N_{0,0}(\varphi) < \infty$ ce qui montre que $\varphi \in L^\infty$. De plus $N_{2,0}(\varphi) < +\infty$ ce qui montre que $\varphi(x) = O_{\pm\infty}\left(\frac{1}{x^2}\right)$, et donc $\varphi \in L^p(\mathbb{R})$ pour tout $p \in [1, +\infty[$.

— On constate pour $\varphi \in \mathscr{S}(\mathbb{R})$ et $(k,p) \in \mathbb{N}^2$,

$$N_{k,p}(\varphi') = N_{k,p+1}(\varphi)$$

ce qui donne $\varphi' \in \mathscr{S}(\mathbb{R})$. De même, par la formule de Leibniz, en notant $m : x \in \mathbb{R} \mapsto x$,

$$N_{k,p}(m\varphi) = \|m^k(m\varphi)^{(p)}\|_\infty \leq \sum_{j=0}^{p} \left\| \binom{p}{j} m^{(j)} \varphi^{(p-j)} m^k \right\|_\infty = N_{k+1,p}(\varphi) + pN_{k,p-1}(\varphi)$$

qui montre que $m\varphi \in \mathscr{S}(\mathbb{R})$. Par récurrence on en déduit que $m^n\varphi \in \mathscr{S}(\mathbb{R})$ si $n \in \mathbb{N}$, et par linéarité $P\varphi \in \mathscr{S}(\mathbb{R})$ pour $P \in \mathbb{C}[X]$.

□

Venons-en à la propriété centrale de l'espace $\mathscr{S}(\mathbb{R})$, qui témoigne de son importance pour la transformation de Fourier :

Théorème 2.39. *L'espace $\mathscr{S}(\mathbb{R})$ est stable par $\mathscr{F}$.*

Démonstration. Soit $\varphi \in \mathscr{S}(\mathbb{R})$ Alors pour tout $q \in \mathbb{N}, m^q\varphi \in L^1(\mathbb{R})$ car $N_{q+2,0}(\varphi) < \infty$. Ainsi par la proposition 2.15, $\widehat{\varphi} \in C^\infty(\mathbb{R})$ et

$$\forall p \in \mathbb{N}, \quad \widehat{\varphi}^{(p)} = \widehat{(-im)^p\varphi}.$$

Maintenant, pour $q \in \mathbb{N}$, on a par la formule de Leibniz et la proposition 2.37 $(m^p\varphi)^{(q)} \in \mathscr{S}(\mathbb{R})$, donc ces fonctions sont dans $L^1(\mathbb{R})$, et par la proposition 2.15,

$$m^k\widehat{\varphi}^{(p)} = m^k\widehat{(-im)^p\varphi} = \frac{(-i)^p}{i^k}\widehat{(m^p\varphi)^{(k)}}, \tag{2.9}$$

et donc étant la transformée de Fourier d'une fonction de $L^1(\mathbb{R})$, $m^k\widehat{\varphi}^{(p)}$ est bien bornée, donc $\widehat{\varphi} \in \mathscr{S}(\mathbb{R})$. □

Le théorème 2.39 a pour conséquence

Corollaire 2.40. *L'opérateur*

$$\mathscr{F} : \mathscr{S}(\mathbb{R}) \longrightarrow \mathscr{S}(\mathbb{R})$$

est une bijection.

Démonstration. Soit $\varphi \in \mathscr{S}(\mathbb{R})$. Alors $\varphi \in L^1(\mathbb{R})$, et $\widehat{\varphi} \in \mathscr{S}(\mathbb{R}) \subset L^1(\mathbb{R})$, donc on peut appliquer la formule d'inversion :

$$\forall x \in \mathbb{R}, \quad \varphi(x) = \widehat{\widehat{\varphi}}(-x) = \mathscr{F}(\mathscr{G}(\varphi))(x). \tag{2.10}$$

Mais en notant $\sigma : \varphi \mapsto (x \mapsto \varphi(-x))$, on a $\mathscr{G} = \sigma \circ \mathscr{F}$; de plus, on constate facilement que $\sigma(\varphi) \in \mathscr{S}(\mathbb{R})$ donc par le théorème 2.39, $\sigma(\mathscr{F}(\varphi)) = \mathscr{G}(\varphi) \in \mathscr{S}(\mathbb{R})$, et donc par (2.10) $\varphi \in \mathscr{F}(\mathscr{S}(\mathbb{R}))$. Autrement dit, $\mathscr{F}$ est surjective. □

68. On parle de fonction à décroissance rapide.

Remarque 2.41 (**)**.** En fait, $\mathscr{F} : \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\mathbb{R})$ est même un isomorphisme, c'est-à-dire que $\mathscr{F} : (\mathscr{S}(\mathbb{R}), d) \to (\mathscr{S}(\mathbb{R}), d)$ est continue, et son inverse, qui n'est autre que $\mathscr{G} : (\mathscr{S}(\mathbb{R}), d) \to (\mathscr{S}(\mathbb{R}), d)$, aussi. Pour le voir, on peut reprendre la formule (2.9) qui donne

$$N_{k,p}(\widehat{\varphi}) \leq C_{k,p} \sum_{\alpha \leq k, \beta \leq p} N_{\alpha,\beta}(\varphi).$$

Remarque 2.42. Comme on l'a vu dans les preuves qui précèdent, l'espace $\mathscr{S}(\mathbb{R})$ est très commode quand il s'agit de calculer avec la transformation de Fourier, car toutes les hypothèses (celles de la formule d'inversion ou de la proposition 2.15) sont satisfaites. L'importance de cet espace va beaucoup plus loin, car il est à la base de la notion de distribution tempérée, voir le paragraphe 3.4.

2.4 Cadre des mesures finies et des probabilités

On suggère la lecture de [Ouv09, Chapitre 12] et [GH13, 10.3].

On peut définir également la transformée de Fourier d'une mesure borélienne finie :

Définition-Proposition 2.43. — Soit μ une mesure borélienne (i.e. définie sur $\mathscr{B}(\mathbb{R})$) telle que $\mu(\mathbb{R}) < +\infty$. Alors on peut définir

$$\forall t \in \mathbb{R}, \quad \widehat{\mu}(t) = \int_{\mathbb{R}} e^{+itx} d\mu(x),$$

la transformée de Fourier de μ. De plus $\widehat{\mu} \in C_b^0(\mathbb{R})$ et

$$\|\widehat{\mu}\|_\infty \leq \mu(\mathbb{R}). \tag{2.11}$$

Autrement dit, on a défini l'opérateur

$$\mathscr{F} : \begin{array}{ccc} \mathcal{M}_f^+(\mathbb{R}) & \longrightarrow & C_b^0(\mathbb{R}) \\ \mu & \longmapsto & \widehat{\mu} \end{array}$$

où on a noté $\mathcal{M}_f^+$ l'ensemble des mesures boréliennes, positives, finies.

— Soit X une variable aléatoire, définie sur un espace de probabilité $(\Omega, \mathscr{A}, \mathbb{P})$ et à valeurs réelles. On appelle fonction caractéristique de X la transformée de Fourier de la loi de X, et on la note φ_X : d'après le théorème de transfert, on a

$$\forall t \in \mathbb{R}, \quad \varphi_X(t) = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(x) = \mathbb{E}(e^{itX}).$$

Démonstration. Á t fixé, la fonction $x \mapsto e^{itx}$ est continue donc borélienne, et μ-intégrable car μ est supposée finie. De plus par le théorème 3.16 donne la continuité de $\widehat{\mu}$, et la borne (2.11) découle du calcul

$$\forall t \in \mathbb{R}, \quad |\widehat{\mu}(t)| \leq \int_{\mathbb{R}} |e^{itx}| d\mu(x) \leq \int_{\mathbb{R}} d\mu(x) = \mu(\mathbb{R}).$$

□

Remarque 2.44. Remarquons qu'on a changé de convention par rapport aux paragraphes précédents, afin d'utiliser la convention habituellement utilisée en probabilités (il y a un signe + au lieu du signe - dans l'exponentielle, il n'y a plus de $\frac{1}{\sqrt{2\pi}}$, et on note t eu lieu de ξ la variable de Fourier). Néanmoins, on conserve les notations $\widehat{\cdot}$ et $\mathscr{F}$; soyez vigilants quant à la convention choisie.

Remarque 2.45 (*)**.** On n'a pas dit ici que $\mathscr{F}$ est une application linéaire, pour la bonne raison que $\mathcal{M}_f^+(\mathbb{R})$ n'est pas un espace vectoriel ; on peut arranger ce problème en étendant la définition précédente à $\mathcal{M}_f(\mathbb{R})$ l'ensemble des mesures signée voire complexes (voir par exemple [GH13, Paragraphe 2.4] ou [Rud98, Chapitre 6]). Notons qu'on peut définir $\mathcal{M}_f(\mathbb{R})$ comme l'espace vectoriel engendré par $\mathcal{M}_f^+(\mathbb{R})$, ou encore dans le cas réel comme l'ensemble des fonctions qui sont différences de deux mesures positives finies :

$$\mathcal{M}_f(\mathbb{R}) = \{\mu_+ - \mu_-, \ \ (\mu_+, \mu_-) \in \mathcal{M}_f^+(\mathbb{R})\}.$$

Remarque 2.46. À ceci près qu'on a changé de convention, la définition 2.43 généralise la définition 2.1 quand $f \in L^1(\mathbb{R})$ est positive (sachant que cette condition de signe est artificielle, voir la remarque 2.45) : en effet, si $f \in L^1(\mathbb{R})$ est positive, on peut considérer la mesure $\mu = fd\lambda$ de densité f par rapport à la mesure de Lebesgue, qui est bien une mesure finie, et alors en respectant les notations de chaque cadre, on a

$$\mu(t) = \int_{\mathbb{R}} e^{+itx} d\mu(x) = \int_{\mathbb{R}} e^{+itx} f(x) d\lambda(x) = \sqrt{2\pi}\widehat{f}(-t).$$

Remarque 2.47 (*Point de vue dualité)**.** On peut présenter un point de vue un peu différent qui consiste à définir une transformation de Fourier pour $T \in C^0_{\to 0}(\mathbb{R})'$ l'ensemble des formes linéaires continues sur $(C^0_{\to 0}(\mathbb{R}), \|\cdot\|_\infty)$: en effet on peut montrer qu'on peut prolonger une telle forme linéaire en une forme linéaire continue sur $(C^0_b(\mathbb{R}), \|\cdot\|_\infty)$ puis poser

$$\forall t \in \mathbb{R}, \quad \widehat{T}(t) = \mu(e_t), \qquad \text{où on a posé } \forall x \in \mathbb{R},\ e_t(x) = e^{itx}.$$

Pour voir le lien avec la définition 2.43, il suffit de voir que si μ est une mesure borélienne finie, alors on peut lui associer $T_\mu \in C^0_{\to 0}(\mathbb{R})$ en considérant

$$\forall \varphi \in C^0_{\to 0}(\mathbb{R}),\ \ T_\mu(\varphi) = \int_{\mathbb{R}} \varphi d\mu.$$

Dans ce cas, on retrouver $\widehat{T_\mu} = \widehat{\mu}$.

Concluons par quelques commentaires :

— Dans le cadre des probabilités, ce point de vue est classique : il consiste à considérer

$$\varphi \in C^0_{\to 0}(\mathbb{R}) \mapsto \mathbb{E}(\varphi(X))$$

pour identifier la loi de X.

— Nous allons en fait approfondir ce point de vue avec la notion de distribution tempérée au paragraphe 3.4.

— Les deux points de vue sont en fait équivalents : en effet une version [69] du théorème de Riesz (qui est un résultat difficile) affirme que l'application

$$\mu \in \mathcal{M}_f(\mathbb{R}) \longmapsto T_\mu \in (C^0_{\to 0}(\mathbb{R}))'$$

est en fait une bijection [70], et même une isométrie [71] surjective.

— [**] On pourrait se demander pourquoi on n'a pas travaillé directement avec $(C^0_b(\mathbb{R}))'$; du point de vue de la définition de la transformation de Fourier, en effet ça ne pose pas de problème. Mais du point de vue de la théorie de la mesure, ça pose deux difficultés :

  — l'espace $C^0_b(\mathbb{R})$ n'est pas séparable, ce qui explique que le théorème de Banach-Alaoglu séquentiel ne s'applique pas [72] dans $C^0_b(\mathbb{R})$,

  — l'espace $C^0_b(\mathbb{R})'$ est strictement plus gros qu'un espace de mesure borélienne : il contient ce qu'on appelle des limites de Banach, qu'on peut construire comme suit. Considérez l'espace $C^0_\ell(\mathbb{R})$ des fonctions réelles ayant une limite en $+\infty$ et considérons $T : \varphi \in C^0_\ell(\mathbb{R}) \mapsto \lim_{+\infty} \varphi$. Alors par le théorème de Hahn-Banach [73] on peut prolonger T en une forme linéaire continue sur $C^0_b(\mathbb{R})$, et il n'existe pas de mesure borélienne μ telle que $T(\varphi) = \int_{\mathbb{R}} \varphi d\mu$ pour tout $\varphi \in C^0_{\to 0}(\mathbb{R})$, car alors cette mesure serait nulle, ce qui est une contradiction.

69. Il y a plusieurs versions du résultat, suivant qu'on regarde uniquement des mesures positive, qu'on travaille sur $C^0(K)'$ (avec K compact) ou sur $(C^0_c(\Omega)'$ (avec Ω ouvert) ou encore $C^0_{\to 0}(\Omega)'$; voir par exemple ce lien ou celui-ci.

70. C'est la partie "surjection" qui est difficile.

71. Attention il faut mettre une norme sur $\mathcal{M}_f(\mathbb{R})$ pour donner un sens au mot isométrie.

72. Bien sûr, cela ne suffit pas à dire que la conclusion du théorème (que toute suite bornée admet une sous-suite qui converge simplement) est fausse ; il faut produire des contre-exemples. Pour cela, on peut par exemple considérer la suite $(T_n := \delta_n)_{n \in \mathbb{N}}$.

73. Ce résultat est hors programme ; il permet de donner l'existence d'un prolongement sans que l'espace de départ soit dense, comme on le fait au corollaire 2.51 du Chapitre III. On se restreint par contre au cas où l'espace d'arrivée est le corps de base ($\mathbb{R}$ ou $\mathbb{C}$), et on perd l'unicité.

Remarque 2.48. Notez qu'on n'a pas dit que $\widehat{\mu} \in C^0_{\to 0}(\mathbb{R})$, car ceci est faux en général ; autrement dit, le lemme de Riemann-Lebesgue ne s'adapte pas à ce cadre plus général. Pour le constater, il suffit d'observer

$$\widehat{\delta_0} = 1,$$

qui est un exemple intéressant à connaître.

La première question naturelle est de se demander si l'opération $\mathscr{F}$ reste injective :

Proposition 2.49. *L'application $\mathscr{F} : \mathcal{M}_f^+(\mathbb{R}) \to C_b^0(\mathbb{R})$ est injective.*

Remarque 2.50. Dans le cadre des probabilités, cet énoncé affirme que la fonction caractéristique de la variable X, φ_X, caractérise la loi de X.

Éléments de preuve : Soit $\varphi \in L^1(\mathbb{R})$. Alors par le théorème de Fubini [74]

$$\int_{\mathbb{R}} \widehat{\varphi} d\mu = \int_{\mathbb{R}} \int_{\mathbb{R}} \varphi(x) e^{+itx} dx d\mu(t) = \int_{\mathbb{R}} \varphi(x) \int_{\mathbb{R}} e^{+itx} d\mu(t) dx = \int_{\mathbb{R}} \widehat{\mu}(x) \varphi(x) dx.$$

La formule d'inversion avec la nouvelle convention affirme que si $\varphi \in C^0(\mathbb{R}) \cap L^1(\mathbb{R})$ est telle que $\widehat{\varphi} \in L^1(\mathbb{R})$, alors

$$\forall x \in \mathbb{R}, \quad \varphi(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{\varphi}(t) e^{-itx} dt = \frac{1}{2\pi} \widehat{\widehat{\varphi}}(-x)$$

et donc

$$\int_{\mathbb{R}} \varphi(x) d\mu = \int_{\mathbb{R}} \widehat{\widehat{\varphi}}(-x) d\mu(x) = \int_{\mathbb{R}} \widehat{\varphi}(-x) \widehat{\mu}(x) dx.$$

Ainsi, si $\widehat{\mu} = \widehat{\nu}$ où $(\mu, \nu) \in \mathcal{M}_f^+(\mathbb{R})$, alors $\int_{\mathbb{R}} \varphi d\mu = \int_{\mathbb{R}} \varphi d\nu$ pour tout $\varphi \in D := \{\phi \in C^0(\mathbb{R}) \cap L^1(\mathbb{R})$ telle que $\widehat{\phi} \in L^1(\mathbb{R})\}$. L'espace D contient $C_c^2(\mathbb{R})$ donc est dense dans $C^0_{\to 0}(\mathbb{R})$. Les résultats classiques [75] de théorie de la mesure affirment en conséquence que $\mu = \nu$. □

Remarque 2.51. Voir [Ouv09, Proposition 12.7] : on peut en fait avoir un théorème d'inversion adapté à ce cadre : si $\mu \in \mathcal{M}_f^+(\mathbb{R})$ est telle que $\widehat{\mu} \in L^1(\mathbb{R})$, alors μ est une mesure de la forme $h d\lambda$, et

$$h(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \widehat{\mu}(t) e^{-itx} dx \text{ pour presque tout } x \in \mathbb{R}.$$

Établissons un équivalent de la proposition 2.15 dans le cadre des probabilités [76] :

Proposition 2.52. *Soit X une variable aléatoire ayant un moment d'ordre $n \in \mathbb{N}$. Alors φ_X est de classe C^n sur $\mathbb{R}$, et*

$$\forall k \in [\![0, n]\!], \quad \varphi_X^{(k)}(t) = i^k \mathbb{E}\left[X^k e^{itX}\right], \quad \textit{et en particulier} \quad \varphi^{(k)}(0) = i^k \mathbb{E}(X^k).$$

Démonstration. Il suffit d'appliquer la version itérée du théorème 3.21 au Chapitre IV (voir la remarque 3.22 au Chapitre IV). □

Remarque 2.53. Il est intéressant de se poser la question de la réciproque [77] :

1. dans [Ouv09, page 204] on trouve : si φ_X est k fois dérivable en 0 avec $k \geq 2$, alors
   — si k est pair, X admet des moments jusqu'à l'ordre k,
   — si k est impair, X admet des moments jusqu'à l'ordre $k-1$.

---

74. Qui s'applique car

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |\varphi(x) e^{+itx}| dx d\mu(t) = \|\varphi\|_{L^1(\mathbb{R})} \mu(\mathbb{R}) < +\infty.$$

75. C'est la partie unicité du théorème de Riesz évoqué à la remarque 2.47.

76. On peut bien sûr l'énoncer dans le cadre des mesures finies, mais comme les applications que nous avons en tête sont focalisées sur les probabilités, on préfère cette formulation.

77. Il m'est d'avis que l'un ou l'autre de ces résultats fait un bon développement. Le premier est probablement un peu court mais peut certainement être facilement complété. Le second est plus technique, et très transverse.

2. [Ouv09, page 205] donne un contre-exemple intéressant expliquant qu'on ne peut pas obtenir l'existence d'un moment d'ordre 1 en supposant φ_X dérivable.

Concluons par quelques remarques et suggestions de lecture dans la direction du lien entre les probabilités et l'analyse de Fourier :

1. Comme pour la transformation de Fourier, on peut définir la transformation de Fourier d'une mesure borélienne finie sur $\mathbb{R}^d$. En particulier, le théorème de Fubini montre

$$\forall t \in \mathbb{R}^2, \quad \widehat{\mu_1 \otimes \mu_2}(t_1, t_2) = \widehat{\mu_1}(t_1)\widehat{\mu_2}(t_2).$$

En terme de probabilités, on a le corollaire

Corollaire 2.54. *Soit (X_1, X_2) deux variables aléatoires réelles. Alors X_1 et X_2 sont indépendantes si et seulement si*

$$\varphi_{(X_1,X_2)}(t_1, t_2) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2).$$

2. On peut définir la convolution de deux mesures $(\mu_1, \mu_2) \in \mathcal{M}_f^+(\mathbb{R})$ par

$$\mu_1 * \mu_2(A) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_A(x+y) d\mu_1(x) d\mu_2(y).$$

Remarquons que cette notion est bien cohérente avec la convolution des fonctions de $L^1(\mathbb{R})$: en effet si $\mu_1 = f d\lambda$ et $\mu_2 = g d\lambda$ avec $(f, g) \in L^1(\mathbb{R})$ et positive[78], alors $\mu_1 * \mu_2$ est une mesure à densité par rappot à $d\lambda$ et de densité $f * g$.
On peut généraliser la proposition 2.12 et montrer $\widehat{\mu_1 * \mu_2}(t) = \mu_1(t)\mu_2(t)$. En terme de probabilités, cela s'énonce[79] :

Corollaire 2.55. *Soit (X_1, X_2) deux variables aléatoires indépendantes. Alors*

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t).$$

On peut appliquer ce résultat pour identifier la loi d'une somme de variables aléatoires indépendantes et ayant des lois normales, ou encore ayant des lois de Poisson.

3. On considère $(\mu_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{M}_f^+(\mathbb{R})$. Si on prend le point de vue de la dualité, il y a 3 espaces de fonctions tests possibles

$$C_c^0(\mathbb{R}) \subset C_{\to 0}^0(\mathbb{R}) \subset C_b^0(\mathbb{R})$$

qui chacun corresondent à une convergence :

$$\mu_n \text{ converge vers } \mu \text{ si } \forall \varphi \in S, \quad \int_{\mathbb{R}} \varphi d\mu_n \xrightarrow[n\to+\infty]{} \int_{\mathbb{R}} \varphi d\mu$$

(a) est appelée convergence vague si $S = C_c^0(\mathbb{R})$
(b) est appelée convergence faible si $S = C_{\to 0}^0(\mathbb{R})$
(c) est appelée convergence étroite si $S = C_b^0(\mathbb{R})$.

Bien sûr, convergence étroite implique convergence faible, qui elle-même implique convergence vague. Mais les réciproques sont fausses : la suite $(\delta_n)_{n\in\mathbb{N}}$ converge vers 0 faiblement, mais pas étroitement. La suite $(n\delta_n)_{n\in\mathbb{N}}$ converge vaguement vers 0, mais pas faiblement. Le premier contre-exemple montre également que l'ensemble des probabilités (mesures boréliennes positives de masse 1) n'est pas fermé pour les deux premières convergences, ce qui explique que les probabilistes privilégient la convergence étroite. D'ailleurs, la convergence en loi de $(X_n)_{n\in\mathbb{N}}$ vers X correspond à la convergence étroite de $(\mathbb{P}_{X_n})_{n\in\mathbb{N}}$ vers $\mathbb{P}_X$.
Pour réconcilier ces convergences, on peut montrer (voir [Ouv09, Prop 14.3])

78. Rappelons que la positivité n'est pas vraiment nécessaire, elle est là pour dire qu'on manipule des mesures positives.
79. Bien sûr, on peut le montrer directement avec le formalisme des probabilités, ce qui est particulièrement simple ici.

Lemme 2.56. *Soit $(\mu_n)_{n\in\mathbb{N}}$ une suite de $\mathcal{M}_f^+(\mathbb{R})$, et $\mu \in \mathcal{M}_f^+(\mathbb{R})$. Alors*

$$\mu_n \text{ converge étroitement vers } \mu \iff \begin{cases} \mu_n \text{ converge vaguement vers } \mu \\ \text{et } \mu_n(\mathbb{R}) = \int_{\mathbb{R}} 1 d\mu_n \xrightarrow[n\to+\infty]{} \mu(\mathbb{R}) \end{cases}.$$

On en vient au théorème de Lévy qui permet de caractériser la convergence en loi :

Théorème 2.57. *Soit $(X_n)_{n\in\mathbb{N}}$ et X des variables aléatoires réelles.*

(a) X_n converge en loi vers $X \iff \varphi_{X_n}$ converge simplement vers φ_X,

(b) [*] plus précisément, si $(\varphi_{X_n})_{n\in\mathbb{N}}$ converge simlplement vers une fonction φ, alors

$$\text{il existe } X \text{ une v.a. telle que } \varphi = \varphi_X \quad \iff \quad \varphi \text{ est continue en } 0.$$

Quelques commentaires sur la preuve :

— Le sens direct de la première partie du résultat est élémentaire,

— Le sens réciproque de la première partie du résultat repose sur la formule d'inversion, la densité de $\{\phi \in C^0(\mathbb{R}) \cap L^1(\mathbb{R}), \widehat{\phi} \in L^1(\mathbb{R})\}$ dans $C^0_{\to 0}(\mathbb{R})$ et le lemme 2.56. Ce résultat est à connaître, notamment il est au cœur de la preuve du théorème central limite.

— La seconde partie du résultat est beaucoup plus délicate, car elle nécessite un argument de compacité. Il existe plusieurs démonstration, décrivons rapidement une preuve reposant sur le théorème de Banach-Alaoglu séquentiel : on voit $(\mathbb{P}_{X_n})_{n\in\mathbb{N}}$ comme une suite de la boule unité de $C^0_{\to 0}(\mathbb{R})'$. Comme $C^0_{\to 0}(\mathbb{R})$ est séparable, on a la convergence de $(\mathbb{P}_{X_{\sigma(n)}})_{n\in\mathbb{N}}$ vers une mesure μ, où σ est une extraction. Il reste à voir que $\mathbb{P}_{X_{\sigma(n)}}(\mathbb{R})$ converge vers $\mu(\mathbb{R})$, ce qui est vrai grâce à la continuité de φ en 0 [80]. Ainsi μ est une mesure de probabilité, et il existe X une v.a. telle que $\mu = \mathbb{P}_X$. On conclut à la convergence totale de la suite (sans extraction) par le fait que toute suite extraite admet pour unique valeur d'adhérence la mesure μ.

— En fait dans la preuve évoquée au point précédent, on a en fait montré un théorème de compacité dans l'espace des mesures de probabilité, appelé théorème de Prokhorov :

Théorème 2.58. *Soit $(\mu_n)_{n\in\mathbb{N}}$ une suite de probabilités. On suppose $(\mu_n)_{n\in\mathbb{N}}$ tendue, c'est-à-dire*

$$\forall \varepsilon > 0, \ \ \exists K_\varepsilon \subset \mathbb{R} \text{ compact tel que } \forall n \in \mathbb{N}, \ \ \mu_n(K_\varepsilon^c) \leq \varepsilon.$$

Alors il existe σ une extraction et μ une mesure de probabilité telle que $\mu_{\sigma(n)}$ converge étroitement vers μ.

L'hypothèse de tension est là pour assurer que la suite $(\mu_n)_{n\in\mathbb{N}}$ ne perd pas de masse "'a l'infini".

Notez qu'on peut donner une preuve différente de celle évoquée au point précédent, si on reste en dimension 1 ; en effet on peut utiliser le théorème de Helly qui donne un critère de compacité pour la convergence simple sur les suites de fonctions croissantes. Voir [GK19].

4. On peut s'intéresser au problème des moments : étant donné deux variables aléatoires X, Y, telles que

$$\forall n \in \mathbb{N}, \ \ \mathbb{E}(X^n) = \mathbb{E}(Y^n)$$

peut-on conclure que X et Y on la même loi ?

(a) la réponse est non en général,

(b) la réponse est oui si les variables aléatoires sont supposées bornées (il suffit d'appliquer le théorème de Weierstrass pour obtenir $\varphi_X = \varphi_Y$)

(c) la réponse reste positive si les variables ont un moment d'ordre exponentiel (et la preuve est la même que le développement classique sur un critère de densité pour les polynômes orthogonaux, voir [BMP05, Exo 3.7]).

80. Voir dans la littérature pour les détails de cette affirmation.

3 Distributions

Notez que dans le programme d'analyse et probabilités de la session 2020, le paragraphe concernant les distributions est le seul à avoir évolué ! La notion de distribution "tout court" est revenue au programme, alors que dans les sessions précédentes, seule la notion de distribution tempérée était au programme. C'est ce qui expliquait la place de cette partie du cours dans le chapitre consacré à l'analyse de Fourier. Mais maintenant que les distributions "tout court" sont à nouveau au programme, précisons qu'une partie de cette section peut être considérée indépendante de l'analyse de Fourier.

En terme de référence, on conseille [Bon06] qui étant destiné à des étudiants en école d'ingénieurs, correspond plutôt bien à ce qui peut être attendu du jury, à savoir plus une manipulation des outils de ce chapitre, qu'une connaissance affinée de leur construction. On peut également citer [Sch97] ; j'apprécie personnellement beaucoup [Zui02].

3.1 Définition et premiers exemples

Les distributions sont une notion de "fonction généralisée". Prenons une fonction $f : I \to \mathbb{R}$ où $I \subset \mathbb{R}$ est un intervalle ouvert : l'idée est de remplacer la connaissance habituelle (à un $x \in I$ on associe $f(x) \in \mathbb{R}$) par la connaissance de

$$\varphi \longmapsto \int_{\mathbb{R}} f(x)\varphi(x)dx \tag{3.1}$$

où $\varphi : I \to \mathbb{R}$ sera dans une certaine classe de fonctions, qu'on appellera les fonctions tests. On peut voir cette valeur comme une moyenne pondérée par φ de f. Cette idée est la même que celle que vous utilisez en probabilité lorsque vous calculez

$$\mathbb{E}(\varphi(X))$$

pour toute fonction φ continue bornée, afin d'identifier la loi de X (voir aussi la remarque 2.47).

L'idée des distributions est donc de regarder un objet abstrait qui à une fonction φ dans l'espace des fonctions tests associe une valeur réelle, et qui satisfait des propriétés basiques comme c'est le cas de la formule (3.1), comme par exemple la linéarité. Il faudra ajouter une certaine propriété de continuité.

Précisons que plus la classe de fonctions tests choisie est petite, plus on aura d'objets qui rentrent dans cette définition. Néanmoins, on veut quand même que la vision où on remplace une fonction f par l'objet défini en (3.1) soit injective.

Ces réflexions amènent à la définition suivante :

Définition 3.1. Soit I un ouvert [81] de $\mathbb{R}$. On dit que T est une distribution sur I si $T : C_c^\infty(I) \to \mathbb{K}$ est une application linéaire qui satisfait :

$$\forall K \text{ compact inclus dans } I,\ \exists C \in \mathbb{R},\ \exists p \in \mathbb{N},$$
$$\forall \varphi \in C_c^\infty(I) \text{ telle que } Supp(\varphi) \subset K,\ |T(\varphi)| \le C \sup_{k\in[\![0,p]\!]} \|\varphi^{(k)}\|_\infty. \tag{3.2}$$

Remarque 3.2. On note en général $\mathcal{D}(I) = C_c^\infty(I)$. Aussi, si T est une distribution et $\varphi \in \mathcal{D}(I)$, on note

$$\langle T, \varphi \rangle = T(\varphi).$$

Ce crochet est appelé crochet de dualité, voir la remarque 3.3.

Remarque 3.3. On note en général $T \in \mathcal{D}'(I)$ quand T est une distribution, en référence à l'idée d'une dualité, c'est-à-dire de l'ensemble des formes linéaires continues sur $\mathcal{D}(I)$. On peut effectivement constater que (3.2) correspond bien à une condition de continuité :

— on peut définir une convergence sur $\mathcal{D}(I)$, à savoir que si $(\varphi_n)_{n\in\mathbb{N}}$ est une suite de $\mathcal{D}(I)$ et $\varphi \in \mathcal{D}(I)$, on dit que $\varphi_n \xrightarrow[n\to+\infty]{\mathcal{D}(I)} \varphi$ si

$$\text{il existe } K \text{ compact tel que } \forall n \in \mathbb{N},\ Supp(\varphi_n) \subset K \text{ et } \forall k \in \mathbb{N},\ \varphi_n^{(k)} \xrightarrow[n\to+\infty]{\|\cdot\|_\infty} \varphi^{(k)}. \tag{3.3}$$

81. On peut penser que I est un intervalle, mais ça n'est pas nécessaire pour le moment.

Et vérifier que si $T : \mathcal{D}(I) \to \mathbb{K}$ est linéaire, alors T est continue[82] si et seulement si elle vérifie (3.2).

— Néanmoins, on ne cherche pas à définir une topologie sur $\mathcal{D}(I)$ qui amène à cette convergence, et donc à justifier "au sens usuel" la notation $\mathcal{D}'(I) = (\mathcal{D}(I))'$; il est effectivement explicité dans le programme qu'on n'attend pas des étudiants de connaître les subtilités liées à ces questions. Ceci est parfaitement compréhensible, car la topologie en question (qui existe !) n'est pas issue d'une métrique, et donc sort du programme de topologie du concours.

Exemple 3.4 (Exemple fondamental : injection de $L^1_{loc}(I)$ dans $\mathcal{D}'(I)$)**.** Reprenons notre exemple introductif, et voyons comment justifier rigoureusement le fait qu'une distribution peut être vue comme une fonction généralisée : à $f \in L^1_{loc}(I)$, on associe l'application

$$[f] : \begin{array}{ccc} \mathcal{D}(I) & \longrightarrow & \mathbb{K} \\ \varphi & \longmapsto & \int_I f(x)\varphi(x)dx \end{array}$$

qui est bien définie car φ est à support compact, bornée, et f est localement intégrable. Il est clair que $[f]$ est linéaire, et elle satisfait :

$$\forall K \text{ compact inclus dans } I,\ \forall \varphi \in C^\infty_c(I) \text{ telle que } Supp(\varphi) \subset K,$$
$$|\langle [f], \varphi\rangle| \leq \int_K |f(x)\varphi(x)|dx \leq \|f\|_{L^1(K)}\|\varphi\|_\infty.$$

Ainsi (3.2) est satisfaite avec $C = \|f\|_{L^1(K)}$[83] et $p = 0$.

On a ainsi défini une façon d'inclure $L^1_{loc}(I)$ dans $\mathcal{D}'(I)$ via l'application

$$\begin{array}{ccc} L^1_{loc}(I) & \longrightarrow & \mathcal{D}'(I) \\ f & \longmapsto & [f] \end{array}.$$

Pour que cette "inclusion" soit légitime, il faut vérifier que l'application précédente est injective. Ceci est le propos du lemme 5.61 au Chapitre IV.

Avec la pratique, on peut faire l'abus de confondre f et $[f]$.

Exemple 3.5. Pour $a \in \mathbb{R}$, considérons l'application

$$\varphi \in \mathcal{D}'(\mathbb{R}) \mapsto \varphi(a).$$

Il s'agit bien d'une application linéaire, et elle satisfait $|\varphi(a)| \leq \|\varphi\|_\infty$ donc c'est une distribution. On l'appelle masse de Dirac en a, et on la note δ_a.

En fait, cet exemple peut être généralisé : si μ est une mesure borélienne sur I, qui est finie sur les compacts de I[84], alors on peut considérer

$$\varphi \in \mathcal{D}'(I) \mapsto \int_I \varphi(x)d\mu(x)$$

qui est une distribution d'ordre 0[85]. Le cas précédent correspond au cas de la mesure de Dirac en a.

Notons qu'on peut également considérer, pour $a \in \mathbb{R}$ et $k \in \mathbb{N}$, l'application

$$\varphi \in \mathcal{D}'(\mathbb{R}) \mapsto \varphi^{(k)}(a).$$

Il s'agit bien d'une application linéaire, et elle satisfait $|\varphi^{(k)}(a)| \leq \|\varphi^{(k)}\|_\infty$ donc c'est une distribution. Ces distributions sont au signe près les dérivées successives de δ_a, voir le paragraphe 3.2.

82. Au sens séquentiel, i.e. $\langle T, \varphi_n\rangle$ converge vers $\langle T, \varphi\rangle$ si $(\varphi_n)_{n\in\mathbb{N}}$ converge vers φ au sens (3.3).

83. Notons qu'il est autorisé que C dépende du compact K.

84. On appelle parfois une telle mesure une mesure de Radon.

85. En fait, en lien avec la remarque 2.47, on peut montrer que si T est une distribution d'ordre 0 sur I, alors il existe une unique mesure de Radon μ telle que

$$\forall \varphi \in \mathcal{D}(I),\ \ \langle T, \varphi\rangle = \int_I \varphi d\mu.$$

Ce résultat est une conséquence simple d'une version du théorème de Riesz (qui n'est lui, pas simple).

Remarque 3.6. Lorsqu'il est possible de choisir p dans (3.2) qui soit indépendant du compact K, on dit que T est une distribution d'ordre fini, et on définit l'ordre de T comme le plus petit entier p qui convient. Dans l'exemple 3.4, les fonctions de $L^1_{loc}(I)$ définissaient des distributions d'ordre 0.

Exercice 3.7. Notons qu'il existe des distributions qui ne sont pas d'ordre fini : vérifiez que l'application

$$\varphi \in \mathcal{D}(I) \longmapsto \sum_{n\in\mathbb{N}} \varphi^{(n)}(n)$$

est bien définie, est une distribution, et qu'elle est d'ordre infini.

Exemple 3.8 (Valeur principale de Cauchy)**.** La fonction $x \in \mathbb{R}^* \mapsto 1/x$ n'est pas localement intégrable sur $\mathbb{R}$. Néanmoins, on peut constater que si $\varphi \in \mathcal{D}(\mathbb{R})$,

$$\int_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]} \frac{\varphi(x)}{x}dx = \int_{[-A,A]\setminus[-\varepsilon,\varepsilon]} \frac{\varphi(x)}{x}dx = \int_{[-A,A]\setminus[-\varepsilon,\varepsilon]} \frac{\varphi(x)-\varphi(0)}{x}dx \tag{3.4}$$

où $A \in \mathbb{R}$ est assez grand pour avoir $Supp(\varphi) \subset [-A, A]$. Comme on a la domination

$$\forall x \in \mathbb{R}^*, \quad \left|\frac{\varphi(x)-\varphi(0)}{x}\right| \leq \|\varphi'\|_\infty$$

le théorème de convergence dominée affirme que le terme intervenant dans (3.4) a une limite, et donc on pose

$$\langle vp(1/x), \varphi\rangle = \lim_{\varepsilon\to 0} \int_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]} \frac{\varphi(x)}{x}dx \left(= \int_{[-A,A]\setminus\{0\}} \frac{\varphi(x)-\varphi(0)}{x}dx \right).$$

Il s'agit d'une distribution car par les calculs précédents on obtient :

$$\forall A \in \mathbb{R},\ \forall \varphi \in C^\infty_c(\mathbb{R}) \text{ telle que } Supp(\varphi) \subset [-A, A], \quad |\langle vp(1/x), \varphi\rangle| \leq 2A\|\varphi'\|_\infty,$$

et on l'appelle distribution “valeur principale de $1/x$”. L'inégalité précédente montre qu'elle est d'ordre fini, au plus 1, et on peut vérifier qu'elle est en fait exactement d'ordre 1 (voir la littérature).

3.2 La dérivée au sens des distributions

Une question naturelle est de se demander comment définir des opérations sur $\mathcal{D}'(I)$. Cette question n'est pas simple en générale. Il n'y a pas de souci par exemple pour définir la somme de deux distributions (et plus généralement $\mathcal{D}'(I)$ est un espace vectoriel), mais par exemple le produit de deux distributions n'a pas de sens en général.

La méthode à retenir, c'est d'essayer de voir comment se comporte l'opération sur une fonction classique f (ou des fonctions classiques f et g si l'opération fait intervenir deux fonctions), notamment en faisant passer l'opération sur f à une opération sur la fonction test φ, et de voir si on peut proposer une définition qui généralise cette opération à $\mathcal{D}'(I)$: on dit qu'on procède par dualité.

Appliquons cette méthode à l'opération qui donne une grande célébrité à la théorie des distributions, la dérivation. Supposons que $f \in C^1(\mathbb{R})$: elle est ainsi dérivable au sens classique, et $f' \in C^0(\mathbb{R}) \subset L^1_{loc}(\mathbb{R})$ [86]. Essayons de réécrire le comportement de la distribution $[f']$ sur une fonction test φ : par intégration par partie, on voit que

$$\forall \varphi \in \mathcal{D}(I), \quad \langle [f'], \varphi\rangle = \int_I f'(x)\varphi(x)dx = -\int_I f(x)\varphi'(x)dx = -\langle [f], \varphi'\rangle. \tag{3.5}$$

Notez qu'il n'y a pas de terme de bord dans l'intégration par parties car φ est à support compact dans I.

Ceci amène à la définition suivante :

Définition-Proposition 3.9 (Dérivée au sens des distributions)**.** Soit $T \in \mathcal{D}'(I)$. On peut définir

$$\forall \varphi \in \mathcal{D}'(I), \quad \langle T', \varphi\rangle := -\langle T, \varphi'\rangle$$

et T' est une distribution, qu'on appelle dérivée au sens des distributions de T.

86. Si on avait seulement supposé f dérivable, il n'est pas clair que f' serait une fonction localement intégrable.

Démonstration. L'application T' est effectivement linéaire, et si K est un compact inclus dans I, comme T est une distribution, il existe $C \in \mathbb{R}$ et $p \in \mathbb{N}$ tel que

$$\forall \varphi \in C_c^\infty(I) \text{ telle que } Supp(\varphi) \subset K,\ \ |\langle T, \varphi\rangle| \leq C \sup_{k \in [\![0,p]\!]} \|\varphi^{(k)}\|_\infty.$$

En conséquence,

$$\forall \varphi \in C_c^\infty(I) \text{ telle que } Supp(\varphi) \subset K,\ \ |\langle T', \varphi\rangle| = |\langle T, \varphi'\rangle| \leq C \sup_{k \in [\![0,p]\!]} \|\varphi^{(k+1)}\|_\infty \leq C \sup_{k \in [\![0,p+1]\!]} \|\varphi^{(k)}\|_\infty.$$

Et ainsi T' est une distribution. □

Remarque 3.10. Bien sûr, si $f \in C^1(\mathbb{R})$, on a donc deux objets à comparer :

1. la dérivée classique f' comme définie au Chapitre I,
2. la dérivée au sens des distributions, à savoir $[f]'$.

D'après la formule (3.5) et le lemme 5.61 au Chapitre IV, ces deux objets coïncident, autrement dit

$$[f'] = [f]'.$$

En pratique, on rappelle qu'on peut s'autoriser à confondre f et $[f]$; de même on peut écrire f' pour désigner la dérivée au sens des distributions de f. Mais attention, il se peut que $[f]'$ ait un sens, mais pas $[f']$ (si f n'est pas dérivable au sens classique).

Remarque 3.11. Le résultat précédent, aussi simple peut-il paraître, à pour conséquence étrange que

toute distribution est dérivable au sens des distributions ;

et même infiniment dérivable puisqu'on peut sans problème itérer le processus.

Prenons le cas d'une fonction $f \in L^1_{loc}(\mathbb{R})$; la distribution $[f]'$ existe forcément, mais attention, on ne dit pas que $[f]'$ est une fonction. A priori $[f]'$ est une distribution. On verra un peu plus loin ce qu'on peut dire si cette distribution est en fait une fonction.

Remarque 3.12. On vérifie sans peine que si T est une distribution d'ordre au plus p, alors T' est une distribution d'ordre au plus $p + 1$;

Exemple 3.13 (Exemple fondamental : la fonction de Heaviside)**.** Si je vous demande "une fonction non-dérivable" (au sens classique), probablement le premier exemple que vous allez me répondre est $x \mapsto |x|$; mais c'est parce que vous avez appris qu'une condition nécessaire pour qu'une fonction soit dérivable, est qu'elle soit continue. Du coup vous avez reformulé (à tort ou à raison) ma question en "une fonction continue non-dérivable".

Il se trouve que l'exemple $f : x \mapsto |x|$ n'est pas très intéressant pour notre propos ; on l'a évoqué plusieurs fois précédemment dans ce chapitre, mais comme il s'agit d'une fonction continue et C^1 par morceaux, on peut tout à fait justifier l'intégration par parties de (3.5), et finalement même si effectivement la fonction n'est pas dérivable en 0, rien n'empêche de considérer la fonction $x \in \mathbb{R}^* \mapsto \text{signe}(x)$, qui est à la fois la dérivée classique de f sur $\mathbb{R}^*$, et également la dérivée au sens des distributions de $[f]$ en tant que distribution sur $\mathbb{R}$ (le fait que la fonction ne soit pas définie en 0 n'a pas d'importance puisqu'ici on la considère comme une fonction de $L^1(\mathbb{R})$). Plus précisément, $[f]' = [\text{signe}]$.

Un exemple plus intéressant est donc de considérer une fonction non continue, par exemple

$$H = \mathbb{1}_{\mathbb{R}_+}$$

qu'on appelle fonction de Heaviside. Elle est dérivable au sens classique sur $\mathbb{R}^*$, et sa dérivée est nulle. D'ailleurs, on comprend que ce serait une mauvaise idée de faire comme précédemment et d'ignorer ce qui se passe en 0, car on a perdu de l'information [87] en dérivant, puisque H n'est pas une fonction constante.

87. On perd toujours de l'information en dérivant, car on est toujours "à une constante additive près" quand on essaie d'intégrer : mais ici on perd plus qu'une constante additive.

Regardons maintenant ce qui se passe au sens des distributions :

$$\forall \varphi \in \mathcal{D}(\mathbb{R}), \quad \langle [H]', \varphi \rangle = -\langle [H], \varphi' \rangle = -\int_0^{+\infty} \varphi'(x)dx = \varphi(0) - \varphi(+\infty) = \varphi(0). \tag{3.6}$$

On constate effectivement que $[H]'$ n'est pas représentable par une fonction[88] Plus précisément, on observe que $[H]' = \delta_0$.

Nous allons généraliser l'exemple 3.13 au cas d'une fonction C^1 par morceaux :

Proposition 3.14 (Formule des sauts)**.** *Soit I un intervalle ouvert, et $f : I \to \mathbb{R}$ une fonction C^1 par morceaux. On note $a_1 < \ldots < a_n$ les discontinuités de f. Alors*

$$[f]' = [f'] + \sum_{i=1}^{n} \Big[f(a_i^+) - f(a_i^-) \Big] \delta_{a_i},$$

où $[f']$ est la distribution associée à la dérivée classique de f, qui est définie sur $I \setminus \{a_i\}_{i \in [\![1,n]\!]}$, et est une fonction continue par morceaux.

Démonstration. Soit $\varphi \in \mathcal{D}(I)$. Alors en posant $a_0 = a$ et $a_{n+1} = b$ où $I =]a, b[$, on a

$$\begin{aligned}
\langle [f]', \varphi \rangle &= -\langle [f], \varphi' \rangle = -\int_I f(x)\varphi'(x)dx \\
&= -\sum_{i=0}^{n} \int_{a_i}^{a_{i+1}} f(x)\varphi'(x) = -\sum_{i=0}^{n} \left([f\varphi]_{a_i^+}^{a_{i+1}^-} - \int_{a_i}^{a_{i+1}} f'(x)\varphi(x)dx \right) \\
&= -\sum_{i=0}^{n} \Big[f(a_{i+1}^-)\varphi(a_{i+1}) - f(a_i^+)\varphi(a_i) \Big] + \int_a^b f'(x)\varphi(x)dx \\
&= -\sum_{i=1}^{n} \Big[f(a_i^-) - f(a_i^+) \Big] \varphi(a_i) + \int_a^b f'(x)\varphi(x)dx \\
&= \sum_{i=1}^{n} \Big[f(a_i^+) - f(a_i^-) \Big] \langle \delta_{a_i}, \varphi \rangle + \langle f', \varphi \rangle
\end{aligned}$$

□

Exemple 3.15. Pour $a \in I$, on peut considérer la dérivée de δ_a, à savoir

$$\varphi \mapsto -\varphi'(a).$$

On ne peut pas vraiment simplifier cette expression.

Exercice 3.16 (Dérivée de $\ln(|x|)$)**.** On considère $f : x \in \mathbb{R}^* \mapsto \ln|x|$.

1. Montrer que $f \in L^1_{loc}(\mathbb{R})$.
2. Montrer que la dérivée classique de f est définie sur $\mathbb{R}^*$, mais n'est pas dans $L^1_{loc}(\mathbb{R})$.
3. Montrer que
$$[f]' = vp(1/x).$$

Dans le Chapitre I, un résultat fondamental est que si une fonction est dérivable sur un intervalle, de dérivée nulle, alors elle est constante. Ce résultat permet de mesurer l'information perdue quand on dérive une fonction (à savoir qu'à une constante additive près, on ne perd par d'information).

Ce résultat a un équivalent dans le cadre des distributions :

88. Cela signifie qu'il n'existe pas $g \in L^1_{loc}(\mathbb{R})$ tel que

$$\forall \varphi \in \mathcal{D}(\mathbb{R}), \quad \langle [H]', \varphi \rangle = \int_{\mathbb{R}} g(x)\varphi(x)dx.$$

La raison est que si une telle fonction existait, en se restreignant aux fonctions φ à support dans $\mathbb{R}^*$, on aurait $g = 0$ presque partout sur $\mathbb{R}^*$, ce qui est pareil que $g = 0$ presque partout sur $\mathbb{R}$. Mais ceci contredit la formule (3.6).

Lemme 3.17. *Soit I un intervalle ouvert. Si $T \in \mathcal{D}'(I)$ est elle que $T' = 0$, alors il existe $c \in \mathbb{R}$ telle que $T = [c]$.*

Démonstration. — Remarquons que si $\varphi \in \mathcal{D}(I)$, alors on peut considérer $x \mapsto \int_{-\infty}^{x} \varphi(t)dt$ qui est bien définie[89], et est de classe C^∞ sur I. Par contre, elle est dans $\mathcal{D}(I)$ si et seulement si $\int_I \varphi dx = 0$.

— Il existe une fonction $\theta \in \mathcal{D}(I) \setminus \{0\}$ telle que $\int_I \theta dx = 1$[90]. Soit $\varphi \in \mathcal{D}(I)$, alors en notant $J = \int_I \varphi(x)dx$, la fonction $\varphi - J\theta$ est dans $\mathcal{D}(I)$ et est d'intégrale nulle, donc par le point précédent, il existe $\psi \in \mathcal{D}(I)$ telle que $\varphi - J\theta = \psi'$, et donc

$$\varphi = J\theta + \psi'.$$

— On suppose $T' = 0$. Alors avec les notations du point précédent, $\langle T, \psi' \rangle = 0$, et donc

$$\langle T, \varphi \rangle = J \langle T, \theta \rangle = \int_I c\varphi$$

où on a posé $c = \langle T, \theta \rangle$, d'où le résultat.

$\square$

Exercice 3.18 (Existence d'une primitive)**.** Soit $T \in \mathcal{D}'(I)$. On introduit $\theta \in \mathcal{D}(I)$ telle que $\int_I \theta dx = 1$. Pour $\varphi \in \mathcal{D}(I)$, on a vu dans la preuve précédente qu'il existe $\psi \in \mathcal{D}(I)$ telle que $\varphi = \left(\int_I \varphi dx\right) \theta + \psi'$. On pose

$$\langle U, \varphi \rangle := -\langle T, \psi \rangle.$$

1. Montrer que U est une distribution sur I.
2. Montrer que $U' = T$. Ainsi toute distribution admet une primitive.

Exercice 3.19 (Grand classique cité dans le rapport)**.**

1. Soit $f \in L^1_{loc}(I)$, et $a \in I$. On pose

$$\forall x \in I, \quad F(x) = \int_a^x f(t)dt.$$

   (a) Montrer que $[F]$ définit une distribution sur I.

   (b) Montrer que $[F]' = [f]$.[91]

2. Soit $T \in \mathcal{D}'(I)$ telle que $T' \in L^1_{loc}(I)$, c'est-à-dire qu'il existe $f \in L^1_{loc}(I)$ telle que $T' = [f]$.

   (a) Montrer qu'il existe $c \in \mathbb{R}$ telle que[92].

$$T = [F + c] \qquad \text{où} \qquad \forall x \in I, F(x) = \int_a^x f(t)dt.$$

   (b) Montrer que si $f \in C^0(I)$, alors T est représentable par une fonction $C^1(I)$.

---

89. L'intégrale est faussement sur $]-\infty, x]$ puisque φ est à support compact dans I.

90. Pour le justifier, on peut utiliser la fonction classique

$$\rho : x \mapsto \exp\left(\frac{-1}{1-x^2}\right) \mathbb{1}_{]-1,1[}(x)$$

qui est de classe C^∞ sur $\mathbb{R}$, et à support dans $[-1, 1]$. Par homothétie et translation, on peut changer ce support : en effet si $\varepsilon > 0$,

$$x \in \mathbb{R} \mapsto \rho(x/\varepsilon)$$

est à support dans $[-\varepsilon, \varepsilon]$, et si $a \in \mathbb{R}$,

$$x \in \mathbb{R} \mapsto \rho((x-a)/\varepsilon)$$

est à support dans $[a - \varepsilon, a + \varepsilon]$. En choisissant $a \in \mathbb{R}$ et $\varepsilon > 0$ on peut faire en sorte que $[a - \varepsilon, a + \varepsilon] \subset I$, ce qui donne la construction attendue.

91. Ainsi, il est facile de montrer, au sens des distributions, que la dérivée de F est f. La même question au sens classique est beaucoup plus difficile : il s'agit du cas uni-dimensionnel du théorème de différentiation de Lebesgue, qui affirme que F est dérivable presque partout, et que sa dérivée est f presque partout.

92. Ce résultat affirme que si T est une distribution, et que sa dérivée T' est représentable par une fonction (localement intégrable), alors T était en fait elle-même représentable par une fonction (qui a la régularité de F, en l'occurrence au moins continue.)

Remarque 3.20 (*Espaces de Sobolev (à exposants entiers)[93] [94]). Pour $m \in \mathbb{N}$, et I un intervalle de $\mathbb{R}$, on peut poser

$$H^m(I) := \{u \in L^2(I),\ \ \forall k \in [\![0, n]\!], u^{(k)} \in L^2(I)\} \tag{3.7}$$

où on a confondu u et $[u]$, et l'écriture $T \in L^2(I)$ lorsque T est une distribution signifie qu'il existe $g \in L^2(I)$ telle que $T = [g]$.

On peut munir cet espace du produit scalaire

$$\forall (u, v) \in H^m(I),\ \ \langle u, v\rangle_{H^m} = \sum_{k=0}^{m} \int_I u^{(k)}\overline{v^{(k)}}dx.$$

Quelques commentaires sur ces espaces :

1. $(H^m(I), \langle\cdot,\cdot\rangle_{H^m})$ est un espace de Hilbert,
2. si $(m, m') \in \mathbb{N}^2$ avec $m \leq m'$, alors
$$C_c^\infty(I) \subset H^{m'}(I) \subset H^m(I) \subset H^0(I) = L^2(I).$$
3. si $I = \mathbb{R}$ et $m = s \in \mathbb{N}$, la définition (3.7) coïncide avec la définition (2.8) de la remarque 2.34, et les normes associées sont équivalentes.
4. on peut court-circuiter la théorie des distributions pour définir ces espaces : par exemple avec $m = 1$, la définition (3.7) est en fait la même que :
$$H^1(I) = \{u \in L^2(I), \exists g \in L^2(I), \forall \varphi \in C_c^\infty(I),\ \int_I u\varphi' dx = -\int_I g\varphi dx\}$$
et dans ce cas, g est en fait le représentant $L^2(I)$ de la distribution $[u]'$.
Comme on est en dimension 1 et du fait de l'exercice 3.19, on peut aussi définir ces espaces via la notion de primitive : étant donné $x_0 \in I$, on a en fait
$$H^1(I) = \{u \in L^2(I),\ \exists (c, g) \in \mathbb{R} \times L^2(I),\ u(x) = c + \int_{x_0}^{x} g(t)dt\}$$
et encore une fois, g est le représentant $L^2(I)$ de $[u]'$.
5. On peut généraliser au cadre $L^p(I)$ en définissant
$$W^{m,p}(I) := \{u \in L^p(I),\ \ \forall k \in [\![0, n]\!], u^{(k)} \in L^p(I)\}$$
qu'on munit de la norme
$$\forall u \in W^{m,p}(I),\ \ \|u\|_{W^{m,p}} = \sum_{k=0}^{m} \|u^{(k)}\|_p.$$
Les espaces $(W^{m,p}(I), \|\cdot\|_{W^{m,p}})$ sont des espaces de Banach.

93. La définition ici peut être considérée indépendamment de celle de la remarque 2.34 qui reposait sur la transformée de Fourier ; l'avantage cette dernière est qu'on pouvait le faire pour des exposants non-entiers. Le désavantage c'est qu'on se restreignait à l'intervalle $\mathbb{R}$. Un autre avantage de la définition donnée ici est qu'on peut facilement adapter à un cadre $L^p(I)$ pour $p \in [1, +\infty]$.

94. Plus généralement, la question de mesurer les propriétés d'une fonction $f \in L^1_{loc}(I)$ par les propriétés de sa dérivée (au sens des distributions) montre le confort que donne la théorie des distributions. Par exemple, on peut montrer sans aucune hypothèse de régularité que si $f \in L^1_{loc}(I)$, alors

$$f' \geq 0 \quad \Longleftrightarrow \quad f \text{ est croissante.}$$

Ce résultat généralise sans hypothèse de régularité un énoncé bien connu du Chapitre I. Notons qu'il faut expliquer ce que signifie $T \geq 0$ quand T est une distribution :

$$T \geq 0 \text{ dans } \mathcal{D}'(I) \quad \text{signifie} \quad \forall \varphi \in \mathcal{D}(I), \varphi \geq 0 \ \Rightarrow\ \langle T, \varphi\rangle \geq 0.$$

Aussi, quand on a écrit "f est croissante", on aurait dû dire que f admet un représentant qui est une fonction croissante.

6. On peut facilement montrer des cas particuliers des injections de Sobolev ; par exemple, on peut montrer que $H^1(I) \subset C^0(\overline{I})$, au sens où une fonction $u \in H^1(I)$ (qui a priori est une classe de fonctions) admet un représentant continu sur $\overline{I}$ [95]. On peut même montrer avec le théorème d'Ascoli que si I est borné, alors l'injection
$$H^1(I) \hookrightarrow C^0(\overline{I})$$
est compacte, c'est-à-dire que si $(u_n)_{n\in\mathbb{N}}$ est une suite bornée [96] d'éléments de $H^1(I)$, alors on peut extraire une sous-suite qui converge uniformément vers une fonction de $C^0(\overline{I})$.
Plus généralement, dans le cadre des espaces $W^{m,p}(I)$, les théorèmes d'injections de Sobolev posent la question (assez technique en générale) de l'inclusion entre les espaces $W^{m,p}(I)$ et les espaces de type Hölder $C^{n,\alpha}$ où $n \in \mathbb{N}$ et $\alpha \in]0,1]$. Un exemple de résultat accessible à l'agrégation est le suivant :
$$W^{1,\infty}(I) = \{f : I \to \mathbb{R} \text{ bornée et lipschitzienne }\} = C^{0,1}(I).$$
7. On peut également définir,
$$H_0^1(I) := \overline{\mathcal{D}(I)}^{\|\cdot\|_{H^1}} = \{u \in H^1(I),\ \exists(\varphi_n) \in \mathcal{D}(I)^{\mathbb{N}},\ \varphi_n \xrightarrow[n\to+\infty]{\|\cdot\|_{H^1}} u\},$$
qui est un sous-espace vectoriel fermé de $H^1(I)$. On peut montrer que si $I =]a,b[$ avec $(a,b) \in \mathbb{R}^2$, alors
$$H_0^1(I) = \{u \in H^1(I),\ u(a) = u(b) = 0\}$$
où on a choisi pour u son représentant continu [97]. Ce résultat rentre dans la famille des théorèmes de traces [98], et les espaces de ce type sont utiles pour résoudre des problèmes d'équations aux dérivées partielles avec condition de bord de type Dirichlet.

3.3 Autres opérations, convergence dans $\mathcal{D}'(I)$

Faisons quelques commentaires sur les différentes opérations possibles sur $\mathcal{D}'(I)$:

1. On a déjà dit que la somme de deux distributions et la multiplication d'une distribution par un scalaire ne pose pas de problème, ainsi $\mathcal{D}'(I)$ est un espace vectoriel.
2. Afin de bien comprendre l'idée de dualité proposée au paragraphe 3.2 pour la dérivation, entrainons sur un autre exemple, la translation : si $f \in L^1_{loc}(\mathbb{R})$, et $\alpha \in \mathbb{R}$, on a posé $\tau_\alpha f : x \mapsto f(x-\alpha)$. On constate alors que
$$\forall \varphi \in \mathcal{D}(\mathbb{R}),\ \ \langle[\tau_\alpha f], \varphi\rangle = \int_{\mathbb{R}} f(x-\alpha)\varphi(x)dx = \int_{\mathbb{R}} f(x)\varphi(x+\alpha)dx = \langle[f], \tau_{-\alpha}\varphi\rangle.$$
Il est donc naturel de définir, si $T \in \mathcal{D}'(\mathbb{R})$,
$$\langle\tau_\alpha T, \varphi\rangle := \langle T, \tau_{-\alpha}\varphi\rangle$$
et de constater que $\tau_\alpha T \in \mathcal{D}'(\mathbb{R})$.
3. La question de la multiplication de deux distributions est complexe : de façon général, on ne peut pas définir le produit de deux distributions [99]. Néanmoins on peut traiter un cas particulier, à savoir la multiplication d'une distribution par une fonction C^∞. Toujours afin de comprendre le

95. On écrit $\overline{I}$ pour signifier que la fonction peut être prolongée continûment à $\overline{I}$, si ce dernier est différent de I.
96. Bien sûr, on veut dire bornée pour la norme $H^1(I)$, c'est-à-dire qu'il existe $M \in \mathbb{R}$ telle que $\forall n \in \mathbb{N}$, $\|u_n\|_{H^1(I)} \leq M$.
97. Ce qui légitime l'écriture $u(a)$, qui n'aurait pas été permise pour une classe de fonctions.
98. Qui sont beaucoup plus délicats si on remplace I par un ouvert de $\mathbb{R}^n$.
99. Sans rentrer dans les détails, la théorie des distributions a eu un impact très rapide sur la théorie des équations aux dérivées partielles linéaires, le mot linéaire témoignant du fait qu'il n'y avait pas de multiplication intervenant dans l'équation ; l'étude des équations aux dérivées partielles non-linéaires est plus élaborée et nécessite souvent de se restreindre à des espaces ou l'opération de produit est légitime.

principe de dualité, commençons par traiter le cas de la multiplication d'une fonction $f \in L^1_{loc}(I)$ par une fonction $g \in C^\infty(I)$. Alors on peut constater

$$\langle [fg], \varphi \rangle = \int_I f(x)g(x)\varphi(x)dx = \langle [f], g\varphi \rangle.$$

Afin d'avoir $g\varphi \in \mathcal{D}(I)$, a priori on a besoin de $g \in C^\infty(I)$, ce qui explique la restriction que l'on a faite sur g.
Plus généralement donc, si $T \in \mathcal{D}'(I)$ et $g \in C^\infty(I)$, il est naturel de définir la distribution gT par

$$\forall \varphi \in \mathcal{D}(I), \quad \langle gT, \varphi \rangle := \langle T, g\varphi \rangle.$$

Un calcul élémentaire basé sur la formule de Leibniz montre que gT ainsi définie est une distribution.

Exercice 3.21. Montrer que

$$\text{“}x.vp(1/x) = 1\text{''}$$

c'est-à-dire que le produit de la fonction $x \mapsto x$ qui est C^∞ par la distribution $vp(1/x)$ est égale à la distribution associée à la fonction constante égale à 1.

4. La convolution n'est pas une opération simple à traiter non plus ; on reviendra sur des cas particuliers dans la suite.
5. On évoque maintenant la notion de convergence au sens des distributions ; on prend un point de vue de dualité comme à la remarque 2.47 en considérant la convergence suivante :

Définition 3.22. Soit $(T_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{D}'(I)$, et $T \in \mathcal{D}'(I)$. On dit que $(T_n)_{n\in\mathbb{N}}$ converge au sens des distributions vers T si

$$\forall \varphi \in \mathcal{D}(I), \ \ \langle T_n, \varphi \rangle \xrightarrow[n\to+\infty]{} \langle T, \varphi \rangle.$$

On peut noter $T_n \xrightarrow[n\to+\infty]{\mathcal{D}'(I)} T$.

C'est une convergence très faible, facile à satisfaire. A titre d'exemples :

— si $(f_n)_{n\in\mathbb{N}}$ est une suite de fonctions de $L^1(I)$ et $f \in L^1(I)$, alors [100]

$$f_n \xrightarrow[n\to+\infty]{L^1(I)} f \quad \Longrightarrow \quad [f_n] \xrightarrow[n\to+\infty]{\mathcal{D}'(I)} [f].$$

— si $(\rho_n)_{n\in\mathbb{N}}$ est une identité approchée, alors

$$\rho_n \xrightarrow[n\to+\infty]{\mathcal{D}'(I)} \delta_0.$$

— on peut aussi traiter des passages à la limite originaux :

Exercice 3.23. Montrer qu'au sens des distributions on a

$$\frac{1}{x \pm i\varepsilon} \xrightarrow[\varepsilon\to 0^+]{\mathcal{D}'(\mathbb{R})} vp(1/x) \mp i\pi\delta_0.$$

— la convergence au sens des distributions étant très flexibles, on peut par exemple se convaincre facilement que l'opération de dérivation sur $\mathcal{D}'(I)$ est continue, c'est-à-dire que

$$T_n \xrightarrow[n\to+\infty]{\mathcal{D}'(I)} T \quad \Longrightarrow \quad T_n' \xrightarrow[n\to+\infty]{\mathcal{D}'(I)} T'.$$

100. *On peut même remplacer $L^1(I)$ par $L^1_{loc}(I)$ dans cet énoncé : la convergence au sens $L^1_{loc}(I)$ signifie la convergence dans $L^1(K)$ pour tout K compact inclus dans I.

Comparez avec la proposition 4.13 au Chapitre I ; ici aucun besoin d'hypothèse. En conséquence, on a aussi que si $(T_n)_{n\in\mathbb{N}}$ est une suite de distributions telle que $\sum_{n\in\mathbb{N}} T_n$ converge au sens des distributions, alors

$$\left(\sum_{n\in\mathbb{N}} T_n\right)' = \sum_{n\in\mathbb{N}} T_n'$$

encore une fois sans condition.

6. Support d'une distribution :

Définition 3.24. Soit $T \in \mathcal{D}'(I)$. On pose

$$\omega_0 := \bigcup_{\omega\in\mathcal{O}} \omega, \text{ où } \mathcal{O} = \{\omega \text{ ouvert }, T_{|\omega} = 0, \text{ i.e. } \forall\varphi \in \mathcal{D}(\omega), \langle T, \varphi\rangle = 0\}$$

et on appelle ω_0 le plus grand ouvert de nullité de T. On pose alors

$$\mathrm{Supp}(T) := I \setminus \omega_0 \quad \text{le support de } T.$$

— un classique est l'étude des distributions dont le support est restreint à un point : on peut montrer que si $T \in \mathcal{D}'(I)$ et $\mathrm{Supp}(T) = \{x_0\}$, alors il existe $n \in \mathbb{N}$ et $(a_0, \dots a_n) \in \mathbb{N}^{n+1}$ tel que

$$T = \sum_{k=0}^{n} a_k \delta_{x_0}^{(k)}.$$

Voir par exemple [Zui02, Théorème 2.4.5].

— on peut également étudier les distributions à support compact, et voir que l'ensemble des distributions à support compact dans I s'assimile au dual de $C^\infty(I)$ (rappelons que $C^\infty(I)$ est un espace métrique, voir le chapitre III).

3.4 Distributions tempérées

3.4.1 Définition

Définition 3.25. Soit $T \in \mathcal{D}'(\mathbb{R})$. On dit que T est une distribution tempérée et on note $T \in \mathscr{S}'(\mathbb{R})$ s'il existe $(k_0, p_0) \in \mathbb{N}^2$ et $C \in \mathbb{R}$ tel que

$$\forall\varphi \in \mathcal{D}(\mathbb{R}), \ \ |\langle T, \varphi\rangle| \le C \sum_{k\le k_0, p\le p_0} N_{k,p}(\varphi). \tag{3.8}$$

Remarque 3.26. On peut donner une définition équivalente, qui ne nécessite pas de connaître $\mathcal{D}'(I)$: on peut en effet dire que $T : \mathscr{S}(\mathbb{R}) \to \mathbb{K}$ est une distribution tempérée si elle est linéaire, et s'il existe $C \in \mathbb{R}$ et $(k_0, p_0) \in \mathbb{N}^2$ tels que

$$\forall\varphi \in \mathscr{S}(I), \ \ |\langle T, \varphi\rangle| \le C \sum_{k\le k_0, p\le p_0} N_{k,p}(\varphi). \tag{3.9}$$

La seule différence étant les ensembles de définitions ($\mathcal{D}(\mathbb{R})$ dans un cas, $\mathscr{S}(\mathbb{R})$ dans l'autre). Pour se convaincre que ces définitions sont équivalentes, il faut constater :

— si $T \in \mathcal{D}'(I)$ et satisfait (3.8), alors par densité de $\mathcal{D}(\mathbb{R})$ dans $\mathscr{S}(\mathbb{R})$ [101] on peut prolonger T en une forme linéaire sur $\mathscr{S}(\mathbb{R})$ qui satisfera (3.9).

— si $T : \mathscr{S}(\mathbb{R}) \to \mathbb{K}$ est linéaire continue et satisfait (3.9), alors sa restriction à $\mathcal{D}(\mathbb{R})$ est une distribution, c'est-à-dire qu'elle satisfait (3.2).

Ainsi, $\mathscr{S}(\mathbb{R})$ étant un espace métrique, on peut vraiment dire que $\mathscr{S}'(\mathbb{R})$ est l'espace dual de $\mathscr{S}(\mathbb{R})$.

101. Cette expression nécessite de manipuler la métrique de $\mathscr{S}(\mathbb{R})$.

Exemple 3.27. Attention, contrairement à $\mathcal{D}'(\mathbb{R})$, les fonctions localement intégrables ne définissent pas forcément des distributions. Il faut imposer une condition de "croissance" raisonnable à l'infini. Plus précisément, si $f \in L^1_{loc}(\mathbb{R})$ est telle qu'il existe $N \in \mathbb{N}$ et $C \in \mathbb{R}$ telle que

$$|f(x)| \leq C(1+|x|^N) \quad \text{p.p.},$$

alors pour tout $\varphi \in \mathscr{S}(\mathbb{R})$

$$\begin{aligned}
\int_{\mathbb{R}} |f(x)\varphi(x)|\,dx &\leq \int_{\mathbb{R}} C(1+|x|^N)|\varphi(x)|dx = \int_{\mathbb{R}} C\frac{(1+|x|^2)(1+|x|^N)}{1+x^2}|\varphi(x)|dx \\
&\leq \int_{\mathbb{R}} C'\frac{(1+|x|^{N+2})}{1+x^2}|\varphi(x)|dx \\
&\leq C''\big(\|\varphi\|_\infty + \sup_{x\in\mathbb{R}} |x^{N+2}\varphi(x)|\big)
\end{aligned}$$

où $C', C'' \in \mathbb{R}^2$. Ainsi $\int_{\mathbb{R}} f\varphi dx$ est bien définie et $[f] \in \mathscr{S}'(\mathbb{R})$.

Exercice 3.28. Il est donc intéressant de trouver une distribution (même une fonction) qui n'est pas une distribution tempérée : montrez que la fonction $x \in \mathbb{R} \mapsto e^{x^2}$ n'est pas une distribution tempérée.

Remarque 3.29 (*). On peut vérifier que les distributions à support compact sont des distributions tempérées. Formellement, on a en fait les inclusions

$$\mathcal{D}(\mathbb{R}) \subset \mathscr{S}(\mathbb{R}) \subset C^\infty(\mathbb{R})$$

qui donnent

$$(C^\infty(\mathbb{R}))' \subset \mathscr{S}'(\mathbb{R}) \subset \mathcal{D}'(\mathbb{R}).$$

Exercice 3.30. Montrer que la distribution $vp(1/x)$ est une distribution tempérée.

Exercice 3.31 (Produit d'une distribution tempérée par une fonction à croissance lente)**.** Soit $f \in C^\infty(\mathbb{R})$ telle que pour tout $p \in \mathbb{N}$, il existe $C_p \in \mathbb{R}$ et $m_p \in \mathbb{N}$ tels que

$$\forall x \in \mathbb{R}, \ \ |f^{(p)}(x)| \leq C_p\,(1+|x|)^{m_p}\,.$$

1. Montrer que si $\varphi \in \mathscr{S}(\mathbb{R})$ alors $f\varphi \in \mathscr{S}(\mathbb{R})$.
2. Montrer que si $T \in \mathscr{S}'(\mathbb{R})$ alors $fT \in \mathscr{S}'(\mathbb{R})$.

3.4.2 Transformation de Fourier sur $\mathscr{S}'(\mathbb{R})$

On applique le principe de dualité évoqué aux paragraphes 3.2 et 3.3 : si $f \in L^1(\mathbb{R})$ et $\varphi \in \mathscr{S}(\mathbb{R})$, alors le théorème de Fubini donne

$$\int_{\mathbb{R}} \widehat{f}\varphi dx = \int_{\mathbb{R}} \varphi(\xi)\int_{\mathbb{R}} f(x)e^{-i\xi x}\frac{dx}{\sqrt{2\pi}}d\xi = \int_{\mathbb{R}} f(x)\int_{\mathbb{R}} \varphi(\xi)e^{-i\xi x}\frac{d\xi}{\sqrt{2\pi}}dx = \int_{\mathbb{R}} f\widehat{\varphi}dx.$$

Définition-Proposition 3.32. Soit $T \in \mathscr{S}'(\mathbb{R})$. Alors on peut définir $\mathscr{F}(T) = \widehat{T}$ par

$$\forall \varphi \in \mathscr{S}(\mathbb{R}), \ \ \langle \widehat{T}, \varphi\rangle := \langle T, \widehat{\varphi}\rangle$$

et alors $\widehat{T} \in \mathscr{S}'(\mathbb{R})$.

Démonstration. Soit $T \in \mathscr{S}'(\mathbb{R})$. Par la remarque 3.26, T est définie [102] sur $\mathscr{S}(\mathbb{R})$, et par le théorème 2.39, si $\varphi \in \mathscr{S}(\mathbb{R})$ alors $\widehat{\varphi} \in \mathscr{S}(\mathbb{R})$, donc $\widehat{T}$ est bien définie. De plus, comme $T \in \mathscr{S}'(\mathbb{R})$, il existe $(k_0, p_0) \in \mathbb{N}^2$ et $C \in \mathbb{R}$ tels que

$$\forall \varphi \in \mathscr{S}(\mathbb{R}), \ \ |\langle T, \varphi\rangle| \leq C\sum_{k\leq k_0, p\leq p_0} N_{k,p}(\varphi).$$

102. (ou prolongeable par continuité sur)

En conséquence,

$$\begin{aligned}\forall \varphi \in \mathscr{S}(\mathbb{R}),\ |\langle \widehat{T}, \varphi\rangle| &= |\langle T, \widehat{\varphi}\rangle| \leq C \sum_{k\leq k_0, p\leq p_0} N_{k,p}(\widehat{\varphi}),\\ &\leq C \sum_{k\leq k_0, p\leq p_0} \|m^k(\widehat{\varphi})^{(p)}\|_\infty = C \sum_{k\leq k_0, p\leq p_0} \|\widehat{(m^p\varphi)^{(k)}}\|_\infty\\ &\leq C \sum_{k\leq k_0, p\leq p_0} \|(m^p\varphi)^{(k)}\|_1 \leq C' \sum_{k\leq k_0, p\leq p_0} \|m^p\varphi^{(k)}\|_1 \text{ par Leibnitz}\\ &\leq C'' \sum_{k\leq k_0+2, p\leq p_0} \|m^p\varphi^{(k)}\|_\infty\end{aligned}$$

où on a utilisé pour la dernière inégalité

$$\int_{\mathbb{R}} |m^p\varphi^{(k)}|dx = \int_{\mathbb{R}} \frac{1+m^2}{1+m^2}|m^p\varphi^{(k)}|dx \leq \int_{\mathbb{R}} \frac{1}{1+|x|^2}dx \|(1+m^2)m^p\varphi^{(k)}\|_\infty.$$

Ainsi $\widehat{T} \in \mathscr{S}'(\mathbb{R})$. □

Théorème 3.33. *L'opérateur*

$$\mathscr{F} : \mathscr{S}'(\mathbb{R}) \to \mathscr{S}'(\mathbb{R})$$

est une bijection, dont l'inverse est

$$\begin{array}{rccl}\mathscr{G} : & \mathscr{S}'(\mathbb{R}) & \longrightarrow & \mathscr{S}'(\mathbb{R})\\ & T & \longmapsto & \mathscr{G}(T) : \Big[\varphi \mapsto \langle T, \mathscr{G}(\varphi)\rangle\Big].\end{array}$$

Démonstration. Le fait que $\mathscr{G}$ soit bien définie sur $\mathscr{S}'(\mathbb{R})$ se montre de façon similaire à la preuve qui précède. On a donc

$$\forall \varphi \in \mathscr{S}(\mathbb{R}), \quad \langle \mathscr{F} \circ \mathscr{G}(T), \varphi\rangle = \langle T, \mathscr{G} \circ \mathscr{F}(\varphi)\rangle = \langle T, \varphi\rangle$$

et donc $\mathscr{F} \circ \mathscr{G} = Id_{\mathscr{S}'(\mathbb{R})}$. De même pour $\mathscr{G} \circ \mathscr{F}$, et donc $\mathscr{F}$ est une bijection. □

Remarque 3.34 (Comparaison avec les cadres $L^1(\mathbb{R}), L^2(\mathbb{R}), \mathcal{M}_f^+(\mathbb{R})$)**.** Remarquons en premier lieu que

$$L^1(\mathbb{R}) \subset \mathscr{S}'(\mathbb{R}), \qquad \text{et} \qquad L^2(\mathbb{R}) \subset \mathscr{S}'(\mathbb{R}), \qquad \text{et} \qquad \mathcal{M}_f^+(\mathbb{R}) \subset \mathscr{S}'(\mathbb{R}).$$

En effet si $f \in L^1(\mathbb{R})$,

$$\forall \varphi \in \mathscr{S}(\mathbb{R}), \quad \left|\int_{\mathbb{R}} f\varphi dx\right| \leq \|f\|_1\|\varphi\|_\infty = \|f\|_1 N_{0,0}(\varphi,)$$

si $g \in L^2(\mathbb{R})$,

$$\forall \varphi \in \mathscr{S}(\mathbb{R}), \quad \left|\int_{\mathbb{R}} g\varphi dx\right| \leq \|g\|_2\|\varphi\|_2 = \|g\|_2\sqrt{\int_{\mathbb{R}} \frac{dx}{1+x^2}}\,(N_{0,0}(\varphi) + N_{1,0}(\varphi))$$

et si $\mu \in \mathcal{M}_f^+(\mathbb{R})$, alors

$$\forall \varphi \in \mathscr{S}(\mathbb{R}), \quad \left|\int_{\mathbb{R}} \varphi d\mu\right| \leq \mu(\mathbb{R})\|\varphi\|_\infty.$$

Le cadre de $\mathscr{F} : \mathscr{S}'(\mathbb{R}) \to \mathscr{S}'(\mathbb{R})$ généralise en fait les trois cadres abordés au paragraphe 2. Pour le constater, on montre :

Proposition 3.35. *Soit $T \in \mathscr{S}'(\mathbb{R})$ et $f \in L^1(\mathbb{R})$ telle que $T = [f]$. Alors $\widehat{T} = \big[\widehat{f}\big]$.*

Démonstration. Soit $\varphi \in \mathscr{S}(\mathbb{R})$. Alors

$$\langle \widehat{T}, \varphi\rangle = \int_{\mathbb{R}} f(x)\widehat{\varphi}(x)dx = \int_{\mathbb{R}} \varphi(\xi)\int_{\mathbb{R}} f(x)e^{-i\xi x}\frac{dx}{\sqrt{2\pi}}d\xi = \langle \big[\widehat{f}\big], \varphi\rangle,$$

où on a appliqué le théorème de Fubini. □

Exemple 3.36. On peut se convaincre que $[1] \in \mathscr{S}'(\mathbb{R})$ car

$$\langle [1], \varphi \rangle = \int_{\mathbb{R}} \varphi dx \leq \int_{\mathbb{R}} \frac{1}{1+x^2} dx \left(\|\varphi\|_\infty + \|m^2 \varphi\|_\infty \right).$$

De plus

$$\forall \varphi \in \mathscr{S}(\mathbb{R}), \qquad \langle \widehat{1}, \varphi \rangle = \langle 1, \widehat{\varphi} \rangle = \int_{\mathbb{R}} \widehat{\varphi} dx = \sqrt{2\pi} \varphi(0) = \sqrt{2\pi} \langle \delta_0, \varphi \rangle$$

où on a appliqué la formule d'inversion pour $\varphi \in \mathscr{S}(\mathbb{R})$. En confondant 1 et [1], on écrit généralement

$$\widehat{1} = \sqrt{2\pi} \delta_0.$$

Remarque 3.37. Les propriétés de $\mathscr{F}$ sur $\mathscr{S}(\mathbb{R})$ vue au paragraphe 2.3 se transposent sur $\mathscr{S}'(\mathbb{R})$: par exemple si $T \in \mathscr{S}'(\mathbb{R})$, alors

$$\mathscr{F}(T') = i\xi \mathscr{F}(T), \qquad \text{i.e.} \quad \widehat{T'} = i\xi \widehat{T}, \qquad \text{et de même} \quad (\widehat{T})' = -i\widehat{mT}$$

ce qui se démontre simplement en utilisant la même formule dans $\mathscr{S}(\mathbb{R})$. Notez qu'on utilise dans les formules précédentes un cas particulier de l'exercice 3.31 pour voir que le produit mT est encore dans $\mathscr{S}'(\mathbb{R})$.

Exercice 3.38 (Transformation de Fourier de $vp(1/x)$)**.** On pose $T = vp(1/x)$, et $\forall x \in \mathbb{R},\ m(x) = x$

1. À partir de l'équation $mT = 1$, calculer $\widehat{T}'$.
2. Montrer que $\widehat{T}$ est impaire, et en déduire

$$\widehat{T} = -i\sqrt{2\pi} H + i\frac{\sqrt{2\pi}}{2}.$$

Exercice 3.39 (Reprise de l'exercice 2.22 dans $\mathscr{S}'(\mathbb{R})$)**.** Soit $c \in \mathbb{R}$.

1. Soit $T \in \mathscr{S}'(\mathbb{R})$ telle que
$$T' = \tau_c(T).$$
2. Montrer que si $c \not\equiv \pi/2 \bmod 2\pi$, alors $T = 0$.

3. *On suppose maintenant $c \equiv \pi/2 \bmod 2\pi$.*
   (a) Montrer que $\operatorname{Supp}(\widehat{T}) \subset \{\pm 1\}$.
   (b) Écrire $\delta_a^{(k)}$ comme la transformée de Fourier d'une fonction, pour $k \in\in \mathbb{N}$ et $a \in \mathbb{R}$.
   (c) Avec le résultat sur les distributions ayant pour support un point (voir page 449), en déduire l'ensemble des distributions tempérées T telles que $T' = \tau_c(T)$, sous l'hypothèse $c \equiv \pi/2 \bmod 2\pi$.
   (d) Dans le cas $c = 0$, comparez et expliquez la différence avec l'ensemble des fonctions $f \in C^\infty(\mathbb{R})$ telles que $f' = f$.

3.5 Sujets de travail

Exercice 3.40 (Formule de Poisson par les distributions)**.** Voir [Zui02, Problème 5 page 191].

1. Soit $(a_n)_{n\in\mathbb{Z}} \in \mathbb{C}^{\mathbb{Z}}$ telle qu'il existe $p \in \mathbb{N}$ tel que $a_n = O(|n|^p)$. Montrer que

$$x \mapsto \sum_{n\in\mathbb{Z}} a_n e^{inx}$$

converge dans $\mathcal{D}'(\mathbb{R})$.

2. On pose
$$\forall x \in [0, 2\pi[, \quad f(x) = \frac{1}{2}(\pi - x)$$
qu'on prolonge à $\mathbb{R}$ en une fonction $f : \mathbb{R} \to \mathbb{R}$ 2π-périodique. Montrer que
$$\forall x \in \mathbb{R}, \quad f(x) = \sum_{n \geq 1} \frac{\sin(nx)}{n}$$
où la convergence a lieu au sens $L^2(\mathbb{R})$.
3. Montrer que la convergence dans $\mathcal{D}'(\mathbb{R})$ et en déduire par dérivation que
$$2\pi \sum_{n \in \mathbb{Z}} \delta_{2n\pi} = \sum_{n \in \mathbb{Z}} e_n$$
où on a posé $e_n : x \in \mathbb{R} \mapsto e^{inx}$ et où les convergences ont lieu dans $\mathcal{D}'(\mathbb{R})$.
4. En déduire que si $\varphi \in \mathcal{D}(\mathbb{R})$ alors
$$\sum_{n \in \mathbb{Z}} \widehat{\varphi}(n) = 2\pi \sum_{n \in \mathbb{Z}} \varphi(2\pi n).$$

Convolution : comme on l'a remarqué précédemment, la convolution n'est pas un sujet facile à définir dans le cadre des distributions : néanmoins, on peut trouver 3 cadres dans lesquelles on peut effectivement définir la convolution :

1. on peut définir la convolution d'une distribution T à support compact avec une fonction $\varphi \in C^\infty(\mathbb{R})$ (ou alors d'une distribution T avec une fonction $\varphi \in \mathcal{D}(\mathbb{R})$) comme la fonction
$$\forall x \in \mathbb{R}, \quad T * \varphi(x) = \langle T, \varphi(x - \cdot) \rangle,$$
qui est de classe C^∞.
2. plus généralement on peut définir la convolution d'une distribution T et d'une distribution à support compact S, qui est alors une distribution ; cela nécessite la notion de produit tensoriel de distributions. Remarquons dans ce cas que δ_0 (qui est une distribution à support compact) agit comme élément neutre, i.e.
$$\forall T \in \mathcal{D}'(\mathbb{R}), \; T * \delta_0 = T.$$
3. on peut affaible la condition que l'un des objets convolés soit à support compact par le fait que les supports des deux objets soient "convolutifs"[103]

Les opérations usuelles de la convolution (la plus importante étant le fait que la dérivée de $T * S$ est $T' * S$ ou encore $T * S'$) peuvent s'adapter à ces cadres. Dans le cadre du programme, il n'est pas attendu de connaissance particulière sur la convolution des distributions. Néanmoins, si vous souhaitez l'utiliser (voir le point suivant), il faut savoir décrire des situations simples.
Solution fondamentale Dans $\mathbb{R}^n$ [104], on appelle
$$T \in \mathcal{D}'(\mathbb{R}^n) \longmapsto \sum_{|\alpha| \leq d} a_\alpha \partial_\alpha T$$
(où $\alpha \in \mathbb{N}^n$, $|\alpha| = \sum_{i=1}^n \alpha_i$ et $\partial_\alpha = \frac{\partial^{|\alpha|}}{\partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}}$) un opérateur différentiel, qu'on peut noter $P(\partial)$ par similitude avec la notation polynomiale.

103. Deux ensembles A et B de $\mathbb{R}^n$ sont dits convolutifs si pour tout $R \in \mathbb{R}$,il existe $\rho(R) \in \mathbb{R}$ tel que
$$\forall (x, y) \in A \times B, \; |x + y| \leq R \implies |x| \leq \rho(R) \text{ et } |y| \leq \rho(R).$$
Par exemple $\mathbb{R}_+$ et $\mathbb{R}_+$ sont convolutifs car si $x + y$ est majoré avec (x, y) positifs, alors x et y sont majorés. Mais $\mathbb{R}_+$ et $\mathbb{R}_-$ ne sont pas convolutifs puisque $x + (-x)$ est borné sans que $(x, -x) \in \mathbb{R}_+ \times \mathbb{R}_-$ ne le soient.
104. On se place désormais dans les distributions sur $\mathbb{R}^n$; tout ce qui précède est facilement adaptable à ce cadre ; on peut néanmoins considérer $n = 1$ par simplicité.

Une solution élémentaire de l'opérateur $P(\partial)$ est une distribution E telle que

$$P(\partial)E = \delta_0.$$

L'utilisation de ce concept est le fait que si $S \in \mathcal{D}'(\mathbb{R})$, alors la solution de l'équation

$$P(\partial)T = S$$

sera $E * S$ si cette convolution est bien définie. En effet, les règles d'opérations sur la convolution donnent

$$P(\partial)(E * S) = (P(\partial)E) * S = \delta_0 * S = S.$$

Exemple 3.41. Dans $\mathbb{R}$, l'unique solution fondamentale de $y \mapsto -y'' + y$ qui est dans $\mathscr{S}'(\mathbb{R})$ est

$$\forall x \in \mathbb{R}, \quad E(x) = \frac{1}{2} e^{-|x|},$$

ce qu'on peut justifier par un calcul simple via la transformation de Fourier dans $\mathscr{S}'(\mathbb{R})$ (voir aussi l'exemple 2.20).

Proposition 3.42 (Solution fondamentale du Laplacien)**.** *On pose*[105]

$$\forall x \in \mathbb{R}^N, \quad \begin{cases} E_1(x) = -\dfrac{1}{2}|x| \\ E_2(x) = -\dfrac{1}{2\pi} \ln |x| \\ E_N(x) = \dfrac{1}{(N-2)|\mathbb{S}^{N-1}|} \dfrac{1}{|x|^{N-2}} & \textit{si } N \geq 3 \end{cases}$$

Alors pour tout $N \in \mathbb{N}^*$ on a

$$-\Delta E_N = \delta_0.$$

La preuve est élémentaire pour $N = 1$. Les cas $N \geq 2$ et $N \geq 3$ peuvent se montrer par la formule de Green[106]. Notez que les cas $N = 1$ et $N = 2$ sont aux cœurs de la partie II de l'épreuve écrite d'Analyse & Probabilités de 2019.

105. Ici $|\mathbb{S}^{N-1}|$ désigne l'hypersurface de $\mathbb{S}^{N-1}$ la sphère unité de $\mathbb{R}^N$.

106. Il existe plusieurs formules de Green ; citons la formule fondamentale (appelée formule de la divergence, formule de Stokes, parfois formule de Green)

$$\int_\Omega \operatorname{div} X = \int_{\partial\Omega} X \cdot \nu d\sigma$$

où $X : \mathbb{R}^n \to \mathbb{R}^n$ est de classe C^1, $\operatorname{div} X = \sum_{i=1}^n \partial_i X_i$, Ω est un ouvert régulier de $\mathbb{R}^n$, ν est le vecteur normal unitaire extérieur sur $\partial\Omega$, et $d\sigma$ est la mesure surfacique. Voir par exemple [Bon06, Page 108].

En conséquence on obtient facilement la généralisation suivante : en écrivant $\operatorname{div}(X\varphi)$ où $\varphi : \mathbb{R}^n \to \mathbb{R}$ est de classe C^1, on obtient

$$\int_\Omega \operatorname{div}(X)\varphi dx = -\int_\Omega X \cdot \nabla\varphi dx + \int_{\partial\Omega} \varphi X \cdot \nu d\sigma$$

ce qui donne, si on particularise à $X = \nabla u$:

$$\int_\Omega (\Delta u)\varphi dx = -\int_\Omega \nabla u \cdot \nabla\varphi dx + \int_{\partial\Omega} \varphi \partial_\nu u d\sigma$$

où $\partial_\nu u = \nabla u \cdot \nu$. En échangeant les rôles de u et φ et en sommant, on obtient également la formule suivante, souvent appelée formule de Green :

$$\int_\Omega (\Delta u) v dx = \int_\Omega u(\Delta\varphi) dx + \int_{\partial\Omega} [(\partial_\nu u)\varphi - u\partial_\nu\varphi] \, d\sigma.$$

Chapitre IX

Topologie 2 et Analyse Fonctionnelle

Dans le Chapitre III, nous avons défini un cadre général d'espaces dans lesquels on peut faire de l'analyse (les espaces métriques, en particulier les espaces vectoriels normés), et afin de faire une première digression vers l'analyse fonctionnelle, nous avons étudié en profondeur l'espace des fonctions continues sur un segment muni de la norme uniforme, qui était à ce moment l'exemple le plus riche. Mais au Chapitre IV, nous avons construit et étudié les espaces $L^p(I)$ ($p \in [1, +\infty]$) issus de la théorie de l'intégrale de Lebesgue, espaces que nous avons abondamment utilisés au Chapitre VIII.

Il est donc temps de rajouter une couche théorique pour prendre un peu de recul sur les notions abordées jusqu'ici, et c'est l'objet de ce chapitre. On peut le découper en deux parties :

1. une première partie qui correspond aux deux premières sections de ce chapitre, à savoir l'étude abstraite des espaces de Hilbert, et l'étude particulière des espaces de Sobolev H^1 et H^1_0 sur un intervalle de $\mathbb{R}$, et leurs applications à la résolution d'EDP (équations aux dérivées partielles) elliptiques en dimension 1,
2. une seconde partie constituée de compléments qui ne sont pas directement inscrits dans le programme [1], mais qui sont des sujets classiques et évoqués à plusieurs reprises dans le rapport de jury.

L'analyse hilbertienne, et plus généralement les thématiques de l'analyse fonctionnelle apparaissent très régulièrement dans les sujets d'écrits (voir par exemple les sujets 2005, 2009, 2011, 2013, 2015, 2018, 2020, 2021). La résolution d'EDP est parfois en filigrane comme fil rouge de certains sujets (comme 2000, 2010, 2019) ; quant aux sujets abordés dans la seconde partie de ce chapitre, et qui sont a priori plus élaborés, on les retrouve quand même dans les sujets d'écrits, même si ceux-ci sont parfois un peu plus anciens (voir 1999, 2009, 2011, 2012).

Pour l'oral, ce qui sera abordé dans ce chapitre est naturellement incontournable pour la leçon **213**, mais on pourra s'en servir dans la plupart des autres leçons, notamment **201, 203, 205, 206, 208, 234, 246, 250, 253**.

Une référence très classique est le livre de H. Brézis [Bré05] ; j'utilise aussi fréquemment [HL99], le classique [Rud98] (Chapitres 4 et 5 principalement) et [Gou98, Annexe A] restent des valeurs sûres. Les livres [Wag12b] et [Rud95] sont également très riches. Pour l'aspect EDP on peut également citer [Zui02], quant à [QZ13, Chapitres VI et VIII] il contient des éléments bien utiles sur les espaces de fonctions et les espaces de Banach. Dans tous les cas, le lecteur doit garder à l'esprit que tous ces livres vont significativement plus loin que le programme de l'agrégation, et il doit choisir ce qui lui convient le mieux dans ces livres.

1 Espaces de Hilbert

Dans cette section nous allons étudier un cas particulier des espaces de Banach pour lesquels, en plus de la possibilité de mesurer une distance entre les objets (grâce à la norme), nous avons une notion d'angle (en particulier d'orthogonalité) qui permet de mélanger des raisonnements "géométriques" et "analytiques". Ces espaces ont une importance majeure en analyse, et en particulier dans le programme

1. Mais ils l'ont été à un moment donné

de l'agrégation, où certains sujets ne sont abordés que dans ce cadre du fait de leur simplicité (la dualité, la notion de projection, etc).

1.1 Produit scalaire, orthogonalité

Dans ce qui suit, on va considérer des espaces vectoriels, et la présentation diffère légèrement selon que le corps de base est $\mathbb{R}$ ou $\mathbb{C}$ [2]. On notera donc $\mathbb{K}$ le corps de base qui désignera $\mathbb{R}$ ou $\mathbb{C}$.

Définition 1.1. Soit H un $\mathbb{K}$-espace vectoriel.

— Si $\mathbb{K} = \mathbb{R}$, on appelle produit scalaire $\langle\cdot,\cdot\rangle : H \times H \to \mathbb{R}$ une forme bilinéaire symétrique définie positive, c'est-à-dire vérifiant :
  i) $\forall\lambda \in \mathbb{R}, \forall(x, y, z) \in H^3, \ \ \langle\lambda x + y, z\rangle = \lambda\langle x, z\rangle + \langle y, z\rangle$,
  ii) $\forall(x, y) \in H^2, \ \ \langle x, y\rangle = \langle y, x\rangle$,
  iii) $\forall x \in H, \ \ \langle x, x\rangle \geq 0, \quad \text{et} \quad \langle x, x\rangle = 0 \ \Leftrightarrow \ x = 0$.

— Si $\mathbb{K} = \mathbb{C}$, on appelle produit scalaire $\langle\cdot,\cdot\rangle : H \times H \to \mathbb{R}$ une forme sesquilinéaire hermitienne définie positive, c'est-à-dire vérifiant :
  i) $\forall\lambda \in \mathbb{C}, \forall(x, y, z) \in H^3, \ \ \langle\lambda x + y, z\rangle = \lambda\langle x, z\rangle + \langle y, z\rangle$,
  ii) $\forall(x, y) \in H^2, \ \ \langle x, y\rangle = \overline{\langle y, x\rangle}$,
  iii) $\forall x \in H, \ \ \langle x, x\rangle \in \mathbb{R}_+, \quad \text{et} \quad \langle x, x\rangle = 0 \ \Leftrightarrow \ x = 0$.

On appelle espace préhilbertien la donnée du couple $(H, \langle\cdot,\cdot\rangle)$ où $\langle\cdot,\cdot\rangle$ est un produit scalaire sur H.

Remarque 1.2. Dans le cas $\mathbb{K} = \mathbb{C}$, l'application $\langle\cdot,\cdot\rangle$ n'est pas bilinéaire mais sesquilinéaire [3] : elle est linéaire en la première variable, et du fait de sa symétrie hermitienne (propriété ii)), elle satisfait

$$\forall\lambda \in \mathbb{C}, \forall(x, y, z) \in H^3, \ \ \langle x, \lambda y + z\rangle = \overline{\lambda}\langle x, y\rangle + \langle x, z\rangle,$$

on dit qu'elle est antilinéaire (ou semi-linéaire) en la seconde variable.

Le fait d'avoir choisi une linéarité "à gauche" et une antilinéarité "à droite" est purement arbitraire, et on pourra trouver l'autre convention (linéarité à droite, antilinéarité à gauche) dans la littérature.

Proposition 1.3. *Soit $(H, \langle\cdot,\cdot\rangle)$ un espace préhilbertien. Alors*

1. *L'application $x \in U \mapsto \sqrt{\langle x, x\rangle}$ est une norme sur H. Sauf précision contraire, on travaillera toujours avec cette norme (et la distance associée) et on la note $\|\cdot\|$; l'espace $(H, \|\cdot\|)$ est donc un espace vectoriel normé.*
2. *Inégalité de Cauchy-Schwarz :*
$$\forall(x, y) \in H^2, \quad |\langle x, y\rangle| \leq \|x\|.\|y\|, \tag{1.1}$$
et il y a égalité dans l'inégalité précédente si et seulement si la famille (x, y) est liée.
3. *Développement d'un carré [4] : Si $\mathbb{K} = \mathbb{R}$,*
$$\forall(x, y) \in H^2, \quad \|x + y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2. \tag{1.2}$$
et si $\mathbb{K} = \mathbb{C}$,
$$\forall(x, y) \in H^2, \quad \|x + y\|^2 = \|x\|^2 + 2\mathrm{Re}(\langle x, y\rangle) + \|y\|^2. \tag{1.3}$$
4. *Identité du parallélogramme :*
$$\forall(x, y) \in H^2, \quad \left\|\frac{x + y}{2}\right\|^2 + \left\|\frac{x - y}{2}\right\|^2 = \frac{1}{2}\left(\|x\|^2 + \|y\|^2\right). \tag{1.4}$$

2. Néanmoins, pour ne pas tout répéter deux fois, il arrivera qu'on donne certaines formules et certains calculs adaptés au cas $\mathbb{K} = \mathbb{C}$; ils seront en fait valables aussi dans le cas réel, simplement on aura des conjugués ou des parties réelles "inutiles".

3. Le préfixe "sesqui" signifie "dans un rapport de un et demi", donc on voit cette application comme "un et demi"-linéaire.

4. On peut rapprocher cette formule du théorème d'Al Kashi qui généralise le théorème de Pythagore

Remarque 1.4. On retiendra de cet énoncé, ainsi que de sa preuve, qu'il est toujours plus commode de travailler avec $\|\cdot\|^2$ plutôt que $\|\cdot\|$. C'est quelque chose à garder en tête quand on est dans un espace préhilbertien.

Remarque 1.5. Les formules (1.2) et (1.3) permettent de faire le "chemin inverse" en exprimant le produit scalaire à partir de la norme : en effet on en déduit si $\mathbb{K}=\mathbb{R}$,

$$\forall (x,y)\in H^2,\quad \langle x,y\rangle = \frac{\|x+y\|^2-\|x\|^2-\|y\|^2}{2}.$$

On peut aussi écrire une variante avec un peu plus de symétrie :

$$\forall (x,y)\in H^2,\quad \langle x,y\rangle = \frac{\|x+y\|^2-\|x-y\|^2}{4}.$$

Dans le cas où $\mathbb{K}=\mathbb{C}$, on peut retrouver de la même façon $\mathrm{Re}(\langle x,y\rangle)$, et en jouant sur $\mathrm{Im}(\langle x,y\rangle)=\mathrm{Re}(\langle -ix,y\rangle)$, on obtient

$$\forall (x,y)\in H^2,\quad \langle x,y\rangle = \frac{(\|x+y\|^2-\|x\|^2-\|y\|^2)}{2}+i\frac{(\|-ix+y\|^2-\|x\|^2-\|y\|^2)}{2}.$$

On appelle parfois ces formules les identités de polarisation. Elles servent notamment à voir l'unicité du produit scalaire dont peut être issue une norme, et elles servent également à "passer" d'une propriété sur les normes à une propriété sur les produits scalaires. Dans des cas explicites, il n'est cependant pas très naturel de déduire l'expression du produit scalaire à partir de la norme et de ces formules : il est en effet souvent plus rapide de "transformer les carrés" qui apparaissent dans l'expression de la norme par des produits xy ou $x\overline{y}$ suivant le corps de base, voir les exemples ci-après. Néanmoins, ces formules serviront par exemple dans la deuxième question de l'exercice 1.8 pour construire le produit scalaire à partir d'une norme abstraite.

Démonstration. Notons $\|x\| := \sqrt{\langle x,x\rangle}$ sans savoir pour le moment qu'il s'agit d'une norme. Cette application est bien à valeurs dans $\mathbb{R}_+$, elle est homogène[5], et si $\|x\|=0$ alors $x=0$, d'après le caractère défini positif de $\langle\cdot,\cdot\rangle$. La seule partie délicate consiste à montrer l'inégalité triangulaire : celle-ci découle de la formule 1.2 et de l'inégalité de Cauchy-Schwarz 1.1 que nous démontrons ci-dessous, en effet :

$$\forall (x,y)\in H^2,\ \ \|x+y\|^2=\|x\|^2+2\langle x,y\rangle+\|y\|^2\leq \|x\|^2+2\|x\|\|y\|+\|y\|^2=(\|x\|+\|y\|)^2.$$

Les formules 1.2 et 1.4 découlent facilement de la bilinéarité et de la symétrie de $\langle\cdot,\cdot\rangle : H\times H\to\mathbb{R}$ dans le cas réel, ou du caractère sesquilinéaire hermitien dans le cas complexe. Il nous reste donc seulement à montrer l'inégalité de Cauchy-Schwarz :

Cas $\mathbb{K}=\mathbb{R}$: Soit $(x,y)\in H^2$. Si y est nul, alors l'inégalité 1.1 est évidente, et il y a même égalité, et la famille $(x,0)$ est bien liée. On suppose donc $y\neq 0$. On introduit

$$P: t\in\mathbb{R}\mapsto \|x+ty\|^2=\langle x+ty,x+ty\rangle = \|x\|^2+2(\langle x,y\rangle)t+(\|y\|^2)t^2.$$

Il s'agit d'une fonction polynomiale de degré 2 (car $y\neq 0$), qui reste positive du fait des propriétés de positivité de $\langle\cdot,\cdot\rangle$. Ainsi le discriminant de ce polynôme est négatif, c'est-à-dire :

$$4(\langle x,y\rangle)^2-4\|x\|^2\|y\|^2\leq 0$$

ce qui donne l'inégalité de Cauchy-Schwarz en passant à la racine carrée. Pour le cas d'égalité[6], il est rapide de voir que si (x,y) est liée, alors il y a égalité dans (1.1), et réciproquement si $|\langle x,y\rangle|=\|x\|\|y\|$ (toujours en supposant $y\neq 0$), alors le polynôme précédent P est de degré 2 et de discriminant nul, il a donc une racine double $t_0\in\mathbb{R}$, et ainsi $P(t_0)=\|x+t_0y\|^2=0$ c'est-à-dire $x+t_0y=0$.

5. Vérifiez le bien, notamment dans le cas $\mathbb{K}=\mathbb{C}$.

6. Rappelons que pour deux éléments (x,y) d'un espace vectoriel E, la famille (x,y) est liée si et seulement si $\Big[y=0$ ou il existe $\lambda\in\mathbb{R}$ tel que $x=\lambda y\Big]$. C'est une erreur classique d'oublier le cas $y=0$!

Cas $\mathbb{K} = \mathbb{C}$: Soit $(x, y) \in H^2$. A nouveau le cas $y = 0$ est facile, on suppose donc $y \neq 0$. On introduit cette fois

$$P : t \in \mathbb{R} \mapsto \|x + ty\|^2 = \langle x + ty, x + ty\rangle = \|x\|^2 + 2\mathrm{Re}(\langle x, y\rangle)t + (\|y\|^2)t^2,$$

et le même raisonnement que ci-dessus donne

$$\mathrm{Re}(\langle x, y\rangle)^2 \leq \|x\|^2.\|y\|^2. \tag{1.5}$$

On sait qu'il existe[7] $\theta \in \mathbb{R}$ tel que $\langle x, y\rangle = |\langle x, y\rangle|e^{i\theta}$, et donc

$$|\langle x, y\rangle| = \langle x, y\rangle e^{-i\theta} = \langle e^{-i\theta} x, y\rangle = \mathrm{Re}(\langle e^{-i\theta} x, y\rangle)$$

où la dernière égalité vient du fait que $|\langle x, y\rangle| = \langle e^{-i\theta} x, y\rangle \in \mathbb{R}$. On applique (1.5) en remplaçant x par $e^{-i\theta} x$ (et sans changer y), ce qui donne

$$|\langle x, y\rangle|^2 = \mathrm{Re}(\langle e^{-i\theta} x, y\rangle)^2 \leq \|e^{-i\theta} x\|^2.\|y\|^2 = \|x\|^2.\|y\|^2.$$

Pour le cas d'égalité, si on suppose $\|x\|\|y\| = |\langle x, y\rangle|$ et $y \neq 0$, alors toujours avec les notations précédentes, on a $\mathrm{Re}(\langle e^{-i\theta} x, y\rangle) = \|x\|.\|y\|$. Ainsi le discriminant du polynôme

$$Q : t \in \mathbb{R} \mapsto \|e^{-i\theta} x + ty\|^2$$

est nul, donc il existe $t_0 \in \mathbb{R}$ tel que $e^{-i\theta} x + t_0 y = 0$, d'où le fait que (x, y) est liée. □

Exemple 1.6. 1. Sur $\mathbb{R}^n$ pour $n \in \mathbb{N}^*$, on a le produit scalaire dit euclidien suivant[8] :

$$\forall (x, y) \in (\mathbb{R}^n)^2, \ \ \langle x, y\rangle = \sum_{i=1}^{n} x_i y_i,$$

et la norme associée est $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$. L'inégalité de Cauchy-Schwarz s'écrit alors

$$\left|\sum_{i=1}^{n} x_i y_i\right| \leq \sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}. \tag{1.6}$$

2. Toujours sur $\mathbb{R}^n$, si on se donne une matrice $A \in S_n^{++}(\mathbb{R})$ et si on note $\langle\cdot,\cdot\rangle$ le produit scalaire euclidien, alors

$$(X, Y) \in (\mathbb{R}^n)^2 \mapsto \langle AX, Y\rangle = \sum_{i,j} a_{ij} x_i y_j$$

est un produit scalaire sur $\mathbb{R}^n$.

3. Sur $M_n(\mathbb{R})$, on a le produit scalaire

$$\forall (A, B) \in M_n(\mathbb{R})^2, \ \ \langle A, B\rangle := \mathrm{Tr}({}^t A.B) = \sum_{i,j} a_{ij} b_{ij}$$

qui est finalement le produit scalaire euclidien si on identifie $M_n(\mathbb{R})$ à $\mathbb{R}^{n^2}$.

7. On peut préférer dire qu'il existe $u \in \mathbb{U}$ (l'ensemble des nombres complexes de module 1) tel que $\langle x, y\rangle = |\langle x, y\rangle|u$, et poursuivre le calcul en utilisant $u\overline{u} = 1$. C'est un peu moins savant que d'utiliser l'exponentielle complexe, mais cette dernière est quand même bien connue et commode à utiliser.

8. Si on note $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ et $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, alors le produit scalaire s'écrit en terme de produit matriciel : $\langle X, Y\rangle = {}^t X.Y$ (on multiplie une matrice de $M_{1,n}(\mathbb{R})$ par une matrice de $M_{n,1}(\mathbb{R})$ ce qui donne bien un réel).

4. Sur l'espace $\ell^2_{\mathbb{R}}(\mathbb{N}) = \{(u_n) \in \mathbb{R}^{\mathbb{N}},\ \sum_{n\in\mathbb{N}} |u_n|^2 < +\infty\}$ on peut définir :
$$\forall (u_n),(v_n) \in \ell^2_{\mathbb{R}}(\mathbb{N}),\ \ \langle u,v\rangle := \sum_{n\in\mathbb{N}} u_n v_n \tag{1.7}$$
qui est un produit scalaire [9] sur $\ell^2_{\mathbb{R}}(\mathbb{N})$ et dont la norme associée est $\|\cdot\|_2$.
5. L'exemple précédent peut être généralisé : si $(E,\mathcal{A},\mu)$ est un espace mesuré, on peut définir
$$\forall (f,g) \in L^2_{\mathbb{R}}(E,\mathcal{A},\mu),\ \ \langle f,g\rangle := \int_E f(t)g(t)d\mu(t)$$
qui est un produit scalaire sur $L^2_{\mathbb{R}}(E,\mathcal{A},\mu)$ dont la norme associée est
$$\|f\|_{L^2} = \sqrt{\int_E f(t)^2 d\mu(t)}.$$
6. On peut adapter tous les exemples précédents au cas où le corps de base est $\mathbb{C}$:
   (a) Sur $\mathbb{C}^n$, pour $n \in \mathbb{N}^*$, le produit scalaire hermitien (aussi dit "canonique") est
   $$\forall (x,y) \in (\mathbb{C}^n)^2,\ \ \langle x,y\rangle = \sum_{i=1}^n x_i\overline{y_i},$$
   et la norme associée est $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$.
   (b) Sur $\mathbb{C}^n$, si on se donne une matrice $A \in H_n^{++}(\mathbb{C})$ c'est-à-dire une matrice hermitienne ($A^* := {}^t\overline{A} = A$) définie positive, et si on note $\langle\cdot,\cdot\rangle$ le produit scalaire hermitien, alors
   $$(X,Y) \in (\mathbb{C}^n)^2 \mapsto \langle AX, Y\rangle = \sum_{i,j} a_{ij}x_i\overline{y_j}$$
   est un produit scalaire sur $\mathbb{C}^n$.
   (c) Sur $M_n(\mathbb{C})$, on a le produit scalaire
   $$\forall (A,B) \in M_n(\mathbb{C})^2,\ \ \langle A,B\rangle := \mathrm{Tr}({}^tA.\overline{B}) = \sum_{i,j} a_{ij}\overline{b_{ij}}.$$
   (d) Sur l'espace $\ell^2_{\mathbb{C}}(\mathbb{N}) = \{(u_n) \in \mathbb{C}^{\mathbb{N}},\ \sum_{n\in\mathbb{N}} |u_n|^2 < +\infty\}$ on définit le produit scalaire :
   $$\forall (u_n),(v_n) \in \ell^2_{\mathbb{C}}(\mathbb{N}),\ \ \langle u,v\rangle := \sum_{n\in\mathbb{N}} u_n\overline{v_n}.$$
   (e) et plus généralement si $(E,\mathcal{A},\mu)$ est un espace mesuré, on définit le produit scalaire sur $L^2_{\mathbb{C}}(E,\mathcal{A},\mu)$:
   $$\forall (f,g) \in L^2_{\mathbb{C}}(E,\mathcal{A},\mu),\ \ \langle f,g\rangle := \int_E f(t)\overline{g(t)}d\mu(t).$$

9. Pour le vérifier il faut surtout vérifier que la formule (1.7) définit bien un réel, c'est-à-dire que la série $\sum_n u_n v_n$ est convergente. Pour ce faire, on utilise l'inégalité de Cauchy-Schwarz :
$$\sum_{n\in\mathbb{N}} |u_n v_n| \le \sqrt{\sum_{n\in\mathbb{N}} u_n^2}\sqrt{\sum_{n\in\mathbb{N}} v_n^2}. \tag{1.8}$$
Attention ici, on pourrait avoir l'impression qu'on "se mort la queue" car on utilise l'inégalité de Cauchy-Schwarz pour montrer que (1.7) définit bien un produit scalaire, alors que (1.1) a pour hypothèse qu'on a a priori un produit scalaire. On peut en fait montrer (1.8) en utilisant (1.6) pour $(|u_n|,|v_n|)_{n\in\mathbb{N}}$, ce qui donne pour tout $N \in \mathbb{N}$:
$$\sum_{n=1}^N |u_n v_n| \le \sqrt{\sum_{n=1}^N u_n^2}\sqrt{\sum_{n=1}^N v_n^2}$$
puis faire tendre N vers $+\infty$.

Remarque 1.7. L'inégalité de Cauchy-Schwarz (1.1) implique en particulier que dans un espace préhilbertien $(H, \langle\cdot,\cdot\rangle)$, la fonction produit scalaire vue de $H \times H$ dans $\mathbb{K}$ est continue [10]. Ainsi si $(x_n) \in H^{\mathbb{N}}$ converge vers $x \in H$ et $(y_n) \in H^{\mathbb{N}}$ converge vers y, alors

$$\langle x_n, y_n\rangle \xrightarrow[n\to\infty]{} \langle x, y\rangle.$$

Exercice 1.8. Etant donné un espace vectoriel normé $(E, \|\cdot\|)$, il peut être naturel de se demander si la norme $\|\cdot\|$ est issue d'un produit scalaire.

1. On considère l'espace mesuré $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda_n)$ [11]. Montrer que la norme $\|\cdot\|_p$ pour $p \in [1, \infty] \setminus \{2\}$ n'est pas issue d'un produit scalaire.
2. (* voir [Gou08a, Exo 9 page 252]) Montrer que si $(E, \|\cdot\|)$ est un espace vectoriel normé dont la norme satisfait l'identité du parallélogramme, c'est-à-dire telle que :

$$\forall (x, y) \in E^2, \quad \left\|\frac{x+y}{2}\right\|^2 + \left\|\frac{x-y}{2}\right\|^2 = \frac{1}{2}\left(\|x\|^2 + \|y\|^2\right),$$

alors la norme de E est issue d'un produit scalaire.

Définition 1.9. Soit $(H, \langle\cdot,\cdot\rangle)$ un espace préhilbertien.

— Soit $(x, y) \in H^2$. On dit que x et y sont orthogonaux si $\langle x, y\rangle = 0$, et on écrit $x \perp y$.

— Soit $A \subset H$. On pose

$$A^\perp := \left\{x \in H, \ \ \forall a \in A, \ \langle x, a\rangle = 0\right\}$$

l'orthogonal de la partie A.

Remarque 1.10. On peut vérifier facilement la propriété : pour (A, B) deux sous-parties de H,

$$A \subset B \ \ \Rightarrow \ \ A^\perp \supset B^\perp.$$

Remarque 1.11 (Théorème de Pythagore)**.** Soit $(H, \langle\cdot,\cdot\rangle)$ un espace préhilbertien. Si $(x_i)_{i\in[\![1,N]\!]}$ est une famille finie de vecteurs deux à deux orthogonaux (c'est-à-dire $\forall i \neq j \in [\![1, N]\!]^2$, $\langle x_i, x_j\rangle = 0$), on a

$$\left\|\sum_{i=1}^N x_i\right\|^2 = \sum_{i=1}^N \|x_i\|^2.$$

Ceci se montre facilement par récurrence en utilisant (1.2). Notons que si on suppose les sommes convergentes, cette égalité est valable également pour une suite $(x_i)_{i\in\mathbb{N}}$, voir la remarque 1.34.

Proposition 1.12. *Soit $(H, \langle\cdot,\cdot\rangle)$ un espace préhilbertien et $A \subset H$. Alors $A^\perp$ est un sous-espace vectoriel fermé de H, et on a*

$$A^\perp = (\mathrm{Vect}(A))^\perp = \overline{A}^\perp, \qquad et \qquad A \subset (A^\perp)^\perp.$$

Démonstration. 1. Comme $0 \in A^\perp$ et

$$\forall (x, y) \in A^\perp, \forall \lambda \in \mathbb{R}, \forall a \in A, \ \ \langle \lambda x + y, a\rangle = \lambda\langle x, a\rangle + \langle y, a\rangle = 0,$$

$A^\perp$ est un sous-espace vectoriel de H. De plus si (x_n) est une suite d'éléments de $A^\perp$ qui converge vers $x \in H$, alors

$$\forall a \in A, \forall n \in \mathbb{N}, \ \ \langle a, x_n\rangle = 0$$

et par continuité du produit scalaire (voir Remarque 1.7) on peut faire tendre n vers $+\infty$ et obtenir que $x \in A^\perp$, ce qui implique que $A^\perp$ est fermé.

10. Rappelons que (mais il est bon de savoir le redémontrer rapidement) si $b : E \times F \to G$ est une application bilinéaire (et E, F, G sont trois espaces vectoriels normés), alors b est continue si et seulement s'il existe $C \geq 0$ telle que

$$\forall (x, y) \in E \times F, \ \|b(x, y)\|_G \leq C\|x\|_E\|y\|_F.$$

Dans le cas $\mathbb{K} = \mathbb{R}$ cela donne bien la continuité du produit scalaire. Dans le cas $\mathbb{K} = \mathbb{C}$, on peut rapidement constater que la caractérisation précédente convient également pour les applications sesquilinéaires.

11. On n'a pas cherché ici à faire de hypothèses optimales, et on peut retenir que le résultat montré ici est vrai "presque" tout le temps. On invite le lecteur curieux à voir sous quelles condition sur $(E, \mathcal{A}, \mu)$ le résultat tombe en défaut.

2. Comme $A \subset \text{Vect}A$ on a déjà $(\text{Vect}A)^\perp \subset A^\perp$. Réciproquement, si $x \in A^\perp$ et $y \in \text{Vect}(A)$ alors il existe $N \in \mathbb{N}^*$, $(a_i)_{i\in[\![1,N]\!]} \in A^N$ et $(\alpha_i)_{i\in[\![1,N]\!]} \in \mathbb{K}^N$ tel que $y = \sum_i \alpha_i a_i$ et alors
$$\langle y, x\rangle = \sum_i \alpha_i \langle a_i, x\rangle = 0$$
d'où $x \in \text{Vect}(A)^\perp$.
De même on a déjà $\overline{A}^\perp \subset A^\perp$, et si $x \in A^\perp$ et $y \in \overline{A}$, alors il existe $(a_n) \in A^{\mathbb{N}}$ telle que $a_n \longrightarrow y$, et en passant à la limite dans $\forall n \in \mathbb{N}, \langle a_n, x\rangle = 0$ on obtient $\langle y, x\rangle = 0$ et donc $x \in \overline{A}^\perp$.
3. Soit $x \in A$. Alors
$$\forall y \in A^\perp, \quad \langle x, y\rangle = 0$$
ce qui montre que $x \in (A^\perp)^\perp$.

□

1.2 Complétude, Théorème de projection et conséquences

Définition 1.13. Soit $(H, \langle\cdot,\cdot,\rangle)$ un espace préhilbertien. On dit que $(H, \langle\cdot,\cdot\rangle)$ est un espace de Hilbert s'il est complet pour la norme induite par $\langle\cdot,\cdot\rangle$.

Remarque 1.14. Dit autrement, un espace de Hilbert est un espace de Banach dont la norme est issue d'un produit scalaire.

Exemple 1.15. Tous les exemples $(\mathbb{R}^n, M_n(\mathbb{R}), \ell^2(\mathbb{N}), L^2(E, \mathcal{A}, \mu)$ et leurs versions complexes) tirés de l'Exemple 1.6 sont des espaces de Hilbert car ils sont tous complets.

Par contre, si on munit l'espace $C^0([0,1])$ du produit scalaire
$$\forall (f,g) \in C^0([0,1])^2, \quad \langle f, g\rangle := \int_0^1 f(t)g(t)dt$$
alors on a un espace préhilbertien non complet. Le “bon” espace [12] complet associé à ce produit scalaire est $L^2([0,1], \mathcal{B}([0,1]), \lambda)$.

Voici le théorème central de cette partie :

Théorème 1.16 (Projection sur un convexe fermé)**.** *Soit $(H, \langle\cdot,\cdot\rangle)$ un espace de Hilbert, et $C \subset H$ une partie convexe, fermée, non vide. Soit $x \in H$. Alors*

— *il existe un unique élément x_0 qui minimise la distance de x à C, c'est-à-dire tel que*
$$\begin{cases} x_0 \in C, \\ \|x - x_0\| = d(x, C) = \min\left\{\|x - c\|, c \in C\right\} \end{cases}$$
— *l'élément x_0 satisfait la caractérisation :*
$$\begin{cases} x_0 \in C, \\ \|x - x_0\| = d(x, C) \end{cases} \Leftrightarrow \begin{cases} x_0 \in C, \\ \forall c \in C, \quad \langle x - x_0, c - x_0\rangle \leq 0 \ \text{si } \mathbb{K} = \mathbb{R} \\ \forall c \in C, \quad \text{Re}(\langle x - x_0, c - x_0\rangle) \leq 0 \ \text{si } \mathbb{K} = \mathbb{C} \end{cases}$$

On note donc $x_0 := p_C(x)$ et on l'appelle projeté de x sur C. L'application p_C s'appelle la projection sur C.

12. Au sens qu'il est complet et qu'il est le plus petit possible, ce qui se voit car $C^0([0,1])$ est dense dans $L^2([0,1])$.

Démonstration. On pose $d = d(x, C) = \inf\{\|x-c\|, c \in C\}$ qui est bien défini (et est réel) puisque C est non vide.

Unicité : Soit x_1, x_2 deux éléments de C tels que $\|x - x_1\| = \|x - x_2\| = d(x, C)$. Par l'identité du parallélogramme (1.4),

$$\begin{aligned} d^2 \leq \left\| x - \frac{x_1 + x_2}{2} \right\|^2 &= \left\| \frac{x - x_1}{2} + \frac{x - x_2}{2} \right\|^2 \\ &= 2\left(\left\| \frac{x - x_1}{2} \right\|^2 + \left\| \frac{x - x_2}{2} \right\|^2 \right) - \left\| \frac{x - x_1}{2} - \frac{x - x_2}{2} \right\|^2 = d^2 - \left\| \frac{x_2 - x_1}{2} \right\|^2 \end{aligned}$$

où la première inégalité s'obtient parce que $(x_1 + x_2)/2 \in C$ par convexité de C. Ainsi $x_1 = x_2$.

Existence : Comme C est non vide, il existe une suite $(x_n) \in C^{\mathbb{N}}$ telle que $\|x - x_n\| \xrightarrow[n\to\infty]{} d$. Nous allons montrer que cette suite converge en montrant qu'elle est de Cauchy : soit $\varepsilon > 0$. Par convergence de la suite $(\|x - x_n\|)_{n\in\mathbb{N}}$, il existe $n_0 \in \mathbb{N}$ tel que

$$\forall n \geq n_0, \quad \|x - x_n\|^2 - d^2 \leq \varepsilon.$$

Alors par l'égalité du parallélogramme on obtient :

$$\forall p \geq n_0, \forall q \geq n_0, \quad \|x_q - x_p\|^2 = \|(x - x_q) - (x - x_p)\|^2 = 2\left(\underbrace{\|x - x_q\|^2 + \|x - x_p\|^2}_{\leq 2d^2 + 2\varepsilon} \right) - \|x_p + x_q - 2x\|^2$$

or

$$\|x_p + x_q - 2x\|^2 = 4\left\| x - \frac{x_p + x_q}{2} \right\|^2 \geq 4d^2$$

puisque $(x_p + x_q)/2 \in C$, ce qui donne finalement :

$$\forall p \geq n_0, \forall q \geq n_0, \quad \|x_q - x_p\|^2 \leq 2\varepsilon.$$

Par complétude de H, la suite (x_n) converge vers un élément noté x_0, qui est bien dans C parce que C est supposé fermé. Par continuité de la norme, on a bien

$$\|x - x_0\| = \lim_{n\to\infty} \|x - x_n\| = d(x, C).$$

Caractérisation : Supposons $\mathbb{K} = \mathbb{R}$, le cas $\mathbb{K} = \mathbb{C}$ se traitant par les mêmes calculs. [13]

$\Rightarrow$ Supposons que $x_0 \in C$ minimise la distance de x à C et soit $y \in C$. Alors pour $t \in [0, 1]$, $(1-t)x_0 + ty \in C$ donc [14]

$$\|x - x_0\|^2 \leq \|x - [(1-t)x_0 + ty]\|^2 = \|x - x_0 + t(x_0 - y)\|^2 = \|x - x_0\|^2 + 2t\langle x_0 - y, x - x_0\rangle + t^2\|x_0 - y\|^2$$

d'où

$$\forall t \in [0, 1], \quad 2t\langle y - x_0, x - x_0\rangle \leq t^2\|x_0 - y\|^2 \leq 0.$$

En divisant par $t > 0$ et en faisant tendre t vers 0 par valeur supérieure, on en déduit $\langle y - x_0, x - x_0\rangle \leq 0$.

$\Leftarrow$ Supposons que $x_0 \in C$ est tel que $\forall y \in C, \langle y - x_0, x - x_0\rangle \leq 0$. Soit $c \in C$. Alors

$$\|x - c\|^2 = \|x - x_0 + x_0 - c\|^2 = \|x - x_0\|^2 + 2\underbrace{\langle x - x_0, x_0 - c\rangle}_{\geq 0} + \|x_0 - c\|^2 \geq \|x - x_0\|^2$$

d'où le fait que x_0 minimise la distance de x à C. □

13. Dans la première partie de la preuve, nous n'avons pas distingué le corps de base car les calculs sont valables dans les deux cas $\mathbb{K} = \mathbb{R}$ et $\mathbb{K} = \mathbb{C}$, principalement parce que tout repose sur l'identité du parallélogramme qui est la même quel que soit le corps de base.

14. Il est important d'essayer de comprendre ce qu'on fait dans cette partie de la preuve. On est en fait en train d'écrire une condition d'optimalité pour le problème $\min\{\|x - c\|, c \in C\}$ (qu'il vaut mieux voir comme $\min\{\|x - c\|^2, c \in C\}$ car les calculs de dérivation sont plus simples), dans la lignée de ce qui est présenté à la section 5 du Chapitre VII. Précisons quand même qu'on ne cherche pas à appliquer directement les énoncés de cette section, d'une part parce qu'on se satisfait de dérivation directionnelle, et surtout parce qu'on n'est ni dans un cas d'optimisation sans contrainte, ni dans un cas de contrainte d'égalité, mais dans une contrainte ($x \in C$) plus spécifique. Néanmoins, on peut voir cette caractérisation comme une condition d'optimalité, et voir qu'il s'agit d'une condition nécessaire et suffisante car le problème d'optimisation considéré est un problème convexe (la fonction optimisée est convexe et l'ensemble des éléments admissibles est convexe).

Exercice 1.17. (* [HL99, page 92]) Avec les notations du Théorème 1.16, montrer que p_C est 1-lipschitzienne.

Il est fréquent d'utiliser le Théorème 1.16 dans le cas particulier où on projette sur un sous-espace vectoriel fermé de H.

Corollaire 1.18. *Soit H un espace de Hilbert, F un sous-espace vectoriel fermé de H, et $x \in H$. Alors*

1. *il existe un unique élément $x_0 \in F$ tel que $\|x - x_0\| = d(x, F)$, et cet élément est caractérisé par le fait :*
$$\begin{cases} x_0 \in F, \\ x - x_0 \in F^\perp. \end{cases}$$
2. *On a l'égalité $H = F \oplus F^\perp$, et p_F est la projection F parallèlement à $F^\perp$, qu'on appelle projection orthogonale sur F.*
3. *On a l'égalité $(F^\perp)^\perp = F$.*

Remarque 1.19. Si F est un sous-espace vectoriel non supposé fermé de H, on peut appliquer le corollaire précédent à $\overline{F}$ qui est bien un sev fermé de H, et on obtient :
$$H = \overline{F} \oplus F^\perp, \quad (F^\perp)^\perp = \overline{F},$$
puisque $\overline{F}^\perp = F^\perp$. En particulier, on obtient le critère de densité suivant, très fréquemment utilisé [15] :
$$F \text{ est dense dans } H \quad \Leftrightarrow \quad F^\perp = \{0\}.$$

Démonstration. 1. On commence par le cas $\mathbb{K} = \mathbb{R}$. Comme F est convexe et fermé, on peut appliquer le Théorème 1.16 qui donne l'existence et l'unicité du projeté, et nous dit que pour un $x \in H$ donné, la projection $x_0 \in F$ est caractérisée par
$$\forall y \in F,\ \langle x - x_0, y - x_0 \rangle \leq 0. \tag{1.9}$$
Si on suppose $x - x_0 \in F^\perp$, alors effectivement $\forall y \in F,\ \langle x - x_0, y - x_0 \rangle = 0 \leq 0$ puisque $y - x_0 \in F$ par structure d'espace vectoriel de F. Réciproquement, si on suppose (1.9), alors comme pour $z \in F$ on a $z + x_0 \in F$, et on en déduit
$$\forall z \in F,\ \langle x - x_0, z \rangle \leq 0$$
et comme pour tout $z \in F$, $-z \in F$, finalement
$$\forall z \in F,\ \langle x - x_0, z \rangle = 0, \qquad \text{i.e.} \qquad x - x_0 \in F^\perp.$$
Dans le cas $\mathbb{K} = \mathbb{C}$, la preuve est très similaire. On suppose que
$$\forall y \in F,\ \mathrm{Re}(\langle x - x_0, y - x_0 \rangle) \leq 0,$$
et on doit montrer $x - x_0 \in F^\perp$ (l'autre sens de l'équivalence étant facile). Comme dans le cas réel, en utilisant $z + x_0$ et $-z$ pour $z \in F$, on déduit que
$$\forall z \in F,\ \mathrm{Re}(\langle x - x_0, z \rangle) = 0,$$
et il reste à voir que si $z \in F$ alors $iz \in F$, et comme $\mathrm{Re}(\langle x - x_0, iz \rangle) = \mathrm{Im}(\langle x - x_0, z \rangle)$, on obtient bien $\langle x - x_0, z \rangle = 0$ pour tout $z \in F$.

2. Ensuite, il suffit d'écrire
$$\forall x \in H,\ \ x = \underbrace{p_F(x)}_{\in F} + \underbrace{(x - p_F(x))}_{\in F^\perp},$$
qui montre $H = F + F^\perp$ et que p_F est bien la projection sur F parallèlement à $F^\perp$, et la somme est directe car $F \cap F^\perp = \{0\}$.

15. Et donc à connaître par cœur.

3. Enfin, on a déjà $F \subset (F^\perp)^\perp$ par la Proposition 1.12. Réciproquement si $x \in (F^\perp)^\perp$, alors $x - p_F(x) \in (F^\perp)^\perp \cap F^\perp$ donc $x = p_F(x) \in F$.

□

Exemple 1.20. [Voir [All12, Th 9.2.6]] La preuve d'existence donnée en prouvant le Théorème 1.16 peut se généraliser : soit $(E, \|\cdot\|)$ un espace de Banach, $C \subset E$ un convexe fermé non vide, et $J : C \to \mathbb{R}$ continue et minorée [16]. On suppose J α-convexe pour $\alpha > 0$, c'est-à-dire :

$$\forall t \in [0,1], \forall (x,y) \in E^2, \quad J(tx + (1-t)y) \leq tJ(x) + (1-t)J(y) - \frac{\alpha}{2} t(1-t)\|x-y\|^2.$$

1. Montrer qu'il existe un unique minimum de J sur C.
2. Si H est un espace de Hilbert, étudier pour quelles valeurs de α la fonction $J : x \in H \mapsto \|x\|^2$ est α-convexe, et en déduire que la première question permet de retrouver l'existence et l'unicité d'un projeté sur un convexe fermé $C \subset H$ (première partie du Théorème 1.16).

Dualité

Une des applications classiques du Théorème de projection 1.16 est de pouvoir identifier le dual d'un espace de Hilbert. Commençons par la remarque suivante, valable sans hypothèse de complétude :

Proposition 1.21. *Soit $(H, \langle\cdot,\cdot\rangle)$ un espace préhilbertien. Alors pour tout $x \in H$ l'application $\varphi_y : x \in H \mapsto \langle x, y\rangle \in \mathbb{K}$ est une forme linéaire continue sur H, et $\|\varphi_y\|_{H'} = \|y\|$ [17].*
Dit autrement, on peut définir

$$\begin{array}{rcl} \Phi : H & \longrightarrow & H' \\ x & \longmapsto & \varphi_y \end{array} \tag{1.10}$$

qui est de plus une isométrie, linéaire si $\mathbb{K} = \mathbb{R}$, anti-linéaire si $\mathbb{K} = \mathbb{C}$.

Démonstration. Pour $y \in H$, la définition du produit scalaire donne la linéarité de φ_y, et l'inégalité de Cauchy-Schwarz nous donne que $\varphi_y \in H'$, et que $\|\varphi_y\|_{H'} \leq \|y\|$. L'inégalité opposée est donnée par $\varphi_y(y) = \|y\|^2$.

Enfin, Φ est bien définie, isométrique, et sa linéaire/anti-linéarité vient à nouveau de la définition du produit scalaire. □

Autrement dit, nous pouvons avoir accès à de nombreuses formes linéaires sur H en regardant le produit scalaire avec un vecteur fixe de H. Le Théorème de Riesz donne une réciproque à cette propriété :

Théorème 1.22 (Théorème de Riesz)**.** *Soit $(H, \langle\cdot,\cdot\rangle)$ un espace de Hilbert. Alors pour tout $\ell \in H'$, il existe un unique $y \in H$ tel que*

$$\forall x \in H, \quad \ell(x) = \langle x, y\rangle,$$

autrement dit l'application Φ définie en (1.10) est une isométrie surjective.

Démonstration. Avec la Proposition 1.21, il reste à montrer le caractère surjectif. Soit donc $\ell \in H'$, qu'on suppose non nulle (si $\ell = 0$ on peut choisir $y = 0$). L'ensemble $F := \mathrm{Ker}(\ell)$ est un sous-espace vectoriel de H, il est fermé car ℓ est continue, et H étant supposé complet, le Corollaire 1.18 nous donne $H = F \oplus F^\perp$.

16. *On peut en fait montrer qu'une fonction α-convexe et continue est nécessairement minorée, et même coercive. On peut voir par exemple [All12, Proposition 9.2.5] ; la preuve utilise le théorème de séparation de Hahn-Banach. Faites-moi savoir si vous connaissez une preuve plus simple.

17. Cette égalité peut se réécrire :

$$\forall y \in H, \quad \|y\| = \sup_{x \in H \setminus \{0\}} \frac{|\langle x, y\rangle|}{\|y\|}.$$

Comme $\ell \neq 0$, il existe $e \in F^{\perp} \setminus \{0\}$, qu'on peut choisir tel que $\|e\| = 1$. Pour tout $x \in H$ on peut constater que [18]

$$x = \underbrace{x - \frac{\ell(x)}{\ell(e)}e}_{\in F} + \underbrace{\frac{\ell(x)}{\ell(e)}e}_{\in F^{\perp}} \tag{1.11}$$

qui est la décomposition de x dans la décomposition $F \oplus F^{\perp}$, et qui donne en prenant le produit scalaire avec e et en utilisant $\langle z, e\rangle = 0$ si $z \in F$,

$$\forall x \in H, \quad \langle x, e\rangle = \frac{\ell(x)}{\ell(e)}$$

et donc $y = \ell(e)e$ dans le cas $\mathbb{K} = \mathbb{R}$ et $y = \overline{\ell(e)}e$ dans le cas $\mathbb{K} = \mathbb{C}$ conviennent. □

Le résultat qui suit est une généralisation du théorème de Riesz dans le cas d'une forme bilinéaire qui n'est pas forcément symétrique. Il sert essentiellement pour la résolution de problèmes aux limites, voir le paragraphe 2.3 et en particulier l'exercice 2.30.

Théorème 1.23 (Lax-Milgram). *Soit $(H, \langle \cdot, \cdot \rangle)$ un espace de Hilbert, $L \in H'$ une forme linéaire continue, et $a \colon H \times H \to \mathbb{R}$ une forme bilinéaire qui satisfait :*

— *continuité : il existe $M > 0$ telle que $|a(u,v)| \leq M\|u\|_H\|v\|_H$, pour tous u, $v \in H$;*

— *coercivité : il existe $\alpha > 0$ telle que $a(u,u) \geq \alpha\|u\|_H^2$, pour tout $u \in H$.*

Alors, il existe un unique $u \in H$ tel que [19]

$$a(u,v) = L(v), \ \textit{pour tout } v \in H. \tag{1.12}$$

Remarque 1.24. Si la forme bilinéaire a est de plus supposée symétrique, alors $a(\cdot,\cdot)$ définit sur H un nouveau produit scalaire, et d'après les hypothèses faites sur a, la norme issue de ce nouveau produit scalaire est équivalente à la norme initiale induite par $\langle \cdot, \cdot \rangle$. Ainsi le théorème de Riesz s'applique (l'espace H muni de ce nouveau produit scalaire est encore complet et la forme linéaire est encore continue pour la nouvelle norme), ce qui donne une preuve alternative du résultat précédent dans ce cas.

Aussi, dans ce cas on peut vérifier [20] (dans le cas où a est symétrique) que l'unique solution du problème (1.12) est aussi solution de

$$\inf_{v \in H} J(v) \text{ avec } \ J(v) = \frac{1}{2}a(v,v) - L(v). \tag{1.13}$$

Démonstration. Pour $u \in H$, on définit $L_u : v \in H \mapsto a(u,v)$. La bilinéarité de a montre que L_u est une forme linéaire, et la continuité de a implique que L_u est continue avec $\|L_u\|_{H'} \leq M\|u\|_H$. Par le théorème de Riesz, il existe $A(u) \in H$ et $f \in H$ qui représentent respectivement L_u et L à l'aide du produit scalaire.

On est donc ramené à démontrer l'existence et l'unicité d'un $u \in H$ tel que

$$A(u) = f, \tag{1.14}$$

c'est-à-dire à montrer que l'application A est bijective de H dans H.

18. L'équation (1.11) est en fait l'argument classique pour voir que le noyau d'une forme linéaire est de co-dimension 1, c'est-à-dire admet un supplémentaire de dimension 1. Si on veut se passer de la redémonstration de ce fait, on peut formuler la preuve ainsi : sachant que ℓ est non nulle, $F = \mathrm{Ker}(\ell)$ est un hyperplan, donc $F^{\perp}$ est de dimension 1 (tous les supplémentaires d'un hyperplan sont de dimension 1), et on peut considérer $e \in F^{\perp}$ de norme 1 qui engendre $F^{\perp}$. Pour $y = \ell(e)e$ (si $\mathbb{K} = \mathbb{R}$ sinon on prend $y = \overline{\ell(e)}e$), on constate que φ_y et ℓ coïncident sur F (elles y sont nulles) et en e, elles sont donc égales sur H.

19. On peut aussi montrer $\|u\|_H \leq \frac{1}{\alpha}\|L\|_{H'}$.

20. Preuve rapide : si u est solution de (1.12), alors $a(u,w) = L(w)$ pour tout $w \in H$ et après calcul $J(u+w) \geq J(u)$. Réciproquement, si $J(u+w) \geq J(u)$ pour tout $w \in H$, on prend $w = tv$ avec $t > 0$ dans (1.12), puis on fait tendre $t \to 0^+$, il vient alors $a(u,v) - L(v) \geq 0$, puis $a(u,v) - L(v) = 0$ en changeant v en $-v$.
En fait, l'équation (1.12) peut être vue comme la condition d'optimalité pour le problème d'optimisation (1.13), c'est-à-dire $dJ(u) = 0$. Il y a équivalence car la fonction J est convexe.

Tout d'abord, on vérifie facilement que l'application $A\colon H \to H$ est linéaire, et elle est de plus continue puisque $\|A(u)\|_H \leq M\|u\|_H$. Par ailleurs, la coercivité de a implique que A est injective car si $A(u) = 0$, alors $0 = \langle A(u), u\rangle = a(u,u) \geq \alpha\|u\|_H^2$, ce qui implique que $u = 0$. Pour montrer la surjectivité, on établit d'abord que $\operatorname{Im} A$ est fermé dans H. Pour ce faire, on considère une suite $(v_n = A(u_n))_{n\in\mathbb{N}}$ de $\operatorname{Im} A$ telle que $v_n \to v$ dans H. Par coercivité de a on en déduit que pour tout $(p,q) \in \mathbb{N}^2$,

$$\begin{aligned}\alpha\|u_q - u_p\|_H^2 \leq a(u_q - u_p, u_q - u_p) &= \langle A(u_q) - A(u_p), u_q - u_p, \rangle \\ &= \langle v_q - v_p, u_q - u_p, \rangle \leq \|v_q - v_p\|_H \|u_q - u_p\|_H,\end{aligned}$$

par l'inégalité de Cauchy-Schwarz. Il s'ensuit que $(u_n)_{n\in\mathbb{N}}$ est de Cauchy dans H, ce qui assure l'existence d'un $u \in H$ tel que $u_n \to u$, et par continuité de A, il vient $v = A(u)$ ce qui montre que $\operatorname{Im} A$ est fermé dans H. D'après le corollaire 1.18, on peut donc décomposer $H = \operatorname{Im} A \oplus (\operatorname{Im} A)^\perp$. Soit $u \in (\operatorname{Im} A)^\perp$: alors pour tout $v \in H$, $\langle u, A(v)\rangle = 0$. En particulier, le choix $v = u$ montre que $\alpha\|u\|_H^2 \leq a(u,u) = \langle u, A(u)\rangle = 0$, soit $u = 0$, donc finalement $(\operatorname{Im} A)^\perp = \{0\}$ et $H = \operatorname{Im} A$, ce qui conclut la démonstration. □

1.3 Bases hilbertiennes

Commençons pas un petit rappel algébrique : dans un espace préhilbertien $(H, \langle\cdot,\cdot\rangle)$, si une famille $(e_i)_{i\in I}$ d'éléments de H est orthonormale, c'est-à-dire si elle satisfait

$$\forall (i,j) \in I^2, \quad \langle e_i, e_j\rangle = \delta_{ij},$$

alors elle est libre. En effet si $J \subset I$ est fini et $\sum_{j\in J} \lambda_j e_j = 0$, alors pour tout $i \in J$, le produit scalaire avec e_i donne $\lambda_i = 0$ [21].

Dans le cadre habituel de l'algèbre linéaire, on aime bien travailler avec des familles libres et génératrices, c'est-à-dire des bases. Rappelons qu'une famille $(e_i)_{i\in I}$ de H est dite génératrice si tout élément de H est une combinaison linéaire **finie** d'éléments de $(e_i)_{i\in I}$ c'est-à-dire si $\operatorname{Vect}((e_i)_{i\in I}) = H$ [22] (et cette écriture devient unique si la famille est de plus libre). Mais nous allons voir que cet outil n'est pas commode pour étudier des espaces de Hilbert qui sont de dimension infinie [23]. Pour cette raison, nous introduisons la définition suivante :

Définition 1.25. Soit H un espace de Hilbert. Une famille $(e_i)_{i\in I}$ de H est appelée base hilbertienne de H si

— elle est orthonormale, c'est-à-dire $\forall (i,j) \in I^2, \quad \langle e_i, e_j\rangle = \delta_{ij}$,

— elle engendre un espace dense dans H, c'est-à-dire :

$$\overline{\operatorname{Vect}((e_i)_{i\in I})} = H. \tag{1.15}$$

Remarque 1.26. Grâce au Corollaire 1.18, la condition $\overline{\operatorname{Vect}((e_i)_{i\in I})} = H$ est équivalente à $\big((e_i)_{i\in I}\big)^\perp = \{0\}$, c'est-à-dire

$$\forall x \in H, \quad \Big[\forall i \in I,\ \langle x, e_i\rangle = 0, \ \Rightarrow \ x = 0\Big].$$

Il est commode de se souvenir de cette caractérisation qui sert souvent.

Remarque 1.27. Attention, contrairement à ce que laisse penser la terminologie, une base hilbertienne n'est pas, en général, une base au sens algébrique du terme. Plus précisément, on sait que dans un evn complet de dimension infinie, il n'existe jamais de base algébrique dénombrable (voir le théorème 3.6), et donc dans le cas des espaces de Hilbert de dimension infinie, les bases hilbertiennes dénombrables (et on se concentrera sur le cas de famille dénombrable dans ce qui suit) ne sont jamais des bases algébriques.

21. Notons qu'on n'a pas vraiment eu besoin de $\langle e_i, e_i\rangle = 1$ mais plutôt de $\langle e_i, e_i\rangle \neq 0$

22. Rappelons que dans le cas où I est infini, on a

$$\operatorname{Vect}((e_i)_{i\in I}) = \left\{ \sum_{j\in J} \lambda_j e_j, \quad J \subset I \text{ fini}, \lambda_j \in \mathbb{R}^J \right\}.$$

23. Rappelons qu'on veut ici travailler dans des espaces fonctionnels, ceux-ci sont de dimension infinie sauf cas atypiques

Remarque 1.28. Il n'est pas rare de manipuler des familles $(e_i)_{i\in I}$ qui sont seulement orthogonales, c'est-à-dire $\forall i \neq j,\ \langle e_i, e_j\rangle = 0$ et si on suppose $\forall i \in I,\ \langle e_i, e_i\rangle \neq 0$, alors la famille $\frac{e_i}{\|e_i\|}$ sera orthonormale. On pourra occasionnellement s'intéresser à des familles dont la version renormalisée est une base hilbertienne, voir plus loin le paragraphe sur les polynômes orthogonaux.

Exemple 1.29. Dans l'espace $\ell^2(\mathbb{N})$, la famille $(e_i)_{i\in\mathbb{N}}$ définie par

$$\forall (i,n) \in \mathbb{N}^2, \quad e_i(n) = \delta_{in}$$

est une base hilbertienne de $\ell^2(\mathbb{N})$. En effet :

1. Pour tout $(i,j) \in \mathbb{N}^2, i \neq j$,
$$\langle e_i, e_j\rangle = \sum_{n\in\mathbb{N}} e_i(n)\overline{e_j(n)} = \delta_{ij}.$$
2. Pour la densité, on peut faire deux preuves :
   (a) Pour tout $u \in \ell^2(\mathbb{N})$, on peut définir $\forall N \in \mathbb{N},\ \ u^{(N)} = \sum_{i=0}^N u(i)e_i$ (c'est-à-dire la troncation de u à $[\![0, N]\!]$), et comme
   $$\|u - u^{(N)}\|_2^2 = \sum_{n>N} |u(n)|^2 \xrightarrow[N\to\infty]{} 0$$
   on a bien $\mathrm{Vect}((e_i)_{i\in\mathbb{N}})$ dense dans $\ell^2(\mathbb{N})$.
   (b) On utilise le critère de densité donné à la remarque 1.28 : soit u tel que pour tout $i \in \mathbb{N}$, $\langle u, e_i\rangle = 0$. Alors comme $\langle u, e_i\rangle = u(i)$, cela implique bien que $u = 0$.

Dans l'énoncé suivant, de façon similaire à ce qu'on fait pour les espaces vectoriels, nous étudions l'existence de base hilbertienne :

Théorème 1.30. *Soit $(H, \langle\cdot,\cdot\rangle)$ un espace de Hilbert, qu'on suppose séparable*[24] *et de dimension infinie. Alors toute base hilbertienne est infinie dénombrable*[25]*, et il en existe.*

Remarque 1.31. Dans le cas d'un espace H de dimension finie, alors il n'est pas difficile de voire que toute base hilbertienne est une base algébrique (elle est libre (donc finie), et comme tout espace de dimension finie est fermé, la condition (1.15) revient à dire que la famille est génératrice), et aussi qu'il en existe : pour cette dernière propriété d'existence, on invoque l'existence d'une base algébrique, et on lui applique le procédé de Gram-Schmidt (dont on donne une variante dans la preuve ci-dessous), qui donne bien l'existence d'une base hilbertienne. Il s'agit en fait dans ce cas d'une base orthonormale, on réserve généralement le terme de base hilbertienne aux familles infinies.

Dans le cas d'un espace de Hilbert de dimension infinie mais non séparable, on peut quand même montrer l'existence d'une base hilbertienne $(e_i)_{i\in I}$, mais cela est plus délicat et recquiert l'axiome du choix (comme c'est le cas dans la preuve d'existence d'une base algébrique dans un espace vectoriel de dimension infinie).

Démonstration. Commençons par construire une base hilbertienne de H. Comme l'espace H est séparable, il existe une famille $(e_i)_{i\in\mathbb{N}}$ dense dans H[26]. Nous allons montrer qu'il existe $(\varepsilon_i)_{i\in\mathbb{N}}$ une famille dénombrable de H telle que

$$\begin{cases} \forall N \in \mathbb{N}, \ \ \mathrm{Vect}(e_0, \cdots, e_N) \subset \mathrm{Vect}(\varepsilon_0, \cdots, \varepsilon_N) \\ \forall (i,j) \in \mathbb{N}^2, \ \ \langle \varepsilon_i, \varepsilon_j\rangle = 0 \end{cases}$$

ce qui permet de conclure que $(\varepsilon_n)_{n\in\mathbb{N}}$ est une base hilbertienne car

$$\mathrm{Vect}((\varepsilon_n)_{n\in\mathbb{N}}) = \bigcup_{N\in\mathbb{N}} \mathrm{Vect}(\varepsilon_0, \cdots, \varepsilon_N) \supset (e_n)_{n\in\mathbb{N}}.$$

24. Rappelons qu'on appelle espace séparable un espace métrique contenant une partie dénombrable et dense, voir le Chapitre III.
25. On veut dire ici que toutes les bases ont le même cardinal, en l'occurrence le cardinal de $\mathbb{N}$.
26. Attention, on ne suppose rien d'autre, en particulier pas la liberté de cette famille ; il y a donc possiblement beaucoup de "redondance" dans cette famille.

Pour construire ε_0, on note $\varphi(0)$ le plus petit entier i tel que $e_i \neq 0$ et on pose $\varepsilon_0 = \frac{e_{\varphi(0)}}{\|e_{\varphi(0)}\|}$.

Pour un certain $N \in \mathbb{N}$, on suppose construits $(\varepsilon_0, \dots, \varepsilon_N)$ orthogonaux et tels que $\mathrm{Vect}(e_0, \cdots, e_N) \subset \mathrm{Vect}(\varepsilon_0, \cdots, \varepsilon_N)$. Comme H est de dimension infinie, on sait qu'il existe $i \in \mathbb{N}$ tel que $e_i \notin \mathrm{Vect}(\varepsilon_0, \cdots, \varepsilon_N)$, et on note $\varphi(N+1)$ l'indice minimal pour cette propriété (qui est nécessairement strictement plus grand que N), et on pose

$$f_{N+1} = e_{\varphi(N+1)} + \sum_{i=0}^{N} \alpha_i \varepsilon_i \text{ avec } \forall i \in [\![0, N]\!],\ \alpha_i = -\langle e_{\varphi(N+1)}, \varepsilon_i \rangle$$

où le choix des α_i permettent d'avoir $\langle f_{N+1}, \varepsilon_i \rangle = 0$ pour tout $i \leq N$. Enfin on pose $\varepsilon_{N+1} = \frac{f_{N+1}}{\|f_{N+1}\|}$, qui est bien de norme 1, orthogonal à ε_i pour $i \in [\![0, N]\!]$, et on a bien $\mathrm{Vect}(e_0, \cdots, e_N, e_{N+1}) \subset \mathrm{Vect}(\varepsilon_0, \cdots, \varepsilon_{N+1})$ car :

— soit $\varphi(N+1) = N+1$ et alors $e_{N+1} = f_{N+1} - \sum_{i=0}^{N} \alpha_i \varepsilon_i$,

— soit $\varphi(N+1) > N+1$ et alors $e_{N+1} \in \mathrm{Vect}(\varepsilon_0, \cdots, \varepsilon_N)$ par définition de $\varphi(N+1)$.

Pour conclure, il nous reste à voir d'une part que comme H est de dimension infinie, une base hilbertienne ne peut pas être finie, d'autre part qu'elle doit être dénombrable. C'est en fait le cas de toute famille orthonormale dans un espace séparable : en effet si $(\varepsilon_i)_{i \in I}$ est orthonormale, alors un calcul rapide donne

$$\forall (i,j) \in I^2 \text{ tel que } i \neq j,\ \ \|\varepsilon_j - \varepsilon_i\| = \sqrt{2}.$$

Par densité de $(e_n)_{n \in \mathbb{N}}$, pour tout $i \in I$ on peut trouver $\sigma(i) \in \mathbb{N}$ tel que $e_{\sigma(i)} \in B(\varepsilon_i, \frac{\sqrt{2}}{2})$, et l'application $i \in I \mapsto \sigma(n) \in \mathbb{N}$ est nécessairement injective vu que les boules $B(\varepsilon_i, \frac{\sqrt{2}}{2})$ sont disjointes, et donc I est dénombrable.

□

Dans le résultat suivant, on explique comment tout élément de H peut s'écrire “en coordonnées dans une base hilbertienne” : c'est très similaire au cas d'une base algébrique, mais la somme est infinie.

Théorème 1.32. *Soit H un espace de Hilbert de dimension infinie, et $(e_n)_{n \in \mathbb{N}}$ une base hilbertienne (infinie dénombrable) de H. Alors*[27]

$$\forall x \in H, \quad x = \sum_{n \in \mathbb{N}} \langle x, e_n \rangle e_n. \tag{1.16}$$

De plus,

$$\begin{array}{rcl} \Phi : H & \longrightarrow & \ell^2(\mathbb{N}) \\ x & \longmapsto & (\langle x, e_n \rangle)_{n \in \mathbb{N}} \end{array}$$

est une isométrie linéaire et surjective, dont l'inverse est

$$\begin{array}{rcl} \Phi^{-1} : \ell^2(\mathbb{N}) & \longrightarrow & H \\ (\alpha_n)_{n \in \mathbb{N}} & \longmapsto & \sum_{n \in \mathbb{N}} \alpha_n e_n \end{array} \tag{1.17}$$

Remarque 1.33. L'égalité (1.16) combinée avec le théorème de Pythagore (voir la remarque **??**) donne l'égalité dite de Parseval :

$$\forall x \in H, \quad \|x\|^2 = \sum_{n \in \mathbb{N}} |\langle x, e_n \rangle|^2. \tag{1.18}$$

Par formule de polarisation (voir la remarque 1.5), on en déduit la version plus générale suivante :

$$\forall (x,y) \in H^2, \quad \langle x, y \rangle = \sum_{n \in \mathbb{N}} \langle x, e_n \rangle \overline{\langle y, e_n \rangle}.$$

27. On entend ici que la série $(\sum_{n=0}^{N} \langle x, e_n \rangle e_n)_N$ est convergente et converge vers x, c'est-à-dire

$$\sum_{n=0}^{N} \langle x, e_n \rangle e_n \xrightarrow[N \to \infty]{} x.$$

Remarque 1.34. Si H est un espace de Hilbert, et $(x_n)_{n\in\mathbb{N}}$ une famille orthogonale de H (c'est-à-dire $\forall(i,j)\in\mathbb{N}^2, i\neq j \Rightarrow \langle x_i, x_j\rangle = 0$), on a la caractérisation [28]

$$\sum_{n\in\mathbb{N}} x_n \quad \text{converge} \quad \Leftrightarrow \quad \sum_{n\in\mathbb{N}} \|x_n\|^2 < +\infty,$$

et dans ce cas

$$\left\|\sum_{n\in\mathbb{N}} x_n\right\|^2 = \sum_{n\in\mathbb{N}} \|x_n\|^2,$$

qui généralise le théorème de Pythagore à une famille infinie.

En effet :

— $\Rightarrow$ On pose $S = \sum_n x_n \in H$ la limite de cette série qu'on suppose convergente. Alors par le Théorème de Pythagore (voir la remarque 1.11), on a

$$\forall N\in\mathbb{N}, \quad \|\sum_{n=0}^{N} x_n\|^2 = \sum_{n=0}^{N} \|x_n\|^2$$

et comme le terme de gauche converge pour $N\to\infty$, la série $(\sum_{n=0}^{N}\|x_n\|^2)_N$ est convergente dans $\mathbb{R}$.

— $\Leftarrow$ On suppose $\sum_{n\in\mathbb{N}}\|x_n\|^2 < +\infty$. Encore avec le théorème de Pythagore, on a pour tout $(p,q)\in\mathbb{N}^2$ avec $p<q$:

$$\left\|\sum_{n=0}^{q} x_n - \sum_{n=0}^{p} x_n\right\|^2 = \left\|\sum_{n=p+1}^{q} x_n\right\|^2 = \sum_{n=p+1}^{q}\|x_n\|^2 = \sum_{n=0}^{q}\|x_n\|^2 - \sum_{n=0}^{p}\|x_n\|^2.$$

Comme $(\sum_{n=0}^{N}\|x_n\|^2)_{N\in\mathbb{N}}$ est de Cauchy dans $\mathbb{R}$ (car elle est convergente), l'inégalité précédente montre que $(\sum_{n=0}^{N} x_n)_{N\in\mathbb{N}}$ est de Cauchy dans H, et donc convergente par complétude.

Démonstration. Soit $x\in H$. Par définition $\mathrm{Vect}(e_n)_{n\in\mathbb{N}}$ est dense dans H, et donc en posant $F_N = \mathrm{Vect}(e_0,\dots,e_N)$ (qui sont des sev fermés de H) pour $N\in\mathbb{N}$, la suite $(p_{F_N}(x))_{N\in\mathbb{N}}$ des projections de x sur F_N (obenues par le Théorème 1.16) converge vers x [29]. Mais classiquement on a l'expression suivante :

$$\forall N\in\mathbb{N}, \quad p_{F_N}(x) = \sum_{n=0}^{N}\langle x, e_n\rangle e_n.$$

En effet, en posant $x_N = \sum_{n=0}^{N}\langle x, e_n\rangle e_n$ on a bien $x_N\in F_N$ et $x - x_N \in F_N^{\perp}$ ce qui caractérise le projeté de x sur F_N d'après le corollaire 1.18 [30]. Ceci montre (1.16).

L'application Φ est bien définie et isométrique par (1.18). L'application définie dans (1.17) est également bien définie par la remarque 1.34, et est bien l'inverse de Φ, en particulier Φ est donc bien bijective. □

Exemple 1.35 (Séries de Fourier)**.** Dans $L^2(\mathbb{T})$, la famille $(e_n : x\mapsto e^{inx})_{n\in\mathbb{Z}}$ est une base hilbertienne. On retrouve la théorie des séries de Fourier dans $L^2(\mathbb{T})$, et le Théorème 1.32 donne le Théorème 1.32 du Chapitre VIII. D'ailleurs, si on observe bien les preuves de ces deux résultats, ce sont en fait les mêmes, l'une dans le cas particulier des séries de Fourier, l'autre dans un cadre abstrait.

28. On pourra comparer avec le sens direct de la Proposition 2.18 au Chapitre III qui donne une condition suffisante de convergence d'une série ; ici on a bien une équivalence, ce qui est beaucoup plus fort (mais bien sûr on a l'hypothèse d'orthogonalité en plus).

29. Prenez bien soin de comprendre le raisonnement ici : par densité, il existe une suite (f_N) qui converge vers x et telle que $f_N\in F_N$ pour tout $N\in\mathbb{N}$. Or $p_{F_N}(x)$ est l'élément de F_N le plus proche de x, donc cette suite converge bien aussi vers x.

30. On peut bien sûr refaire l'argument : pour tout $y\in F_N$,

$$\|x-y\|^2 = \left\|\underbrace{x-x_N}_{\in F_N^{\perp}} + \underbrace{x_N - y}_{\in F_N}\right\|^2 = \|x-x_N\|^2 + \|x_N - y\|^2 \geq \|x-x_N\|^2.$$

Polynômes orthogonaux

Etant donnés I un intervalle de $\mathbb{R}$ et $\omega : I \to]0, \infty[$ une fonction qui sera appelée poids, on travaille dans l'espace $H = L^2(I, \omega) := L^2(I, \mathcal{B}(I), \omega(x)dx) = \{f : I \to \mathbb{R}$ mesurable telle que $\int_I |f(x)|^2\omega(x)dx < +\infty\}$ qui est naturellement associé au produit scalaire :

$$\forall (f, g) \in L^2(I, \omega), \ \ \langle f, g\rangle = \int_I f(x)g(x)\omega(x)dx,$$

et dont la norme induite est :

$$\forall f \in L^2(I, \omega), \ \ \|f\|_2 = \sqrt{\int_I |f(x)|^2\omega(x)dx}.$$

On suppose que $\mathbb{R}[X] \subset L^2(I, \omega)$, c'est-à-dire [31]

$$\forall x \in \mathbb{N}, \ \ \int_I |x|^n\omega(x)dx < +\infty.$$

Définition-Proposition 1.36. Il existe une unique famille de polynômes $(P_n)_{n\in\mathbb{N}}$ réels tels que

$$\begin{cases} \forall i \neq j, \ \ \langle P_i, P_j\rangle = 0 \\ \forall n \in \mathbb{N}, \ \ \deg(P_n) = n \quad \text{et} \quad P_n \text{ est unitaire.} \end{cases}$$

On l'appelle famille de polynômes orthogonaux associée à l'espace $L^2(I, \omega)$.

Remarque 1.37. Si on note $(P_n)_{n\in\mathbb{N}}$ une famille de polynômes orthogonaux associée à l'espace $L^2(I, \omega)$. Alors la famille $(\frac{P_n}{\|P_n\|_2})$ est la famille obtenue en appliquant le procédé de Gram-Schmidt à la base canonique $(1, X, X^2, \ldots)$ de $\mathbb{R}[X]$. En effet comme $\forall n \in \mathbb{N}$, $\deg(P_n) = n$, on a bien $\forall N \in \mathbb{N}$, $\mathrm{Vect}((P_n)_{n\in[\![0,N]\!]}) = \mathbb{R}_N[X] = \mathrm{Vect}((X^n)_{n\in[\![0,N]\!]})$. Il arrive qu'on préfère appeler cette famille renormalisée la famille des polynômes orthogonaux (voir par exemple [Gou08a, page 104]).

Démonstration. Pour l'existence, on considère la famille (R_n) obtenue en appliquant le procédé de Gram-Schmidt à la base canonique $(1, X, X^2, \ldots)$ de $\mathbb{R}[X]$. La famille obtenue est orthonormale et telle que $\forall n \in \mathbb{N}, \deg(R_n) = n$. Pour la rendre unitaire, on divise R_n par son coefficient dominant.

Pour l'unicité, on considère deux familles (P_n) et (Q_n) qui conviennent. Par l'absurde, si elles ne sont pas égales, il existe $n_0 \in \mathbb{N}$ tel que $P_{n_0} \neq Q_{n_0}$, et on choisit n_0 minimal pour cette propriété. Comme $\mathrm{Vect}(Q_0, \ldots, Q_{n_0}) = \mathbb{R}_{n_0}[X]$, il existe $\lambda_0, \ldots, \lambda_{n_0} \in \mathbb{R}^{n_0+1}$ tels que

$$P_{n_0} = \sum_{k=0}^{n_0} \lambda_k Q_k = \lambda_{n_0} Q_{n_0} + \sum_{k=0}^{n_0-1} \lambda_k P_k.$$

Pour $i \in [\![0, n_0 - 1]\!]$, en prenant le produit scalaire avec P_i on obtient $\lambda_i = 0$. Ainsi $P_{n_0} = \lambda_{n_0} Q_{n_0}$, et comme les deux polynômes sont unitaires, on a en fait $\lambda_{n_0} = 1$ ce qui constitue une contradiction et conclut la preuve. □

Exemple 1.38. Citons quelques exemples de tels polynômes orthogonaux [32] :

1. Pour $I =]-1, 1[$ et $\omega : x \mapsto 1$, les polynômes orthogonaux associés sont appelés polynômes de Legendre.
2. Pour $I =]-1, 1[$ et $\omega : x \mapsto \frac{1}{\sqrt{1-x^2}}$ et $\omega : x \mapsto \sqrt{1 - x^2}$, les polynômes orthogonaux associés sont respectivement appelés polynômes de Tchebychev de première et de seconde espèce.
3. Pour $I = \mathbb{R}$ et $\omega : x \mapsto e^{-x^2}$, les polynômes orthogonaux associés sont appelés polynômes de Hermite.

31. Stricto sensu, on ne demande la propriété que pour les puissances paires de x, mais cela revient au même puisque $\forall n \in \mathbb{N}, \forall x \in \mathbb{R}, |x|^{2n+1} \leq 1 + |x|^{2n+2}$.

32. Chacune de ces familles constitue un sujet d'étude en soit, notamment on peut avoir des formules plus ou moins explicites pour les calculer ; on les rencontre également occasionnellement dans des sujets d'écrits, voir par exemple les sujets 2015, 2010, 2007 et une étude approfondie dans le sujet 1991.

4. Pour $I = \mathbb{R}_+$ et $\omega : x \mapsto e^{-x}$, les polynômes orthogonaux associés sont appelés polynômes de Laguerre.

La question qui reste en suspens est donc de savoir si la famille $(\frac{P_n}{\|P_n\|_2})_{n\in\mathbb{N}}$ constitue une base hilbertienne de $L^2(I,\omega)$, ce qui revient à se demander si $\mathrm{Vect}((P_n)_{n\in\mathbb{N}}) = \mathbb{R}[X]$ est dense dans $L^2(I,\omega)$. D'après le Corollaire 1.18 (voir la Remarque 1.19), cela revient à savoir si [33]

$$\forall f \in L^2(I,\omega), \quad \left[\forall n \in \mathbb{N}, \int_I x^n f(x)\omega(x)dx = 0 \overset{?}{\Longrightarrow} f = 0\right] \tag{1.19}$$

Il est difficile de donner des conditions nécessaires et suffisantes dans cette direction. On donne deux conditions suffisantes dans le résultat suivant :

Théorème 1.39. *1. On suppose $I =]a,b[$ avec $a < b$ deux réels, et $\omega : [a,b] \to]0,\infty[$ continue* [34]. *Alors $\mathbb{R}[X] \subset L^2(I,\omega)$ et il est dense dans $L^2(I,\omega)$.*

2. On suppose I un intervalle de $\mathbb{R}$ (non nécessairement borné), et $\omega : I \to]0,\infty[$ mesurable et telle qu'il existe $\alpha > 0$ tel que

$$\int_I e^{\alpha|x|}\omega(x)dx < +\infty. \tag{1.20}$$

Alors $\mathbb{R}[X] \subset L^2(I,\omega)$ et il est dense dans $L^2(I,\omega)$.

Remarque 1.40. Le second point implique le premier, mais la preuve du premier point via le théorème de densité de Weierstrass est très rapide, la preuve du second étant plus élaborée.

Le premier point montre que les polynômes de Legendre et les polynômes de Tchebychev de seconde espèces forment une base hilbertienne (après renormalisation). Le second point montre que les autres polynômes orthogonaux proposés précédemment (polynômes de Tchebychev de première espèce, de Hermite, de Laguerre) sont aussi des bases hilbertiennes.

Démonstration. 1. Pour tout $P \in \mathbb{R}[X]$, par le théorème des bornes, ω et P sont bornées sur $[a,b]$, donc $\int_a^b |P(x)|^2\omega(x)dt < +\infty$, c'est-à-dire $\mathbb{R}[X] \subset L^2(]a,b[,\omega)$.

Nous allons montrer directement la densité sans utiliser le critère de la remarque 1.19. Soit $f \in L^2(]a,b[,\omega)$ et $\varepsilon > 0$. Par définition on a $f\sqrt{\omega} \in L^2(]a,b[)$ [35], et par le théorème 5.38 du Chapitre IV, $C^0([a,b])$ est dense dans $L^2(]a,b[)$, donc il existe $g \in C^0([a,b])$ tel que $\|f\sqrt{\omega} - g\|_{L^2(]a,b[} \leq \varepsilon$. De plus, comme ω est continue et ne s'annule pas, $g/\sqrt{\omega}$ est continue sur $[a,b]$, donc par le théorème de Weierstrass, il existe $Q \in \mathbb{R}[X]$ tel que $\left\|Q - \frac{g}{\sqrt{\omega}}\right\|_\infty \leq \varepsilon$.

Alors

$$\begin{aligned}
\|f - Q\|_{L^2(]a,b[,\omega)} &= \|f\sqrt{\omega} - Q\sqrt{\omega}\|_{L^2(]a,b[)} \leq \|f\sqrt{\omega} - g\sqrt{\omega}\|_{L^2(]a,b[)} + \|g\sqrt{\omega} - Q\sqrt{\omega}\|_{L^2(]a,b[)} \\
&\leq \varepsilon + \|\sqrt{\omega}\|_\infty . \left\|\frac{g}{\sqrt{\omega}} - Q\right\|_{L^2(]a,b[)} \leq \varepsilon + \sqrt{b-a}\|\sqrt{\omega}\|_\infty . \left\|\frac{g}{\sqrt{\omega}} - Q\right\|_\infty \\
&\leq (1 + \sqrt{b-a}\|\sqrt{\omega}\|_\infty)\varepsilon
\end{aligned}$$

d'où la densité de $\mathbb{R}[X]$.

2. Soit $n \in \mathbb{N}$. Par croissance comparée, $\lim_{x\to\pm\infty} |x|^{2n}e^{-\alpha|x|} = 0$, donc par continuité la fonction $x \in \mathbb{R} \mapsto |x|^n e^{-\alpha|x|}$ est bornée sur $\mathbb{R}$. Ainsi

$$\int_I |x|^{2n}\omega(x)dx = \int_I |x|^{2n}e^{-\alpha|x|}\omega(x)e^{\alpha|x|}dx < +\infty$$

donc $\mathbb{R}[X] \subset L^2(I,\omega)$.

33. Attention, on écrit $f = 0$ au sens de l'espace $L^2(I,\omega)$, cela signifie en fait f est nulle presque partout sur I (il n'y a pas d'ambiguité à savoir si on veut dire presque partout pour la mesure de Lebesgue ou pour la mesure $\omega(x)dx$, puisque comme ω ne s'annule pas, ces deux propriétés sont équivalentes).
34. Ou si on préfère, $\omega :]a,b[\to \mathbb{R}$ continue et prolongeable par continuité en a et en b.
35. Quand on n'écrit pas de poids, on sous-entend que le poids est 1

Pour montrer la densité, on va cette fois montrer (1.19). Soit donc $f \in L^2(I,\omega)$ telle que $\forall n \in \mathbb{N},\ \int_I f(x)x^n\omega(x)dx = 0$. On définit

$$\varphi = \begin{cases} f\omega \text{ sur } I, \\ 0 \text{ sur } \mathbb{R}\setminus I. \end{cases}$$

On a alors $\varphi \in L^1(\mathbb{R})$, en effet [36]

$$\int_{\mathbb{R}} |\varphi(x)|dx = \int_I |f(x)|\omega(x)dx = \int_I |f(x)|\sqrt{\omega(x)}\sqrt{\omega(x)}dx \leq \sqrt{\int_I |f(x)|^2\omega(x)dx}\sqrt{\int_I \omega(x)dx}$$

par l'inégalité de Cauchy-Schwarz [37]. Cela nous permet de considérer $\widehat{\varphi}$ la transformée de Fourier de φ, et on va chercher à montrer qu'elle est nulle.

Grâce à (1.20), on peut en fait plus généralement [38] considérer $F : z \mapsto \int_{\mathbb{R}} \varphi(x)e^{zx}dx$, qui est bien définie et holomorphe sur $D_\alpha := \{z \in \mathbb{C},\ \ |\mathrm{Re}(z)| < \alpha/2\}$. En effet on a la domination [39] :

$$\forall z \in D_\alpha, \forall x \in \mathbb{R},\ \ |\varphi(x)e^{zx}| = |\varphi(x)|e^{\mathrm{Re}(z)x} \leq |\varphi(x)|e^{\frac{\alpha}{2}|x|}$$

avec $x \mapsto |\varphi(x)|e^{\frac{\alpha}{2}|x|}$ intégrable sur $\mathbb{R}$ puisque

$$\int_{\mathbb{R}} |\varphi(x)|e^{\frac{\alpha}{2}|x|}dx = \int_I |f(x)|e^{\frac{\alpha}{2}|x|}\omega(x)dx \leq \sqrt{\int_I e^{\frac{\alpha}{2}|x|}\omega(x)dx}.\sqrt{\int_I |f(x)|\omega(x)dx} < +\infty. \quad (1.21)$$

Ainsi par le Théorème d'holomorphie des intégrales à paramètres [40] (Théorème 3.42 au Chapitre **??**complexe), F est bien définie et holomorphe sur D_α, et

$$\forall n \in \mathbb{N},\ \ F^{(n)}(z) = \int_{\mathbb{R}} \varphi(x)e^{zx}x^n dx = \int_I f(x)e^{zx}x^n\omega(x)dx$$

et donc par hypothèse $F^{(n)}(0) = 0$ pour tout $n \in \mathbb{N}$, ce qui implique par le théorème de prolongement analytique 2.36 du Chapitre II que F est nulle sur tout D_α. Mais $\widehat{\varphi}$ est à une constante multiplicative près la restriction à l'axe des imaginaires purs de F (axe qui est bien dans D_α), donc finalement $\widehat{\varphi}$, et par injectivité de la transformation de Fourier (Corollaire 2.6 au Chapitre VIII), on a bien $\varphi = 0$ presque partout sur $\mathbb{R}$, c'est-à-dire $f = 0$ presque partout sur I (car ω ne s'annule pas sur I), et ceci conclut la démonstration.

□

36. Voici une variante pour justifier $\varphi \in L^1(\mathbb{R})$:

$$\int_I |f(x)|\omega(x)dx \leq \int_I \frac{1+|f(x)|^2}{2}\omega(x)dx \leq \frac{1}{2}\left(\int_I \omega(x)dx + \int_I |f(x)|^2\omega(x)dx\right) < +\infty.$$

37. On a utilisé l'inégalité de Cauchy-Schwarz dans $L^2(I)$ (sans poids), mais on aurait pu l'utiliser dans $L^2(I,\omega)$ ce qui évite cette "décomposition" $\omega = \sqrt{\omega}\sqrt{\omega}$.

38. Les lignes qui précèdent pour montrer $\varphi \in L^1(\mathbb{R})$ sont redondantes, ce qu'on montre ici étant plus fort. Pédagogiquement ce n'est pas inutile de commencer par $\varphi \in L^1(\mathbb{R})$, mais dans le cadre d'un développement et sous contrainte de temps, on peut passer cet argument.

39. On a utilisé l'inégalité simple mais commode : $\forall y \in \mathbb{R}, e^y \leq e^{|y|}$, qui implique d'ailleurs $\forall z \in \mathbb{C}, |e^z| \leq e^{|z|}$, inégalité qu'on peut également montrer directement avec la définition en série entière de l'exponentielle.

40. On peut procéder autrement et viser à justifier l'interversion suivante :

$$\int_I f(x)e^{zx}\omega(x)dx = \int_I f(x)\sum_{n\in\mathbb{N}} \frac{z^n x^n}{n!}\omega(x)dx = \sum_{n\in\mathbb{N}}\left(\int_I f(x)x^n\omega(x)dx\right)\frac{z^n}{n!} = 0.$$

Cette interversion est possible car

$$\int_I |f(x)|\sum_{n\in\mathbb{N}} \frac{|z^n x^n|}{n!}|\omega(x)|dx = \int_I |f(x)|e^{|zx|}\omega(x)dx < +\infty$$

où on utilise un calcul similaire à (1.21), et la Proposition 3.12 du Chapitre IV sur l'interversion série-intégrale.

Échantillonage de Shannon. Terminons par une piste de développement qui mélange la transformation de Fourier, les espaces de Hilbert, et les bases hilbertiennes. On pose

$$\forall x \in \mathbb{R}, \quad \operatorname{sinc}(x) = \frac{\sin(x)}{x}.$$

Théorème 1.41. *On pose $I = [-\pi, \pi]$, et*[41]

$$BL^2 := \left\{ u \in L^2(\mathbb{R}), \ \ \widehat{u} = 0 \text{ p.p. sur } \mathbb{R} \setminus I \right\}.$$

1. *L'espace BL^2 est un espace de Hilbert pour le produit scalaire usuel de $L^2(\mathbb{R})$, et toute fonction de BL^2 admet un représentant continu*[42]*.*
2. *L'application*
$$\begin{array}{rccc} \Phi : & BL^2 & \longrightarrow & \ell^2(\mathbb{Z}) \\ & u & \longmapsto & (u(n))_{n \in \mathbb{Z}} \end{array}$$
est une isométrie.
3. *La famille $(\sqrt{2\pi}\operatorname{sinc}(\pi(\cdot - n)))_{n\in\mathbb{Z}}$ est une base hilbertienne de BL^2 et pour tout $u \in BL^2$, on a*
$$u = \sum_{n\in\mathbb{Z}} u(n) \operatorname{sinc}(\pi(\cdot - n))$$
où la convergence de la série a lieu dans $L^2(\mathbb{R})$ et en norme uniforme.

Remarque 1.42. Le caractére remarquable du résultat est de pouvoir reconstruire le signal u à partir de ses seules valeurs en les entiers. Bien sûr, ceci est rendu possible par l'hypothèse $\widehat{u}$ à support dans $[-\pi, \pi]$.

Remarque 1.43. On en déduit également

$$\forall u \in BL^2, \ \ \|u\|^2_{L^2(\mathbb{R})} = \sum_{n\in\mathbb{Z}} |u(n)|^2.$$

Remarque 1.44. On peut adapter le résultat lorsque I est remplacé par un intervalle $[-a, a]$ avec $a \in]0, +\infty[$: en effet pour tout $u \in L^2(\mathbb{R})$ tel que $\operatorname{Supp}(\widehat{u}) \subset [-a, a]$, on peut montrer

$$u(x) = \sum_{n\in\mathbb{Z}} u\left(\frac{n\pi}{a}\right) \operatorname{sinc}(ax - n\pi)$$

avec convergence L^2 et uniforme. Ainsi il suffit de connaître u sur des valeurs distantes de moins de π/a pour reconstruire toutes les valeurs de u.

Pour montrer ceci, on peut soit adapter la preuve ci-dessous, soit se ramener au cas $a = \pi$ par homothétie.

Démonstration. [Wil95]

— On va montrer que BL^2 est un sev fermé de $L^2(\mathbb{R})$, il sera ainsi complet et donc un espace de Hilbert pour le produit scalaire usual de $L^2(\mathbb{R})$. Soit donc $(u_n)_{n\in\mathbb{N}}$ une suite de BL^2 qui converge en norme $L^2(\mathbb{R})$ vers $u_\infty \in L^2(\mathbb{R})$. Par continuité de la transformation de Fourier sur $L^2(\mathbb{R})$ (voir le Théorème de Plancherel 2.28 au Chapitre VIII), la suite $(\widehat{u_n})_{n\in\mathbb{N}}$ converge vers $\widehat{u_\infty}$ en norme $L^2(\mathbb{R})$, et comme $\|u_n \mathbb{1}_{\mathbb{R}\setminus I} - u_\infty \mathbb{1}_{\mathbb{R}\setminus I}\|_2 \leq \|u_n - u_\infty\|_2$, on a $\|u_\infty \mathbb{1}_{\mathbb{R}\setminus I}\|_2 = 0$ d'où $u_\infty \in BL^2$.

— Montrons maintenant que $BL^2 \subset C^0_b(\mathbb{R})$ et

$$\forall u \in BL^2, \ \ \|u\|_\infty \leq \|u\|_2. \tag{1.22}$$

41. Autrement dit, $BL^2 = \{u \in L^2(\mathbb{R}), \ \operatorname{Supp}(u) \subset I\}$.
42. On choisira donc ce représentant par défaut. Cela nous permet dans ce qui suit de parler de la valeur de $u \in BL^2$ en les entiers.

Soit $u \in BL^2$. Par le théorème de Plancherel [43], on peut appliquer la formule d'inversion du Chapitre VIII qui affirme

$$u(x) = \widehat{\widehat{u}}(-x) \text{ pour presque tout } x \in \mathbb{R}$$

mais comme $L^2(I) \subset L^1(I)$ on a $\widehat{u} \in L^1(I)$, donc $\widehat{\widehat{u}} \in C_b^0(\mathbb{R})$ (Définition/proposition 2.1 au Chapitre VIII), l'égalité précédente a lieu pour tout $x \in \mathbb{R}$ et [44]

$$\|u\|_\infty \leq \frac{1}{\sqrt{2\pi}}\|\widehat{u}\|_{L^1(I)} \leq \|\widehat{u}\|_{L^2(I)} = \|\widehat{u}\|_{L^2(\mathbb{R})}.$$

— Posons pour $n \in \mathbb{N}$, $e_n : x \in \mathbb{R} \mapsto e^{inx}\mathbb{1}_{[-\pi,\pi]}(x)$. On constate que :

$$\begin{aligned}\forall \xi \in \mathbb{R}\setminus\mathbb{Z}, \quad \widehat{e_n}(\xi) &= \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} e^{-i\xi x}e^{inx}dx = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} e^{-i\xi x}e^{inx}dx \\ &= \frac{1}{\sqrt{2\pi}}\left[\frac{e^{ix(n-\xi)}}{n-\xi}\right]_{-\pi}^{\pi} = \sqrt{2\pi}.\mathrm{sinc}(\pi(\xi-n))\end{aligned}$$

Mais on observe que $u \in BL^2 \mapsto (\widehat{u})_{|I} L^2(I)$ est une isométrie bijective, donc si on pose $\varepsilon_n = \sqrt{2\pi}.\mathrm{sinc}(\pi(\cdot - n)) = \widehat{e_n}$, alors $(\varepsilon_n)_{n\in\mathbb{Z}}$ est une base hilbertienne de BL^2 : en effet par la théorie $L^2(\mathbb{T})$ des séries de Fourier, $(e_n)_{n\in\mathbb{Z}}$ est une base hilbertienne de $L^2([-\pi,\pi])$ [45], donc on peut en déduire [46] :

$$\forall (n,m) \in \mathbb{N}^2, \quad \langle \varepsilon_n, \varepsilon_m\rangle_{L^2(\mathbb{R})} = \langle e_n, e_m\rangle_{L^2(I)} = \delta_{nm}$$

et si $v \in BL^2$ est tel que $\forall n \in \mathbb{N}$, $\langle v, \varepsilon_n\rangle_{L^2(\mathbb{R})}$, alors il existe $u \in L^2(I)$ tel que $v = (\widehat{u})_{|I}$ et alors $\forall n \in \mathbb{N}$, $\langle v, \varepsilon_n\rangle_{L^2(\mathbb{R})} = \langle u, e_n\rangle_{L^2(I)} = 0$ donc $u = 0$ (par densité de $\mathrm{Vect}((e_n)_n)$ dans $L^2(I)$), donc finalement $v = 0$, ce qui prouve bien la densité de $\mathrm{Vect}((\varepsilon_n)_n)$ dans BL^2.

— Par le théorème 1.32, on a

$$u = \sum_{n\in\mathbb{Z}} \langle u, \varepsilon_n\rangle_{L^2(\mathbb{R})}\varepsilon_n$$

avec convergence $L^2(\mathbb{R})$. Par (1.22), la convergence a aussi lieu au sens uniforme, donc on peut prendre la valeur en $n_0 \in \mathbb{Z}$, ce qui donne [47]

$$u(n_0) = \sum_{n\in\mathbb{Z}} \langle u, \varepsilon_n\rangle_{L^2(\mathbb{R})}\varepsilon_n(n_0) = \sqrt{2\pi}\langle u, \varepsilon_{n_0}\rangle_{L^2(\mathbb{R})}$$

où on utilise $\mathrm{sinc}(k\pi) = \delta_{k0}$ pour tout $k \in \mathbb{Z}$, ce qui conclut la démonstration.

□

2 Espaces H^1 et EDP en dimension 1

Dans ce paragraphe toutes les fonctions considérées seront à valeurs dans $\mathbb{R}$.

43. Attention à la subtilité ici : le théorème d'inversion “du cadre $L^1(\mathbb{R})$” a pour hypothèse $u \in L^1(\mathbb{R})$ et $\widehat{u} \in L^1(\mathbb{R})$, alors qu'ici on a $u \in L^2(\mathbb{R})$ et $\widehat{u} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$. Mais le théorème de Planchel affirme que $\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ est une bijection et que $\mathcal{F}\circ\mathcal{F} = \sigma$ où $\sigma : u \mapsto (x \mapsto u(-x))$. Ici $u \in L^2(\mathbb{R})$ donc on peut bien écrire $u(x) = \widehat{\widehat{u}}(-x)$ pour presque tout x.

44. On peut en fait en déduire beaucoup plus, à savoir que u est développable en série entière de rayon infini ; on peut aussi estimer les normes infinies des dérivées à partir de la norme $L^2(\mathbb{R})$.

45. Rappelons que par périodisation, on peut voir une fonction de $L^2([-\pi,\pi])$ comme une fonction de $L^2(\mathbb{T})$ c'est-à-dire une fonction 2π-périodique et localement intégrable.

46. On montre ici que l'image d'une base hilbertienne par une isométrie bijective (entre deux espaces de Hilbert) est une base hilbertienne.

47. Voici une autre preuve pour cette dernière égalité :

$$\langle u, \varepsilon_{n_0}\rangle_{L^2(\mathbb{R})} = \langle \widehat{u}, \widehat{\varepsilon_{n_0}}\rangle_{L^2(\mathbb{R})} = \langle \widehat{u}, \widehat{\widehat{e_{n_0}}}\rangle_{L^2(\mathbb{R})} = \langle \widehat{u}, e_{-n_0}\rangle_{L^2(\mathbb{R})} = \sqrt{2\pi}u(n_0)$$

où on a utilisé la formule d'inversion pour affirmer $\widehat{\widehat{e_{n_0}}} = e_{-n_0}$ ainsi que $\langle \widehat{u}, e_{-n_0}\rangle_{L^2(\mathbb{R})} = \sqrt{2\pi}u(n_0)$.

2.1 Définitions et étude de l'espace $H^1(I)$

Nous avons vu au paragraphe précédent l'importance des espaces de Hilbert, et nous avons vu que l'espace $L^2(I)$ est un exemple typique de tel espace. Dans ce paragraphe nous allons définir un espace qui sera là pour "remplacer" l'espace $C^1([a,b])$ des fonctions continûment dérivable par l'espace $H^1(]a,b[)$ qui aura un rôle similaire (attester des fonctions "dérivables"), mais qui en prime sera un espace de Hilbert. Il s'agit d'un espace de Sobolev, on donnera au paragraphe 2.4 quelques généralisations.

On se donne $I =]a,b[$ un intervalle ouvert de $\mathbb{R}$, non nécessairement borné, c'est-à-dire $a \in \mathbb{R}\cup\{-\infty\}$ et $b \in \mathbb{R}\cup\{\infty\}$. Nous allons voir plusieurs définitions équivalentes de l'espace $H^1(I)$:

Définition-Proposition 2.1 (Définition via les distributions)**.** L'espace de Sobolev $H^1(I)$ est l'espace des fonctions de $L^2(I)$ dont la dérivée au sens des distributions est représentable par une fonction de $L^2(I)$, c'est à dire

$$H^1(I) = \{u \in L^2(I), \;\; u' \in L^2(I)\}.$$

C'est un espace vectoriel, et on peut le munir du produit scalaire :

$$\forall (u,v) \in H^1(I), \;\; \langle u,v\rangle_{H^1} = \int_I [u(t)v(t) + u'(t)v'(t)]\,dt. \tag{2.1}$$

Remarque 2.2. Ainsi la norme H^1 est

$$\forall u \in H^1(I), \;\; \|u\|_{H^1} = \sqrt{\int_I [u(t)^2 + u'(t)^2]\,dt}.$$

De plus, si (u_n) est une suite de $H^1(I)$ et $u \in H^1(I)$, alors

$$u_n \xrightarrow[n\to+\infty]{H^1} u \quad \Leftrightarrow \left[u_n \xrightarrow[n\to+\infty]{L^2} u \;\text{ et }\; u'_n \xrightarrow[n\to+\infty]{L^2} u'\right]$$

Démonstration. Pour montrer que $H^1(I)$ est un espace vectoriel, nous montrons que c'est un sous-espace vectoriel de $L^2(I)$: en effet $0 \in H^1(I)$, et pour tout $(u,v) \in H^1(I)$ et $\lambda \in \mathbb{R}$, $(\lambda u + v) \in L^2(I)$ et $(\lambda u+v)' = \lambda u' + v' \in L^2(I)$, et donc $(\lambda u + v) \in H^1(I)$.

Il faut vérifier que (2.1) est bien défini et est un produit scalaire. Pour $(u,v) \in H^1(I)$, on a d'après l'inégalité de Cauchy-Schwarz dans $L^2(I)$ que $(uv) \in L^1(I)$ et $(u'v') \in L^1(I)$ donc (2.1) est bien définie. On observe que l'application définie par (2.1) est bilinéaire symétrique et positive, et si $\langle u,u\rangle_{H^1} = 0$ pour $u \in H^1(I)$, alors en particulier $\|u\|_{L^2} = 0$ donc u est nulle en tant que classe d'équivalence de $L^2(I)$. □

Remarque 2.3. Ici u' est compris au sens des distributions, et signifie précisément, avec les notations de l'exemple 3.4 au Chapitre VIII :

$$\exists g \in L^2(I), \quad [u]' = [g]$$

c'est-à-dire qu'il existe $g \in L^2(I)$ telle que

$$\forall \varphi \in C_c^\infty(I), \quad \langle [u]',\varphi\rangle := -\int_I u(t)\varphi'(t) = \int_I g(t)\varphi(t)dt = \langle [g],\varphi\rangle$$

où $\langle\cdot,\cdot\rangle$ désigne le crochet de dualité dans $\mathcal{D}'(I)\times\mathcal{D}(I)$.

La remarque précédente permet d'aboutir à la définition suivante, qui traduit la précédente sans faire référence aux distributions ou à la dérivation au sens des distributions :

Définition-Proposition 2.4 (La même définition mais sans distributions)**.** L'espace de Sobolev $H^1(I)$ est l'espace des fonctions de $L^2(I)$ telles qu'il existe $g \in L^2(I)$ telle que

$$\forall \varphi \in C_c^\infty(I), \quad \int_I u(t)\varphi'(t) = -\int_I g(t)\varphi(t)dt.$$

Une telle fonction g est alors unique (modulo l'égalité presque partout), et on notera alors $u' = g$ et qu'on appelle dérivée faible [48].

48. On peut aussi l'appeler dérivée au sens des distributions, mais l'objectif de cette définition est justement de ne pas faire référence aux distributions.

Démonstration. Il faut seulement vérifier l'unicité de g. C'est une conséquence du Corollaire 5.61 du Chapitre IV. □

Remarque 2.5. Les fonctions de $H^1(I)$ sont donc des fonctions de $L^2(I)$ pour lesquelles on a abstraitement une intégration par parties faisant intervenir une fonction $L^2(I)$ qui joue le rôle de la dérivée de u, quand bien même on n'a pas supposé de dérivabilité "classique" de u.

Exemple 2.6. Si on suppose $I =]a, b[$ borné, alors les fonctions de classe C^1 sur $[a, b]$ [49] sont dans $H^1(I)$: en effet elles sont dans $L^2(I)$ (car elles sont bornées sur $[a, b]$ puisqu'elles sont continues sur le compact $[a, b]$), et leur dérivée au sens classique est bien dans $L^2(I)$ (elle est continue sur $[a, b]$) et satisfait bien

$$\forall \varphi \in C_c^\infty(I), \quad \int_I u(t)\varphi'(t) = -\int_I u'(t)\varphi(t)dt.$$

En l'occurence la dérivée faible coïncide avec la dérivée classique de u.

Plus généralement, les fonctions qui sont continues sur $[a, b]$ et C^1 par morceaux sur $[a, b]$ sont dans $H^1(I)$, car la formule d'intégration par parties reste valable pour ces fonctions (et que les fonctions continues par morceaux sur $[a, b]$ sont dans $L^2(]a, b[)$. Dans ce cas la dérivée faible est égale à la dérivée classique là où elle est définie, c'est-à-dire sur $[a, b]$ privé des points où u' n'est pas dérivable (qui est un ensemble fini, donc de mesure nulle).

Définition-Proposition 2.7 (Une autre définition équivalente via les primitives)**.** L'espace $H^1(I)$ est l'ensemble des fonctions $u \in L^2(I)$ telles qu'il existe $c \in \mathbb{R}$ et $g \in L^2(I)$ telles que

$$u(x) = c + \int_{x_0}^x g(t)\,dt \quad \text{pour presque tout } x \in I, \tag{2.2}$$

où $x_0 \in I$ est un point donné. Cette définition ne dépend pas du choix de x_0, et la fonction g est à la fois indépendante de x_0 et unique [50], on la notera donc u'. [51]

Remarque 2.8. L'intégrale dans (2.2) est comprise au sens de

$$\int_{x_0}^x g(t)\,dt = \begin{cases} \int_{[x_0,x]} g(t)\,dt & \text{si } x \geq x_0, \\ -\int_{[x,x_0]} g(t)\,dt & \text{si } x < x_0, \end{cases}$$

où les intégrales de droite sont des intégrales de Lebesgue. Notons également que ces expressions ont bien un sens pour $g \in L^2(I)$, car alors $g \in L^1_{\text{loc}}(I)$ s'intègre sur tout compact.

Remarque 2.9. Une conséquence importante de la définition-proposition 2.7 est la suivante : si $u \in H^1(I)$ est telle que $u' \in C^0(\overline{I})$ (que ce u' soit compris via la définition 2.1 ou la définition 2.7) alors nécessairement $u \in C^1(\overline{I})$. En effet la formule (2.2) montre que u a un représentant de classe C^1 sur $\overline{I}$.

Démonstration.
1. Commençons par vérifier que la définition 2.2 ne dépend pas du choix de x_0. Si $u \in L^2(I)$ est tel que

$$u(x) = c + \int_{x_0}^x g(t)\,dt$$

pour presque tout x avec $g \in L^2(I)$ et $x_0 \in I$, alors pour $x_1 \in I$ on a par la relation de Chasles

$$u(x) = \left(c + \int_{x_0}^{x_1} g(t)\,dt\right) + \int_{x_1}^x g(t)\,dt, \quad \text{pour presque tout } x \in I,$$

et donc on retrouve bien (2.2) avec la constante $c + \int_{x_1}^{x_0} g_1(t)\,dt$.

---

49. Attention, on prend bien la peine ici d'avoir des fonctions définies et C^1 jusqu'en a et b. Si on considère des fonctions de $C^1(]a, b[)$ alors elles ne sont pas forcément de carré intégrable au voisinage de a et de b. On écrit donc bien $C^1([a, b]) \subset H^1(]a, b[)$. Néanmoins, si $\varphi \in C^1(]a, b[)$ et qu'on suppose $(\varphi, \varphi') \in L^2(I)^2$ alors bien sûr $\varphi \in H^1(I)$.

50. Il y a aussi unicité de la constante c pour un x_0 donné, par contre la constante c dépend de x_0.

51. On ne le précise pas mais avec cette définition on peut également voir que $H^1(I)$ est un espace vectoriel et qu'on peut le munir du produit scalaire (2.1).

2. Montrons à la fois que g ne dépend pas de x_0 et est unique : On suppose
$$u(x) = c_0 + \int_{x_0}^{x} g_0(t)dt = c_1 + \int_{x_1}^{x} g_1(t)dt, \text{ pour presque tout } x \in I,$$
pour $(c_0, c_1) \in \mathbb{R}^2$, $(x_0, x_1) \in I^2$ et $(g_0, g_1) \in L^2(I)^2$. Alors en appliquant à nouveau la relation de Chasles, on en déduit
$$\int_{x_0}^{x} \Big[g_1(t) - g_0(t)\Big]dt = C \text{ pour presque tout } x \in I$$
où $C \in \mathbb{R}$ est une constante. Nous allons montrer que ceci implique $g_0 = g_1$ presque partout. Classiquement, le membre de gauche de cette égalité est une fonction continue de x [52], et donc l'égalité précédente est en fait valable pour tout $x \in I$. On en déduit avec la relation de Chasles :
$$\forall (x, y) \in I \times I \text{ tels que } y \geq x, \int_I [g_1 - g_0](t)\mathbb{1}_{[x,y]}(t)\, dt = 0.$$
or toute fonction en escalier à support compact dans I est combinaison linéaire de fonctions caractéristiques d'intervalles de la forme $[x, y]$ avec $(x, y) \in I$ et $x \leq y$, donc pour toute fonction en escalier à support compact φ,
$$\int_I (g_1 - g_0)(t)\varphi(t)\, dt = 0. \tag{2.3}$$
Toute fonction continue à support compact sur I est limite uniforme d'une suite de fonctions en escalier à support dans le même compact, donc (2.3) est en fait valable pour tout $\varphi \in C_c^0(I)$. Le corollaire 5.61 du Chapitre IV permet alors de conclure que $g_0 = g_1$ presque partout, d'où l'unicité attendue.
3. Nous montrons maintenant l'équivalence avec la définition 2.4 : soit $u \in L^2(I)$ tel qu'il existe $c \in \mathbb{R}$, $x_0 \in I$ et $g \in L^2(I)$ tel que
$$u(x) = c + \int_{x_0}^{x} g(t)dt \text{ pour presque tout } x \in I.$$
Montrons alors que la dérivée au sens des distributions de u est $g \in L^2(I)$ [53] : pour cela, on écrit :
$$\forall \varphi \in C_c^\infty(I), \ \ \langle u', \varphi\rangle_{\mathcal{D}'\times\mathcal{D}} = -\int_I u(t)\varphi'(t)dt = -\int_I \left(\int_{x_0}^{t} g(s)ds + c\right)\varphi'(t)dt.$$
La constante c disparait car $\int_I \varphi'(t)dt = 0$. Ensuite on écrit, en notant $I =]a, b[$,
$$\begin{aligned} -\int_I \left(\int_{x_0}^{t} g(s)ds\right)\varphi'(t)dt &= \int_a^{x_0}\int_a^{x_0} \mathbb{1}_{[t,x_0]}(s)g(s)\varphi'(t)dsdt - \int_{x_0}^{b}\int_{x_0}^{b} \mathbb{1}_{[x_0,t]}(s)g(s)\varphi'(t)dsdt \\ &= \int_a^{x_0}\int_a^{x_0} \mathbb{1}_{]a,s]}(t)\varphi'(t)dtg(s)ds - \int_{x_0}^{b}\int_{x_0}^{b} \mathbb{1}_{[s,b[}(t)\varphi'(t)dtg(s)ds \\ &= \int_I \varphi(s)g(s)ds \end{aligned}$$
où on a appliqué le théorème de Fubini : ceci se justifie car φ est à support compact, donc il existe $K \subset I$ intervalle compact, dont on peut supposer qu'il contient x_0 et tel que $\varphi = 0$ sur $I \setminus K$, et alors
$$\int_a^{x_0}\int_a^{x_0} \left|\mathbb{1}_{[t,x_0]}(s)g(s)\varphi'(t)\right| dsdt \leq \int_a^{x_0}\int_a^{x_0} |\mathbb{1}_K(s)g(s)\varphi'(t)|\, dsdt \leq \|\varphi'\|_{L^1(I)}\|g\|_{L^1(K)}$$
or $g \in L^1_{loc}(I)$ (voir l'exercice 5.18 au Chapitre IV).

52. Le faire en exercice, par exemple avec le théorème de convergence dominée. Cet argument est valable sous la seule hypothèse que $f \in L^1_{loc}(I)$, mais si on a de plus $f \in L^2(I)$ comme c'est le cas ici, on a même que la fonction $x \mapsto \int_{x_0}^{x} f(t)dt$ est uniformément continue, et même 1/2-hölderienne : en effet
$$\forall (x, y) \in I^2, \ \left|\int_{x_0}^{x} f(t)dt - \int_{x_0}^{y} f(t)dt\right| \leq \left|\int_x^y |f(t)|dt\right| \leq \sqrt{\int_I f^2(t)dt}\sqrt{|x - y|}.$$

53. Il s'agit en fait de l'exercice 3.19 au Chapitre VIII.

4. Réciproquement, supposons que $u \in L^2(I)$ soit tel que pour tout $\varphi \in C_c^\infty(I)$ on ait

$$\int_I u\varphi'\, dx = -\int_I g\varphi\, dx$$

pour un certain $g \in L^2(I)$, c'est-à-dire que la dérivée au sens des distributions de u est g. Posons alors

$$v(x) = \int_{x_0}^{x} g(t)\, dt$$

pour tout $x \in I$, où $x_0 \in I$ est fixé. Nous avons vu au point précédent que la dérivée au sens des distributions de v est g, donc par le lemme 3.17 du Chapitre VIII, il existe $c \in \mathbb{R}$ tel que $u = c + v$, d'où le résultat.

□

Proposition 2.10. *L'espace $(H^1(I), \langle\cdot,\cdot\rangle_{H^1})$ est un espace de Hilbert.*

Démonstration. D'après les définitions-propositions 2.1 ou 2.7 (voir la note 51 dans ce second cas), il reste à montrer que $H^1(I)$ est complet : soit $(u_n)_{n\in\mathbb{N}}$ une suite de Cauchy dans $H^1(I)$. De la définition de cette norme, il résulte que les suites $(u_n)_{n\in\mathbb{N}}$ et $(u_n')_{n\in\mathbb{N}}$ sont de Cauchy dans $L^2(I)$. Ce dernier est complet par le Théorème de Riesz-Fischer 5.22 du Chapitre IV, donc il existe $u \in L^2(I)$ et $g \in L^2(I)$ telles que $u_n \to u$ et $u_n' \to g$ dans $L^2(I)$. Il reste à conclure que $u' = g$ afin d'avoir $u_n \to u$ dans l'espace $H^1(I)$ (voir la Remarque 2.2), ce qu'on peut faire de deux méthodes différentes suivant qu'on parte de la définition 2.1 ou de la définition 2.7 :

1. Avec la définition 2.1, on peut se donner $\varphi \in C_c^\infty(I)$, et comme

$$\forall n \in \mathbb{N},\quad \int_I u_n'(t)\varphi(t)dt = -\int_I u_n(t)\varphi'(t)dt,$$

on peut passer à la limite pour obtenir $\int_I g(t)\varphi(t)dt = -\int_I u(t)\varphi'(t)dt$, et ce grâce aux inégalités

$$\forall n \in \mathbb{N},\quad \left|\int_I u_n'(t)\varphi(t)dt - \int_I g(t)\varphi(t)dt\right| \le \|\varphi\|_{L^2(I)}\|u_n' - g\|_{L^2(I)},$$
$$\text{et}\quad \left|\int_I u_n(t)\varphi'(t)dt - \int_I u(t)\varphi'(t)dt\right| \le \|\varphi'\|_{L^2(I)}\|u_n - u\|_{L^2(I)}.$$

Ceci [54] montre que $u' = g$ et conclut la preuve.

2. Partons maintenant de la définition 2.7. Pour $x_0 \in I$, on a maintenant l'existence de $(c_n) \in \mathbb{R}^{\mathbb{N}}$ tels que [55]

$$u_n(x) = c_n + \int_{x_0}^{x} u_n'(t)\, dt, \quad \text{pour presque tout } x \in I. \tag{2.4}$$

Comme I est un ouvert contenant x_0, il existe $\alpha > 0$ tel que $J_\alpha = [x_0 - \alpha, x_0 + \alpha] \subset I$. On intègre (2.4) sur cet intervalle et on en déduit que

$$c_n = \frac{1}{2\alpha}\Big(\int_{J_\alpha} u_n(x)\, dx - \int_{J_\alpha}\int_{x_0}^{x} u_n'(t)\, dtdx\Big),$$

puis avec l'inégalité de Cauchy-Schwarz,

$$\forall (p,q) \in \mathbb{N}^2,\quad |c_q - c_p| \le \frac{1}{2\alpha}\big(\sqrt{2\alpha}\|u_q - u_p\|_{L^2(I)} + (2\alpha)^{3/2}\|u_q' - u_p'\|_{L^2(I)}\big).$$

La suite $(c_n)_{n\in\mathbb{N}}$ est en particulier de Cauchy donc converge vers $c \in \mathbb{R}$ par complétude de $\mathbb{R}$.

54. On vient en fait de démontrer que la convergence L^2 implique la convergence au sens des distributions (ce qui n'est pas surprenant car la convergence au sens des distributions est une convergence "très faible").

55. Attention, ici u_n' est pris au sens de la définition 2.7. Il se trouve qu'il s'agit effectivement de la dérivée au sens des distributions de u_n, puisqu'on a montré que les définitions 2.1 et 2.7 coïncident, mais on montre ici la propriété de complétude en ignorant ce fait.

Toujours par l'inégalité de Cauchy-Schwarz, pour tout $x \in I$,

$$\left| \int_{x_0}^{x} u_n'(t)\,dt - \int_{x_0}^{x} g(t)\,dt \right| \leq \sqrt{|x - x_0|}\|u_n' - g\|_{L^2(I)} \xrightarrow[n\to\infty]{} 0.$$

On en déduit donc que

$$c_n + \int_{x_0}^{x} u_n'(t)\,dt \to c + \int_{x_0}^{x} g(t)\,dt$$

pour tout $x \in I$. D'après le corollaire 5.24 du Chapitre IV (réciproque partielle du théorème de convergence dominée), il existe σ une extraction telle que $u_{\sigma(n)} \to u$ presque partout dans I. Comme (2.4) a lieu pour tout n en dehors d'un ensemble de mesure nulle, [56] on en déduit que pour presque tout $x \in I$,

$$u(x) = c + \int_{x_0}^{x} g(t)\,dt,$$

ce qui montre bien que $u \in H^1(I)$ avec $u' = g$, et donc par définition $u_n \to u$ dans $H^1(I)$.

□

On a vu à l'exemple 2.6 que $C^1([a,b]) \subset H^1(]a,b[)$, qui fournit une condition suffisante pour être dans $H^1(]a,b[)$. Voyons maintenant une condition nécessaire, qui se traduit pas une propriété de régularité des fonctions de $H^1(I)$:

Théorème 2.11. *Soit $u \in H^1(I)$. Alors il existe u admet un représentant continu sur $\overline{I}$, c'est-à-dire qu'il existe*[57] *$\widetilde{u} \in C^0(\overline{I})$ tel que $\widetilde{u} = u$ presque partout sur I. Si de plus I est un intervalle borné, alors il existe une constante $C > 0$ telle que, pour tout $u \in H^1(I)$,*

$$\|\widetilde{u}\|_\infty \leq C\|u\|_{H^1(I)}.$$

Remarque 2.12. On est donc amené à écrire $H^1(I) \subset C^0(\overline{I})$ [58]. Ceci est un abus (car a priori $H^1(I)$ est constitué de classes d'équivalence de fonctions), mais qui se justifie car il y a en fait une injection de $H^1(I)$ dans $C^0(\overline{I})$ via l'application $u \mapsto \widetilde{u}$.

En pratique, quand on considèrera une classe de fonctions de $H^1(I)$, on considèrera toujours son représentant continu, et donc on notera u à la place de $\widetilde{u}$.

Par exemple, la formule (2.2) devient

$$\forall x \in I, \quad u(x) = u(x_0) + \int_{x_0}^{x} u'(t)dt.$$

Démonstration. Soit $u \in H^1(I)$. On utilise la définition-proposition 2.7, qui affirme que [59]

$$u(x) = c + \int_{x_0}^{x} u'(t)dt \quad \text{pour presque tout } x \in I.$$

On peut poser $\widetilde{u}(x) = c + \int_{x_0}^{x} u'(t)dt$. Cette fonction est bien continue sur I (voir la note 52).

Il reste à voir que $\widetilde{u}$ se prolonge par continuité sur $\overline{I}$, c'est-à-dire en a si $a > -\infty$ et en b si $b < +\infty$. Dans le cas $b < +\infty$ par exemple, il suffit de voir que $u' \in L^2([x_0, b[) \subset L^1([x_0, b[)$ donc on a bien

$$\widetilde{u}(x) \xrightarrow[x\to b]{} c + \int_{x_0}^{b} u'(t)dt,$$

56. A priori cet ensemble de mesure nulle dépend de n, mais comme on l'a expliqué à la remarque 3.9 au Chapitre IV, comme une union dénombrable d'ensembles de mesure nulle est encore de mesure nulle, on peut trouver un tel ensemble qui ne dépend pas de n.

57. Ce représentant est bien sûr unique, puisque si deux fonctions continues sont égales presque partout, elles sont égales partout.

58. Ce résultat est un cas particulier de ce qu'on appelle les injections de Sobolev, qui consistent à trouver la régularité la plus fine qu'on peut obtenir sur les fonctions des espaces de Sobolev.

59. Soit on part de la définition (2.1) et on considère ceci comme une proposition, soit on part de la définition 2.7 et alors il s'agit simplement de la définition de $u \in H^1(I)$.

et de même en a si $a > -\infty$.[60]

Supposons maintenant I borné et notons $|I|$ sa longueur. Par la relation de Chasles, on a

$$\forall (x,y) \in I^2, \widetilde{u}(x) = \widetilde{u}(y) + \int_y^x u'(t)\, dt, \quad \text{d'où} \quad |\widetilde{u}(x)| \leq |\widetilde{u}(y)| + \int_I |u'(t)|\, dt.$$

Avec l'inégalité classique $(a+b)^2 \leq 2(a^2+b^2)$ et l'inégalité de Cauchy-Schwarz, on en déduit

$$\forall (x,y) \in I^2, \;\; |\widetilde{u}(x)|^2 \leq 2|\widetilde{u}(y)|^2 + 2\Big(\int_I |g(t)|\, dt\Big)^2 \leq 2|\widetilde{u}(y)|^2 + 2|I|^2 \|g\|^2_{L^2(I)}.$$

On intègre cette inégalité par rapport à y et on obtient

$$\forall x \in I, \;\; |I|.|\widetilde{u}(x)|^2 \leq 2\|u\|^2_{L^2(I)} + 2|I|^2\|u'\|^2_{L^2(I)}.$$

d'où le résultat. □

Remarque 2.13 (Injection compacte)**.** Dans le cas où $I =]a,b[$ est borné, on peut en fait dire que l'injection $H^1(]a,b[) \subset C^0([a,b])$ est compacte. Cela signifie que si (u_n) est une suite bornée dans $H^1(]a,b[)$, c'est-à-dire qu'il existe $M \in \mathbb{R}_+$ tel que $\|u_n\|_{H^1(I)} \leq M$ pour tout $n \in \mathbb{N}$, alors il existe une suite extraite $(u_{\sigma(n)})_{n\in\mathbb{N}}$ et une fonction $u \in C^0([a,b]))$ telles que $u_{\sigma(n)} \to u$ uniformément sur $[a,b]$.

En effet, d'après le Théorème 2.11, la suite $(u_n)_{n\in\mathbb{N}}$ est bornée dans $C^0([a,b])$, et par ailleurs, comme à la note 52, l'inégalité de Cauchy-Schwarz donne

$$|u_n(x) - u_n(y)| \leq \sqrt{|x-y|}\|u_n'\|_{L^p(I)} \leq M|x-y|^{1/2},$$

et on en déduit que la suite $(u_n)_{n\in\mathbb{N}}$ est uniformément équicontinue. La conclusion suit donc d'une application immédiate du Théorème d'Ascoli 3.7 du Chapitre III.

On conclut ce paragraphe par deux derniers résultats sur l'espace $H^1(I)$:

Théorème 2.14. *Soit I un intervalle borné. Alors $C^\infty(\bar{I})$ est dense dans $H^1(I)$.*

Démonstration. Soit $u \in H^1(I)$. D'après la définition-proposition 2.7, il existe $c \in \mathbb{R}$ tel que pour tout[61] $x \in I$,

$$u(x) = c + \int_{x_0}^x u'(t)\, dt.$$

D'après le théorème 5.45 du Chapitre IV, il existe une suite $(g_n)_{n\in\mathbb{N}}$ de fonctions $C_c^\infty(I)$ telle que $g_n \to u'$ dans $L^2(I)$. On pose alors pour tout $x \in I$,

$$u_n(x) = c + \int_{x_0}^x g_n(t)\, dt$$

de sorte que $u_n \in C^\infty(\bar{I})$. De plus, par construction, $u_n' = g_n \to u'$ dans $L^2(I)$ et par l'inégalité de Cauchy-Schwarz, on a pour tout $x \in I$,

$$|u_n(x) - u(x)| \leq \int_I |u_n'(t) - u'(t)|\, dt \leq \sqrt{|I|}\|u_n' - u'\|_{L^2(I)},$$

si bien que $u_n \to u$ uniformément, donc à plus forte raison dans $L^2(I)$ puisque I est borné. On en déduit bien que $u_n \to u$ dans $H^1(I)$. □

60. Notons qu'on pouvait argumenter différemment : en effet d'après la note 52 la fonction $x \mapsto \int_{x_0}^x u'(t)dt$ est uniformément continue, elle admet donc bien un prolongement par continuité sur $\bar{I}$ (voir l'exemple 1.89, application du Théorème 1.87 au Chapitre III).

61. Voir la remarque 2.12.

Théorème 2.15 (Formule de Leibnitz et intégration par parties)**.** *Soit I un intervalle ouvert, et (u, v) deux fonctions de $H^1(I)$. Alors $uv \in H^1(I)$ avec $(uv)' = u'v + uv'$* [62]*. De plus,* [63]

$$\forall (x, y) \in I^2, \quad \int_x^y u'(t)v(t)\,dt = -\int_x^y u(t)v'(t)\,dt + u(y)v(y) - u(x)v(x).$$

Démonstration. On commence par remarquer que par le théorème 2.11, la fonction u est bornée, et donc $(uv) \in L^2(I)$. On se donne $[a, b]$ compact inclus dans I. D'après le théorème 2.14, il existe des suites $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ de fonctions $C^\infty([a, b])$ telles que $u_n \to u_{|]a,b[}$ et $v_n \to v_{|]a,b[}$ dans $H^1(]a, b[)$. Par suite, le théorème 2.11 montre que $u_n \to u_{|[a,b]}$ uniformément, donc $u_n v_n \to (uv)_{|]a,b[}$ dans $L^2(a, b)$. De même, comme $u_n \to u_{|]a,b[}$ uniformément et $v_n' \to v'_{|]a,b[}$ dans $L^2(a, b)$, on a que $u_n v_n' \to (uv')_{|]a,b[}$ dans $L^2(a, b)$, et de même $u_n' v_n \to (u'v)_{|]a,b[}$ dans $L^2(a, b)$. Or par la formule de Leibnitz dans le cas classique, $(u_n v_n)' = u_n' v_n + u_n v_n'$, donc $(u_n v_n)' \to (u'v + uv')_{|]a,b[}$ dans $L^2(a, b)$. On en déduit que la suite $(u_n v_n)$ est de Cauchy dans $H^1(]a, b[)$ qui est complet (Théorème 2.10), donc elle converge dans $H^1(]a, b[)$ vers une limite qui ne peut être que $(uv)_{|]a,b[}$, qui est sa limite dans $L^2(a, b)$. Ainsi $(uv) \in H^1(]a, b[)$ avec $(uv)' = u'v + uv'$ sur $]a, b[$. Comme de plus u et v sont bornées sur I, on a en fait $u'v + uv' \in L^2(I)$, donc finalement $(uv) \in H^1(I)$ et $(uv)' = u'v + uv'$.

De plus, d'après la formule d'intégration par parties usuelle, on a pour tout $n \in \mathbb{N}$,

$$\int_a^b u_n'(t)v_n(t)\,dt = -\int_a^b u_n(t)v_n'(t)\,dt + u_n(b)v_n(b) - u_n(a)v_n(a).$$

D'une part $u_n(b)v_n(b) - u_n(a)v_n(a) \to u(b)v(b) - u(a)v(a)$, et d'autre part $u_n v_n' \to uv'$ dans $L^2(a, b)$, ce qui donne par l'inégalité de Cauchy-Schwarz $\int_a^b u_n v_n'\,dt \to \int_a^b uv'\,dt$. De même $\int_a^b u_n' v_n\,dt \to \int_a^b u'v\,dt$, on obtient donc le résultat en passant à la limite quand $n \to +\infty$. □

2.2 Espace $H_0^1(]a, b[)$

Définition 2.16. Soit I un intervalle ouvert de $\mathbb{R}$. On définit $H_0^1(I)$ comme l'adhérence de l'espace des fonctions $C_c^\infty(I)$ dans $H^1(I)$.

Par construction, $H_0^1(I)$ est un sous-espace fermé dans $H^1(I)$. En particulier $H_0^1(I)$ est un espace de Hilbert pour le produit scalaire de $H^1(I)$.

L'énoncé suivant fournit une caractérisation de $H_0^1(I)$ dans le cas où I est borné.

Théorème 2.17. *Soit $a < b$ deux réels et $u \in H^1(]a, b[)$. Alors*

$$u \in H_0^1(]a, b[) \quad \Leftrightarrow \quad u(a) = u(b) = 0.$$

Remarque 2.18. Autrement dit, on a la définition équivalente (dans le cas $-\infty < a < b < +\infty$) :

$$H_0^1(]a, b[) = \{u \in H^1(]a, b[), u(a) = u(b) = 0\}.$$

Remarque 2.19. Dans le cas $I = \mathbb{R}$, on a en fait $H_0^1(\mathbb{R}) = H^1(\mathbb{R})$, c'est-à-dire que $C_c^1(\mathbb{R})$ est dense dans $H^1(\mathbb{R})$, voir par exemple [Bré05].

Démonstration. Soit $u \in H_0^1(]a, b[)$. Par définition, il existe alors une suite $(u_n)_{n\in\mathbb{N}}$ de fonctions $C_c^\infty(]a, b[)$ telle que $u_n \to u$ dans $H^1(]a, b[)$. Comme $\mathrm{supp}(u_n) \subset]a, b[$ pour tout $n \in \mathbb{N}$, on en déduit que $u_n(a) = u_n(b) = 0$ et comme d'après le Théorème 2.11, (u_n) converge uniformément vers u sur $[a, b]$, on a bien $u(a) = u(b) = 0$.

Réciproquement [64], soit $u \in H^1(]a, b[)$ telle que $u(a) = u(b) = 0$. Alors pour tout $x \in [a, b], u(x) = \int_a^x u'(t)dt$ et en particulier $\int_a^b u'(t)dt = 0$. Par le théorème 5.45 du Chapitre IV il existe une suite

62. Autrement dit, la formule de Leibnitz classique vaut aussi pour les dérivées faibles dans l'espace H^1. Attention, une telle règle ne peut avoir de sens en général dans les distributions, pour la bonne raison qu'on ne peut pas toujours définir facilement une multiplication entre deux distributions. Ici tous les produits qui interviennent sont des produits de fonctions

63. Attention, ici on utilise sans le dire le théorème 2.11. Rappelons en effet que si $u \in L^2(I)$, on ne peut pas écrire $u(x)$ pour un x fixé, car cette valeur dépend du représentant de u. Mais pour $u \in H^1(I)$, alors $u(x)$ est la valeur en x du représentant continu de x.

64. Cette preuve est tirée de [HL99, page 305]. On peut trouver un argument différent dans [Bré05, Théorème 8.12].

$g_n \in C_c^\infty(]a,b[)$ qui converge vers u' dans $L^2(a,b)$. On introduit une fonction $\theta \in C_c^\infty(]a,b[)$ telle que $\int_a^b \theta(t)dt = 1$, et on pose [65]

$$u_n(x) = \int_a^x g_n(t)dt - \left(\int_a^b g_n(t)dt\right) \int_a^x \theta(t)dt.$$

Ces fonctions sont bien C^∞ et à support compact dans $]a,b[$, elles convergent uniformément vers u (donc dans $L^2(a,b)$), et pour tout $n \in \mathbb{N}$, $u_n' = g_n - \left(\int_a^b g_n(t)dt\right)\theta(x)$ qui converge vers u' dans $L^2(a,b)$ puisque $\int_a^b g_n(t)dt \to \int_a^b u'(t)dt = 0$, ce qui conclut la preuve. □

Proposition 2.20 (Inégalité de Poincaré). *Soit $I =]a,b[$ un intervalle borné de $\mathbb{R}$.*

i) Pour tout $u \in H_0^1(I)$,

$$\|u\|_{L^2(I)} \leq \frac{b-a}{\sqrt{2}}\, \|u'\|_{L^2(I)}.$$

ii) Sur $H_0^1(I)$

$$\|u\|_{H_0^1(I)} = \sqrt{\int_a^b |u'(t)|^2\, dt} = \|u'\|_{L^2(a,b)} \tag{2.5}$$

définit une norme qui est équivalente à la norme $H^1(I)$ [66].

Démonstration. i) Soit $u \in H_0^1(I)$. On a pour tout $x \in [a,b]$, $u(x) = \int_a^x u'(t)\, dt$ donc par l'inégalité de Cauchy-Schwarz,

$$|u(x)| \leq \sqrt{|x-a|}\|u'\|_{L^2(I)}. \tag{2.7}$$

En élevant au carré puis en intégrant en $x \in [a,b]$, il vient

$$\int_a^b |u(t)|^2\, dt \leq \frac{(b-a)^2}{2}\|u'\|^2_{L^2(I)}$$

d'où le résultat.

ii) La formule (2.5) est bien positivement homogène, positive, et satisfait l'inégalité triangulaire. Il reste à voir qu'elle est définie positive. Si $u \in H_0^1(]a,b[)$ est tel que $\|u'\|_{L^2(a,b)} = 0$, alors $u' = 0$ presque partout, et par suite $u(x) = \int_a^x u'(t)dt$ est nul pour tout x [67]. Ainsi (2.5) est bien une norme. De plus pour tout $u \in H_0^1(I)$, on a d'une part $\|u\|_{H_0^1(I)} \leq \|u\|_{H^1(I)}$, et d'autre part par l'inégalité de Poincaré on a également

$$\|u\|^2_{H^1(I)} = \|u\|^2_{L^2(I)} + \|u'\|^2_{L^2(I)} \leq \Big(\frac{(b-a)^2}{2} + 1\Big)\|u'\|^2_{L^2(I)},$$

ce qui conclut à l'équivalence des normes.

□

Remarque 2.21. Notons que l'inégalité de Poincaré et les résultats d'équivalence de norme ne nécessitent l'annulation de u qu'à une seule des extrémités de I.

Remarque 2.22. On dispose donc de deux produits scalaires différents sur $H_0^1(I)$,

$$\langle u,v\rangle_{H^1(I)} = \int_a^b (uv + u'v')\, dt \text{ et } \langle u,v\rangle_{H_0^1(I)} = \int_a^b u'v'\, dt.$$

Ces deux produits scalaires induisent des normes équivalentes, donc la même structure topologique. Néanmoins, ce sont deux structures hilbertiennes différentes : elles ne voient pas les mêmes angles droits. Deux éléments de $H_0^1(I)$ peuvent en effet être orthogonaux pour l'un des produits scalaires, mais pas pour l'autre. Par conséquent, le théorème de Riesz donne lieu à deux identifications différentes du dual de $H_0^1(I)$. Il faut donc préciser laquelle on utilise.

65. On retrouve des idées similaires dans la preuve du lemme 3.17 du Chapitre VIII.

66. Si on préfère le dire ainsi,

$$\langle u,v\rangle_{H_0^1(I)} = \int_a^b u'(t)v'(t)\, dt \tag{2.6}$$

est un produit scalaire sur $H_0^1(]a,b[)$, dont la norme associée est équivalente à la norme de $H^1(]a,b[)$.

67. On pouvait aussi invoquer l'inégalité de Poincaré pour conclure que $u = 0$.

2.3 Problème de Dirichlet en dimension 1

Soit $I =]a, b[$ un intervalle borné, $c \in C^0(\bar{I})$ et $f \in C^0(\bar{I})$. L'objectif de ce paragraphe est de résoudre des problèmes du type

$$\begin{cases} -u''(x) + c(x)u(x) = f(x) & \text{pour tout } x \in \bar{I}, \\ u(a) = u(b) = 0. \end{cases} \tag{2.8}$$

où $u \in C^2(\bar{I})$ est l'inconnue. On appelle un tel problème un *problème aux limites* : on impose des conditions aux extrémités (aux « limites ») de l'intervalle, qui sont ici a et b. La condition porte sur les valeurs de la fonction elle-même, on parle dans ce cas particulier de *conditions de Dirichlet*, et même plus précisément de conditions de Dirichlet homogènes pour spécifier que les valeurs au bord sont nulles.

Il convient de bien distinguer un tel problème d'un problème de Cauchy pour la même équation différentielle. En effet, un problème de Cauchy prend la forme

$$\begin{cases} -u''(x) + c(x)u(x) = f(x) & \text{pour tout } x \in \bar{I}, \\ u(a) = u'(a) = 0, \end{cases} \tag{2.9}$$

par exemple. Ces deux problèmes peuvent sembler assez similaire, mais l'esprit de la résolution de (2.8) que nous allons présenter est radicalement différente de la résolution de (2.9) qui sera exposée au Chapitre X [68]. En l'occurence, le théorème de Cauchy-Lipschitz linéaire assure l'existence et l'unicité de la solution du problème de Cauchy (2.9) dans tous les cas, mais il ne dit directement rien sur le problème aux limites (2.8). D'ailleurs, il y a des cas où ce problème aux limites n'a pas de solution [69]. Nous allons voir dans ce paragraphe comment les outils sur les espaces de Hilbert et les espaces de Sobolev présentés précédemment permettent d'étudier l'existence et l'unicité de la solution à des problèmes du type (2.8). [70]

Commençons par voir une formulation équivalente à (2.8) dans le cas où on suppose $u \in C^2(\bar{I})$:

Proposition 2.23. *Soit $u \in C^2(\bar{I})$ telle que $u(a) = u(b) = 0$. La fonction u est solution de (2.8) si et seulement si*

$$\forall \varphi \in C_c^\infty(I), \int_I (u'\varphi' + cu\varphi)\, dx = \int_I f\varphi\, dx. \tag{2.11}$$

Démonstration. Si u est solution de (2.8), on multiplie l'équation différentielle par $\varphi \in C_c^\infty(I)$, ce qui donne

$$-u''(x)\varphi(x) + c(x)u(x)\varphi(x) = f(x)\varphi(x),$$

en chaque point x de $\bar{I}$, puis on intègre le résultat sur I, d'où

$$-\int_I u''\varphi\, dx + \int_I cu\varphi\, dx = \int_I f\varphi\, dx.$$

On fait une intégration par parties sur la première intégrale et l'on obtient bien la formule (2.11) puisque $\varphi(a) = \varphi(b) = 0$.

---

68. La ressemblance entre (2.8) et (2.9) est en quelque sorte accidentelle, elle vient du fait qu'on traite ici seulement le cas de (2.8) lorsque x est dans un espace de dimension 1. La généralisation naturelle de ce problème dans $\mathbb{R}^N$ avec $N \geq 2$ est :

$$\begin{cases} -\Delta u(x) + c(x)u(x) = f(x) & \text{pour tout } x \in \Omega, \\ u(x) = 0 & \text{pour tout } x \in \partial\Omega \end{cases} \tag{2.10}$$

où Ω est un ouvert borné de $\mathbb{R}^N$, $f : \overline{\Omega} \to \mathbb{R}$ est donnée, et $u : \overline{\Omega} \to \mathbb{R}$ est l'inconnue (rappelons que $\Delta u = \sum_{i=1}^N \partial_{ii}^2 u$). Ce problème est une équation aux dérivées partielles.

69. C'est le cas par exemple pour $I =]0, 1[$, $c(x) = -\pi^2$ et $f(x) = 1$

70. Notons qu'on pourrait se passer de ces stratégies en exploitant le fait qu'on est en dimension 1. Néanmoins la résolution de (2.10) se traite avec la même stratégie que celle qu'on expose ici ; le cas de la dimension $N \geq 2$ amène des difficultés supplémentaires qui expliquent que nous n'aborderons pas ces cas ici (il faudrait commencer par supposer que Ω est un ouvert régulier, ce qui n'est pas une notion au programme, et ensuite la preuve de régularité des solutions (que nous ferons dans le cas de la dimension 1) est nettement plus délicate en dimension supérieure), mais les aspects "analyse fonctionnelle" tels que nous les présentons ici sont exactement les mêmes en dimension supérieure.

Réciproquement, si $u \in C^2(\bar{I})$ satisfait (2.11), alors après intégration par parties en sens inverse, il vient que pour tout $\varphi \in C_c^\infty(I)$,

$$\int_I (-u'' + cu - f)\varphi \, dx = 0,$$

ce qui implique que $-u'' + cu - f = 0$ presque partout sur I d'après le Corollaire 5.61 du Chapitre IV. Du fait que $-u'' + cu - f$ est continue sur $\bar{I}$, on obtient en réalité que $-u'' + cu - f = 0$ partout sur $\bar{I}$. Si de plus par hypothèse $u(a) = u(b) = 0$, on voit que u est bien solution de (2.8). □

On se rend compte que la formulation (2.11) ne recquiert pas autant de régularité sur u que la formulation (2.8). Cela nous conduit naturellement à la définition de solution faible du problème (2.8). On suppose maintenant seulement que $c \in L^\infty(I)$ et $f \in L^2(I)$.

Définition-Proposition 2.24. Soit $u \in H_0^1(I)$, nous dirons que u est *solution faible*[71] de (2.8) si

$$\forall \varphi \in C_c^\infty(I), \int_I (u'\varphi' + cu\varphi) \, dx = \int_I f\varphi \, dx. \tag{2.12}$$

Ceci est en fait équivalent à[72]

$$\forall v \in H_0^1(I), \int_I (u'v' + cuv) \, dx = \int_I fv \, dx. \tag{2.13}$$

Ces formulations sont appelées formulation variationnelle ou formulation faible de (2.8).

Remarque 2.25. La différence avec ce qui précède est que l'on ne suppose plus u de classe C^2, et les conditions aux limites de Dirichlet sont incorporées dans l'espace $H_0^1(I)$.

Démonstration. Comme $C_c^\infty(I) \subset H_0^1(I)$, si u est solution de (2.13), il est a fortiori solution de (2.12).

Réciproquement, soit u vérifiant (2.12). Pour tout $v \in H_0^1(I)$, il existe par définition une suite $\varphi_n \in C_c^\infty(I)$ telle que $\varphi_n \to v$ pour la norme $H^1(I)$, c'est-à-dire que $\varphi_n \to v$ et $\varphi_n' \to v'$ dans $L^2(I)$. Ainsi, par l'inégalité de Cauchy-Schwarz :

$$\left| \int_I u'v' \, dx - \int_I u'\varphi_n' \, dx \right| = \left| \int_I u'(v' - \varphi_n') \, dx \right| \leq \|u'\|_{L^2(I)} \|v' - \varphi_n'\|_{L^2(I)} \xrightarrow[n\to\infty]{} 0.$$

De même, toutes les autres intégrales faisant intervenir φ_n dans (2.12) passent à la limite et on obtient bien (2.13). □

Dans l'énoncé suivant, on regroupe les différents résultats qu'on peut obtenir sur le problème (2.8) :

Théorème 2.26. *Soit $I =]a, b[$ un intervalle borné et $f \in L^2(I)$ et $c \in L^\infty(I)$.*

1. *On suppose $c \geq 0$ p.p. Alors il existe une unique solution faible $u \in H_0^1(I)$ de (2.13).*
2. *On suppose $c \in C^0(\bar{I})$ et $f \in C^0(\bar{I})$, et $u \in H_0^1(I)$ solution de (2.13).*[73] *Alors $u \in C^2(\bar{I})$ et est solution de (2.8).*

Remarque 2.27. En combinant les deux items de cet énoncé, on conclut que si $c \in C^0(\bar{I})$ et $c \geq 0$, alors il existe une unique solution $u \in C^2(\bar{I})$ au problème (2.8).

71. Notons que si on décide d'intégrer par parties deux fois, on arrive à la formulation

$$\forall \varphi \in C_c^\infty(I), \int_I (u(-\varphi'' + c\varphi) \, dx = \int_I f\varphi \, dx.$$

On dit dans ce cas que l'équation est satisfaite "au sens des distributions", mais nous n'utiliserons pas ce point de vue dans la suite, car nous préférerons avoir une "symétrie" entre l'espace où on cherche l'inconnue u et l'espace des fonctions tests.

72. Notons que si $(u, v) \in H^1(I)$, alors $uv \in L^1(I)$ et $u'v' \in L^1(I)$, donc (2.13) (et a fortiori (2.12)) ont bien un sens.

73. Si on ne fait pas ces hypothèses de régularité sur c et f, on peut quand même montrer que $u \in H^2(I)$ et satisfait

$$-u'' + cu = f \quad \text{p.p. sur } I.$$

Démonstration. 1. On considère l'espace $H_0^1(I)$ muni du produit scalaire (2.6), qui est un espace de Hilbert. On définit $a\colon H_0^1(I)\times H_0^1(I)\to\mathbb{R}$ et $L\colon H_0^1(I)\to\mathbb{R}$ par

$$\forall (u,v)\in H_0^1(I),\quad a(u,v)=\int_a^b (u'v'+cuv)\,dx,\qquad \text{et}\qquad L(v)=\int_a^b fv\,dx.$$

D'une part, d'après l'inégalité de Cauchy-Schwarz, on a pour tout $v\in H_0^1(]a,b[)$,

$$|L(v)|\le \|f\|_{L^2(a,b)}\|v\|_{L^2(a,b)}\le C\|f\|_{L^2(a,b)}\|v\|_{H_0^1(]a,b[)},$$

où $C=\frac{1}{\sqrt{2}}$ est la constante de l'inégalité de Poincaré 2.20, ce qui montre que L est une forme linéaire continue sur $H_0^1(]a,b[)$.

D'autre part on a pour tout $u\in H_0^1(]a,b[)$,

$$\|u\|^2_{H_0^1(]a,b[)}\le a(u,u)\le \|u'\|^2_{L^2(a,b)}+\|c\|_{L^\infty(a,b)}\|u\|^2_{L^2(a,b)}\le \left(1+C^2\|c\|_{L^\infty(a,b)}\right)\|u\|^2_{H_0^1(]a,b[)}. \tag{2.14}$$

Comme a est une application bilinéaire symétrique, (2.14) montre que a est un produit scalaire dont la norme est équivalente à la norme (2.6) de $H_0^1(]a,b[)$. Ainsi, on peut munir $H_0^1(]a,b[)$ de ce produit scalaire, et par équivalence avec la norme habituelle, l'espace obtenu est encore un espace de Hilbert, et L est encore une forme linéaire continue pour cette nouvelle norme.

Ainsi, d'après le théorème 1.22 de Riesz, il existe une unique fonction $u\in H_0^1(]a,b[)$ tel que $a(u,v)=L(v)$ pour tout $v\in H_0^1(]a,b[)$.

2. Soit $u\in H_0^1(I)$ solution de (2.13). Alors

$$\forall\varphi\in C_c^1(I),\quad \int_I u'\varphi'\,dx=\int_I (f-cu)\varphi\,dx.$$

Or $f-cu\in C^0(\bar I)$, on en déduit donc que $u'\in H^1(]a,b[)$ et $-u''=f-cu\in C^0(\bar I)$, ce qui d'après la remarque 2.9 montre que $u\in C^2(\bar I)$, et par conséquent $-u''=f-cu$ partout sur $\bar I$.

□

Remarque 2.28. Notons qu'ici, la forme bilinéaire a est symétrique. D'après la remarque 1.24, u solution de (2.13) est aussi l'unique point de minimum de la fonctionnelle $J\colon H_0^1(]a,b[)\to\mathbb{R}$ définie par

$$J(v)=\frac{1}{2}a(v,v)-L(v)=\frac{1}{2}\int_a^b\left((v')^2+cv^2\right)dx-\int_a^b fv\,dx.$$

Remarque 2.29 (Condition de Dirichlet non homogène)**.** Soient α et $\beta\in\mathbb{R}$. On pourrait chercher à résoudre

$$\begin{cases}-u''(x)+c(x)u(x)=f(x) & \text{pour tout } x\in[0,1],\\ u(0)=\alpha,\, u(1)=\beta.\end{cases} \tag{2.15}$$

qu'on appelle problème avec condition de Dirichlet non homogènes (le cas homogène correspondant à $\alpha=\beta=0$). En posant $u_0 : x\mapsto (1-x)\alpha+x\beta$, on constate que $u\in C^2(\bar I)$ est solution de (2.15) si $v=u-u_0$ est solution de (2.8) en remplaçant f par $f-u_0c$.

Exercice 2.30 (Généralisation avec un terme d'ordre 1)**.** Soient $b\in C^1([0,1])$, $c\in C^0([0,1])$ et $f\in C^0([0,1])$ données. On s'intéresse maintenant à trouver une fonction $u\in C^2([0,1])$ telle que

$$\begin{cases}-u''(x)+b(x)u'(x)+c(x)u(x)=f(x) & \text{pour tout } x\in[0,1],\\ u(0)=u(1)=0.\end{cases} \tag{2.16}$$

1. Montrer que si $u\in C^2([0,1])$ satisfait $u(0)=u(1)=0$, alors u est solution de (2.16) si et seulement si

$$\int_0^1 (u'v'+bu'v+cuv)\,dx=\int_0^1 fv\,dx \qquad \text{pour tout } v\in H_0^1(]0,1[). \tag{2.17}$$

2. On suppose $c - \frac{1}{2}b' \geq 0$. Montrer qu'il existe une unique fonction $u \in H_0^1(]0,1[)$ telle que (2.17).[74]
3. Montrer que cette solution est bien de classe C^2 sur $[0,1]$ et est donc solution de (2.16).

Exercice 2.31 (Conditions aux limites de Neumann)**.** Soient $c \in C^0([0,1])$, $f \in C^0([0,1])$. On cherche une fonction $u \in C^2([0,1])$ telle que[75]

$$\begin{cases} -u''(x) + c(x)u(x) = f(x) & \text{pour tout } x \in [0,1], \\ u'(0) = 0, \quad u'(1) = 0. \end{cases} \tag{2.18}$$

1. Soit $u \in C^2([0,1])$. Montrer que u est solution de (2.18) si et seulement si[76]

$$\forall \varphi \in C^\infty([0,1]), \int_0^1 (u'\varphi' + cu\varphi)\, dx = \int_0^1 f\varphi\, dx. \tag{2.19}$$

   puis qu'on peut remplacer $C^\infty([0,1])$ par $H^1(]0,1[)$ dans l'équation précédente.
2. On suppose qu'il existe une constante $c_0 > 0$ telle que $c \geq c_0$. Montrer qu'il existe une unique $u \in H^1(]0,1[)$ tel que (2.19).
3. Montrer que $u \in C^2([0,1])$ et est solution de (2.18).

2.4 Autres espaces de Sobolev

Dans ce paragraphe, I désigne un intervalle ouvert de $\mathbb{R}$.

1. **Espaces de Sobolev d'exposant p et d'ordre 1 :** Pour $p \in [1, \infty]$, on peut définir :

$$\begin{aligned} W^{1,p}(I) &= \{u \in L^p(I),\ u' \in L^p(I)\} \\ &= \left\{u \in L^p(I),\ \exists g \in L^p(I), \forall \varphi \in C_c^\infty(I),\ \int_I u\varphi' dt = -\int_I g\varphi dt\right\} \\ &= \left\{u \in L^p(I),\ \exists (c,g) \in \mathbb{R} \times L^p(I), u(x) = c + \int_{x_0}^x g(t)dt \text{ pour presque tout } x \in I\right\} \end{aligned}$$

   qu'on peut munir de la norme[77]

$$\|u\|_{W^{1,p}} = \|u\|_{L^p} + \|u'\|_{L^p}.$$

   Ces espaces généralisent l'espace $H^1(I)$ qui correspond en fait au cas $p = 2$: $H^1(I) = W^{1,2}(I)$. La plupart des résultats du paragraphe 2.1 s'adaptent à ce cas (avec des preuves très similaires, en utilisant l'inégalité de Hölder à la place de l'inégalité de Cauchy-Schwarz) : en particulier $(W^{1,p}(I), \|\cdot\|_{W^{1,p}(I)})$ est un espace de Banach, ses fonctions ont des représentants continus[78] et

74. On pourra procéder comme pour la preuve du Théorème 2.26, mais en appliquant le théorème de Lax-Milgram plutôt que le théorème de Riesz à l'application bilinéaire (non symétrique)

$$a : \begin{array}{ccc} H_0^1(]0,1]) \times H_0^1(]0,1[) & \longrightarrow & \mathbb{R} \\ (u,v) & \longmapsto & \int_0^1 (u'v' + bu'v + cuv)dx \end{array}.$$

75. Les conditions aux limites portent cette fois sur les dérivées de u, on parle de condition de Neumann homogène et nous allons les prendre en compte de façon très différente de ce que nous avons fait pour les conditions de Dirichlet. En particulier, il n'est pas question d'inclure ces conditions dans l'espace fonctionnel, car une fonction de H^1 a sa dérivée dans L^2, et on ne peut pas considérer des valeurs ponctuelles d'une fonction de L^2. Les conditions seront prises en compte directement dans la formulation variationnelle (2.19), via la possibilité de prendre des fonctions tests qui ne sont pas nulles au bord de l'intervalle sur lequel on travaille.

76. Ici l'espace des fonctions test est $C^\infty([0,1])$ alors que pour le problème de Dirichlet, on prenait $C_c^\infty(]0,1[)$.

77. On peut aussi considérer la norme $\left(\|u\|_p^p + \|u'\|_p^p\right)^{1/p}$ qui est équivalente. D'ailleurs dans le cas $p = 2$, c'est ce qu'on a fait pour $H^1(I) = W^{1,2}(I)$ car c'est cette seconde expression qui est issue d'un produit scalaire, ce qui n'est pas le cas de $\|u\|_2 + \|u'\|_2$.

78. En fait les fonctions de $W^{1,p}$ sont hölderiennes d'exposant $1 - 1/p$ (lipschitziennes si $p = \infty$). De plus, si I est borné et $p \in]1, \infty]$, l'injection $W^{1,p}(I) \subset C^0(\overline{I})$ est compacte. Attention pour ce dernier résultat il faut ici exclure $p = 1$, voir [Bré05].

il est stable par produit. Dans le cas $p \in [1, +\infty[$[79], si I est borné alors $C^\infty(\bar{I})$ est dense dans $W^{1,p}(I)$, et on peut définir $W_0^{1,p}(I)$ comme l'adhérence de $C_c^\infty(I)$ pour la norme de $W^{1,p}(I)$.
Ces espaces permettent de résoudre des équations du type (2.8) mais ayant des termes non linéaires (c'est-à-dire faisant par exemple intervenir des termes du type u^α avec $\alpha \neq 1$).
Dans le cas $p = \infty$, on peut montrer que les fonctions de $W^{1,\infty}(I)$ sont exactement les fonctions bornées et lipschitziennes[80] :

$$W^{1,\infty}(I) = \{u \in L^\infty(I), \ \ \exists k \in \mathbb{R}_+, \forall (x,y) \in I, \ |u(y) - u(x)| \leq k|y - x|\}.$$

Ce résultat n'est pas très simple à montrer, il utilise la représentation du dual de $L^1(I)$ comme étant $L^\infty(I)$[81]. Il est joli de voir que ces arguments donnent le résultat suivant dont l'énoncé se comprend sans faire référence aux espaces de Sobolev ou aux distributions :

Théorème 2.32 (*). *Soit I un intervalle de $\mathbb{R}$ et $f : I \to \mathbb{R}$ lipschitzienne. Alors il existe $g \in L^\infty(I)$ telle que pour tout $x_0 \in I$,*

$$\forall x \in I, \ \ f(x) = f(x_0) + \int_{x_0}^{x} g(t)dt.$$

2. **Espaces de Sobolev d'exposant 2 et d'ordre m :** Pour $m \in \mathbb{N}^*$, on peut définir

$$H^m(I) = \{u \in L^2(I), \forall k \in [\![1, m]\!], u^{(k)} \in L^2(I)\}$$

qu'on peut munir du produit scalaire

$$\forall (u,v) \in H^m(I), \ \ \langle u, v\rangle_{H^m} = \sum_{k=0}^{m} \int_I u^{(k)} v^{(k)} dt,$$

qui est également un espace de Hilbert, et dont les fonctions sont de classe C^{m-1} sur $\bar{I}$.
On peut généraliser cette construction comme au point précédent et définir $W^{m,p}(I)$ pour $p \in [1, \infty]$ et $m \in \mathbb{N}$.

3 Compléments de topologie*

3.1 Théorème de Baire et applications

Le théorème de Baire a été retiré du programme de l'agrégation il y a quelques années, mais sa démonstration est relativement simple, et il reste un sujet classique qui a de nombreuses applications qu'on peut légitimement intégrer dans une leçon.

Avant de se lancer dans des considérations plus élaborée, on conseille l'exercice préliminaire suivant :

Exemple 3.1. Soit (X, d) un espace métrique.

1. Montrer qu'un ensemble $U \subset X$ est dense si et seulement si U rencontre tout ouvert non vide de X.
2. Etant donnés U et V deux ouverts denses de X, montrer que $U \cap V$ est aussi un ouvert dense de X. Est-ce encore vrai si U et V ne sont plus nécessairement ouverts et sont seulement supposés denses ?
3. Montrer que dans l'espace métrique $(\mathbb{Q}, |\cdot|)$, on peut trouver une famille dénombrable d'ouverts denses tels que l'intersection de ces ouverts est vide.

79. Comme d'habitude, le cas $p = +\infty$ est exclu en ce qui concerne les résultats de densité. En fait, si on suppose I borné, l'adhérence de $C^\infty(\bar{I})$ dans $W^{1,\infty}(I)$ est $C^1(\bar{I})$. On laisse le lecteur vérifier qu'il y a dans $W^{1,\infty}(I)$ des fonctions qui ne sont pas C^1.

80. Rappelons qu'on devrait dire "ayant un représentant lipschitzien" mais qu'on fait l'abus de confondre l'élément de $W^{1,\infty}(I)$ qui est une classe de fonction, avec son représentant continu (qui est ici lipschitzien).

81. Ce qui se démontre comme corollaire du Théorème de Riesz 1.22 appliqué à l'espace $L^2(I)$.

Théorème 3.2 (Théorème de Baire/Lemme de Baire)**.** *Soit (X, d) un espace métrique complet. Alors toute intersection dénombrable d'ouverts denses dans X est dense dans X : si $(O_n)_{n\in\mathbb{N}}$ est une famille dénombrable d'ouverts tels que $\forall n \in \mathbb{N}, \overline{O_n} = X$, alors*

$$\overline{\bigcap_{n\in\mathbb{N}} O_n} = X.$$

Remarque 3.3.

1. Attention, $\bigcap_{n\in\mathbb{N}} O_n$ n'est plus nécessairement un ouvert, car on a pris une union infinie. Un ensemble qui peut s'écrire comme intersection dénombrable d'ouverts est appelé un G_δ [82].
2. Il est intéressant de noter que le simple fait que $\bigcap_{n\in\mathbb{N}} O_n$ soit non vide est déjà un résultat non trivial, même si le théorème dit "beaucoup plus" en disant que cet ensemble est en fait dense. Il ne sera pas rare dans les applications qu'on n'utilise seulement que cette intersection est non vide.
3. On utilise parfois le théorème de Baire en "passant au complémentaire" : dans un espace métrique complet, toute union dénombrable de fermés d'interieur vide, est d'intérieur vide : si $(F_n)_{n\in\mathbb{N}}$ est une famille dénombrable de fermés tels que $\forall n \in \mathbb{N},\ \mathring{F_n} = \emptyset$, alors
$$\overset{\circ}{\overbrace{\bigcup_{n\in\mathbb{N}} F_n}} = \emptyset.$$
4. Enfin, on utilise parfois les contraposées suivantes : étant donné (X, d) un espace métrique complet,
   — si (O_n) est une famille dénombrable d'ouverts, et que $\bigcap_{n\in\mathbb{N}} O_n$ n'est pas dense dans X, alors il existe n_0 tel que O_{n_0} n'est pas dense dans X,
   — si (F_n) est une famille dénombrable de fermés, et que $\bigcup_{n\in\mathbb{N}} F_n$ est d'intérieur non vide, alors il existe n_0 tel que F_{n_0} est d'intérieur non vide.

Remarque 3.4. Dans un espace métrique, on dit d'un ensemble :

— qu'il est gras ou de deuxième catégorie s'il contient une réunion dénombrable d'ouverts denses,
— qu'il est maigre ou de première catégorie s'il est contenu dans une intersection dénombrable de fermés d'intérieur vide.

Si l'espace métrique est complet, le théorème de Baire s'exprimer alors :

— une partie est grasse si elle contient un G_δ dense,
— une partie est maigre si elle est contenue dans un F_σ d'intérieur vide.

Démonstration. Soit (O_n) une famille dénombrable d'ouverts denses dans X. On pose $G = \bigcap_n O_n$, et on se donne U un ouvert non vide de X. On vise à montrer que $U \cap G$ est non vide. Comme U est un ouvert non vide, il existe $x_0 \in U$ et $r_0 > 0$ tel que $\overline{B}(x_0, r_0) \subset U$.

On considère maintenant l'ouvert $B(x_0, r_0) \cap O_0$ qui est non vide (car O_0 rencontre tout ouvert non vide par densité), donc il existe $x_1 \in X$ et $r_1 > 0$ tel que $\overline{B}(x_1, r_1) \subset B(x_0, r_0) \cap O_0$. Quitte à réduire r_1 on peut également supposer $r_1 < \frac{r_0}{2}$.

On construit ainsi de façon récurrente, en utilisant à chaque étape $n \in \mathbb{N}$ le fait que O_n est un ouvert dense, deux suites $(x_n) \in X^{\mathbb{N}}$ et $(r_n) \in (\mathbb{R}_+^*)^{\mathbb{N}}$ telles que

$$\forall n \in \mathbb{N},\ \ \overline{B}(x_{n+1}, r_{n+1}) \subset B(x_n, r_n) \cap O_n, \qquad r_{n+1} < \frac{r_n}{2}.$$

Ainsi la suite $(\overline{B}(x_n, r_n))$ est une suite décroissante de fermés dont le diamètre tend vers 0, dans l'espace complet X, donc par la propriété des fermés emboités (Proposition 1.78 au Chapitre III), il existe x_∞ qui est dans tous les $\overline{B}(x_n, r_n)$ pour $n \in \mathbb{N}$, et donc par suite, $x_\infty \in U \cap G$ [83], ce qui conclut la démonstration. □

82. De même, un ensemble qui peut s'écrire comme une réunion dénombrable de fermés est appelée un F_σ.
83. On a $x_\infty \in U$ car $x_\infty \in \overline{B}(x_0, r_0) \subset U$ et $x_\infty \in G$ car pour tout $n \in \mathbb{N}$, $x_\infty \in \overline{B}(x_{n+1}, r_{n+1}) \subset B(x_n, r_n) \cap O_n \subset O_n$.

Applications :

Exercice 3.5. 1. Montrer qu'un espace métrique complet est soit fini soit non dénombrable.[84]

2. Montrer que $\mathbb{R}^N$ avec $N \geq 2$ n'est pas réunion dénombrable de droites.[85]
3. Montrer que $\mathbb{R} \setminus \mathbb{Q}$ ne peut pas s'écrire comme la réunion dénombrable de fermés. En déduire que $\mathbb{Q}$ ne peut pas s'écrire comme intersection dénombrable d'ouverts.

Voici une application très classique du théorème de Baire sur les espaces de Banach :

Théorème 3.6. *Un espace vectoriel normé admettant une base (algébrique) infinie dénombrable ne peut pas être complet.*

Démonstration. Supposons que E est un espace vectoriel normé et que $(e_n)_{n\in\mathbb{N}}$ est une base algébrique de E[86]. On pose
$$\forall n \in \mathbb{N}, \quad F_n = \mathrm{Vect}(e_0, \cdots, e_n).$$
Pour tout $n \in \mathbb{N}$, F_n est un sev strict de E donc F_n est d'intérieur vide[87], et comme il est de dimension finie, il est fermé dans E. Comme $E = \bigcup_n F_n$ est d'intérieur non vide, le théorème de Baire ne peut pas s'appliquer, ce qui signifie que E n'est pas complet. □

Remarque 3.7. Dit autrement, un espace de Banach de dimension infinie n'a jamais de base algébrique dénombrable. Cela explique pourquoi la notion de base algébrique n'est pas commode, voir le paragraphe 1.3 sur les bases hilbertiennes.

Exemple 3.8. L'espace $\mathbb{R}[X]$ n'est donc complet pour aucune norme.

Exercice 3.9. Soit E un espace de Banach, et $T \in \mathrm{L}(E)$ tel que
$$\forall x \in E, \quad \exists n \in \mathbb{N}, \quad T^n(x) = 0.$$
Montrer que T est nilpotent[88].

Exercice 3.10 (Exo 3.a page 400 de [Gou08b])**.** Soit $f : \mathbb{R}_+ \to \mathbb{R}$ une fonction continue telle que
$$\forall x \in]0, +\infty[, \quad \lim_{n\to+\infty} f(nx) = 0.$$

1. Soit $\varepsilon > 0$. Pour $n \in \mathbb{N}$, on pose
$$F_n = \{x \in \mathbb{R}_+, \ \forall p \in [\![n, +\infty[\![, |f(px)| \leq \varepsilon\}.$$
Montrer qu'il existe $n_0 \in \mathbb{N}$ tel que F_{n_0} est d'intérieur non vide.
2. Conclure que $\lim_{x\to+\infty} f(x) = 0$.

Exercice 3.11. Soit I un intervalle de longueur finie dans $\mathbb{R}$, et $E \subset L^1(I)$ un sev fermé de $L^1(I)$ tel que
$$\forall f \in E, \quad \exists p > 1, \quad f \in L^p(I).$$

1. pour $p > 1$ et $N \in \mathbb{N}$, on pose
$$E_{p,N} = \{f \in E \cap L^p(I), \quad \|f\|_{L^p(I)} \leq N\}.$$
Montrer que $E_{p,N}$ est un fermé de E (muni de la norme L^1).
2. Montrer qu'il existe $p_0 > 1$ tel que $E \subset L^{p_0}(I)$.

84. On retrouve que $\mathbb{R}$ n'est pas dénombrable puisqu'il est complet (pour la norme "valeur absolue").
85. Ca marche avec des cercles aussi...
86. Rappelons qu'un ensemble infini dénombrable est en bijection avec $\mathbb{N}$.
87. Dans un evn, tout sev strict est d'intérieur vide. Le faire en exercice !
88. Il s'agit d'un exercice classique. Comme pour le Théorème de Banach-Steinhaus 3.20 on cherche une uniformité : dans l'hypothèse l'exposant n dépend de x, et ici on cherche à montrer qu'il existe un exposant qui convient pour tous les x.

Un peu dans la veine de l'Exercice 3.9, on a les caractérisations suivantes des polynômes[89] :

Théorème 3.12. *1. Soit $f : \mathbb{C} \to \mathbb{C}$ holomorphe et telle que*

$$\forall z \in \mathbb{C}, \;\; \exists n \in \mathbb{N}, \;\; f^{(n)}(z) = 0.$$

Alors f est une fonction polynomiale.

2. Soit $f : \mathbb{R} \to \mathbb{R}$ de classe C^∞ et telle que

$$\forall x \in \mathbb{R}, \;\; \exists n \in \mathbb{N}, \;\; f^{(n)}(x) = 0. \tag{3.1}$$

Alors f est une fonction polynomiale.

Démonstration. 1. Soit $f : \mathbb{C} \to \mathbb{C}$ holomorphe. On pose pour $n \in \mathbb{N}$, $F_n = \{z \in \mathbb{C}, \;\; f^{(n)}(z) = 0\}$ qui est un fermé de $\mathbb{C}$. Par hypothèse, $\mathbb{C} = \bigcup_n F_n$, et comme $\mathbb{C}$ est complet on peut appliquer le Théorème de Baire, qui donne l'existence de n_0 tel que F_{n_0} est d'intérieur non vide. Ainsi $f^{(n_0)}$ s'annule sur un ouvert non vide de $\mathbb{C}$, donc par le théorème des zéros isolés (et connexité de $\mathbb{C}$) on en déduit $f^{(n_0)} \equiv 0$ et donc f est une fonction polynomiale.

2. Le cas réel est plus délicat, on l'appelle Théorème de Sunyer i Balaguer, on renvoie à [Gou08b, page 402].

□

Remarque 3.13. En fait on peut généraliser le résultat précédent (et adapter la preuve) au cas où on remplace (3.1) par

$$\forall x \in \mathbb{R}, \;\; \exists n \in \mathbb{N}, \;\; f^{(n)}(x) \in D,$$

où D est un sous-ensemble dénombrable de $\mathbb{R}$.

Applications aux ensembles de continuité des fonctions : Commençons par une petite digression qui n'est pas directement liée au Théorème de Baire, sauf au point 1 de l'exemple 3.16.

Proposition 3.14. *Soit $f : X \to Y$ une application entre deux espaces métriques. Alors l'ensemble des points de discontinuité de f est une réunion dénombrable de fermés de X (c'est-à-dire un F_σ).*

Remarque 3.15. Par passage au complémentaire, l'ensemble des points de continuité de f est une intersection dénombrable d'ouverts de X, c'est-à-dire un G_δ.

Démonstration. [90]Pour $f : X \to Y$ et $x \in X$ on pose[91]

$$\omega_f(x) = \inf_{\delta > 0} \left\{ \sup_{(x,y) \in B(x_0,\delta)} \{d_Y(f(x), f(y))\} \right\}.$$

— Remarquons déjà que f est continue en x_0 si et seulement si $\omega_f(x_0) = 0$: en effet si f continue en x_0 alors pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que $\forall x \in B(x_0, \delta)$, $d_Y(f(x), f(x_0)) \leq \varepsilon$, et donc

$$\forall (x, y) \in B(x_0, \delta), \;\; d_Y(f(x), f(y)) \leq d_Y(f(x), f(x_0)) + d_Y(f(x_0), f(y)) \leq 2\varepsilon,$$

ce qui donne $\omega_f(x_0) \leq 2\varepsilon$, et comme ceci est vrai pour tout $\varepsilon > 0$, on a bien $\omega_f(x_0) = 0$. Réciproquement si $\omega_f(x_0) = 0$, alors pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que

$$\sup_{(x,y) \in B(x_0,\delta)} d_Y(f(x), f(y)) \leq \varepsilon$$

et en prenant en particulier $y = x_0$ on obtient que f est continue en x_0.

89. A ma connaissance ces caractérisations sont plutôt folkloriques ; si vous en connaissez des applications intéressantes, faites-le moi savoir !

90. Cette preuve s'avère relativement astucieuse si on n'a aucune indication : en effet en utilisant la définition classique de la continuité ponctuelle, on a du mal à obtenir le résultat (essayez !). Pour cette raison, on va d'abord montrer que f est continue en x_0 ssi

$$\forall \varepsilon > 0, \exists \delta > 0, \forall (x, y) \in B(x_0, \delta), \;\; d_Y(f(x), f(y)) < \varepsilon,$$

variante de la définition de la continuité qui permettra de prouver le résultat.

91. Il s'agit de l'oscillation de f en x. On a les définitions équivalentes :

$$\omega_f(x) = \inf_{\delta > 0} \{\mathrm{diam}(f(B_X(x_0, \delta))\} = \lim_{\delta \to 0^+} \mathrm{diam}(f(B_X(x_0, \delta)) = \inf\{\mathrm{diam}(f(U)), U \text{ voisinage de } x_0\}.$$

— On montre maintenant que $\{x_0 \in X,\ \omega_f(x_0) < \varepsilon\}$ est un ouvert [92]. Soit x_0 tel que $\omega_f(x_0) < \varepsilon$. Il existe $\delta > 0$ tel que $\sup_{(x,y)\in B(x_0,\delta)} d_Y(f(x), f(y)) < \varepsilon$. On prend $x_1 \in B(x_0, \delta)$ et on pose $\delta' = \delta - d_X(x_0, x_1) > 0$ de sorte que $B(x_1, \delta') \subset B(x_0, \delta)$. Alors

$$\sup_{(x,y)\in B(x_1,\delta')} d_Y(f(x), f(y)) \leq \sup_{(x,y)\in B(x_0,\delta)} d_Y(f(x), f(y)) < \varepsilon$$

donc $\omega_f(x_1) < \varepsilon$ et donc $B(x_0, \delta) \subset \{x \in X,\ \omega_f(x) < \varepsilon\}$, cet ensemble est donc bien un ouvert.

— Ainsi l'ensemble des points de continuité de f peut s'écrire $\bigcap_{n\in\mathbb{N}^*} \left\{\omega_f(x) < \frac{1}{n}\right\}$ qui est donc une intersection dénombrable d'ouverts.

□

Exemple 3.16. 1. Il n'existe pas de fonction $f : \mathbb{R} \to \mathbb{R}$ qui soit continue sur $\mathbb{Q}$ et discontinue sur $\mathbb{R} \setminus \mathbb{Q}$. En effet $\mathbb{Q}$ ne peut pas s'écrire comme une intersection dénombrable d'ouverts, voir l'exercice 3.5.

2. Remarquons qu'il existe bien une fonction $f : \mathbb{R} \to \mathbb{R}$ qui est continue sur $\mathbb{R} \setminus \mathbb{Q}$ et discontinue sur $\mathbb{Q}$: on peut prendre par exemple la fonction de Thomae :

$$\begin{array}{rcl} T : \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \begin{cases} 0 \text{ si } x \notin \mathbb{Q} \\ \frac{1}{q} \text{ si } x = \frac{p}{q} \text{ avec } (p,q) \in \mathbb{Z} \times \mathbb{N}^* \text{ et } pgcd(p,q) = 1 \end{cases} \end{array}$$

3. Plus généralement on peut montrer que si A est un F_σ de $\mathbb{R}$ alors on peut construire $f : \mathbb{R} \to \mathbb{R}$ telle que l'ensemble des points de discontinuité de f est exactement A, voir [Laz].

Comme on l'a vu au Chapitre I, si une suite (f_n) de fonctions continues converge uniformément vers une fonction f, alors cette dernière est continue. Ce résultat ne persiste plus lorsqu'on a seulement une convergence simple, mais le théorème de Baire permet néanmoins de montrer qu'une limite simple de fonctions continues ne peut pas avoir trop de discontinuités.

Théorème 3.17. *Soit X et Y deux espaces métriques, et $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions continues de X vers Y. On suppose que X est complet et que (f_n) converge simplement vers $f : X \to Y$. Alors f est continue sur un sous-ensemble dense de X.*

Démonstration. — On fixe $\varepsilon > 0$, et pour tout $n \in \mathbb{N}$ on pose

$$F_{n,\varepsilon} = \{x \in X,\ \forall p \geq n, \forall q \geq n, d(f_p(x), f_q(x)) \leq \varepsilon\},$$

qui est un fermé comme intersection de fermés (on utilise la continuité des $(f_p)_{p\geq n}$.
Comme (f_n) converge simplement, pour tout $x \in X$ la suite $(f_n(x))_{n\in\mathbb{N}}$ est de Cauchy, et donc $\bigcup_n F_{n,\varepsilon} = E$.

— On va en déduire que $\Omega_\varepsilon := \bigcup_n \left(\overset{\circ}{F}_{n,\varepsilon}\right)$ est dense dans X. Posons pour cela $G_\varepsilon = X \setminus \Omega_\varepsilon$. Alors

$$G_\varepsilon = G_\varepsilon \cap \bigcup_n F_{n,\varepsilon} = \bigcup_n (G_\varepsilon \cap F_{n,\varepsilon})$$

or pour tout $n \in \mathbb{N}$, $F_{n,\varepsilon} \cap G_\varepsilon$ est un fermé et

$$\overset{\circ}{\overbrace{G_\varepsilon \cap F_{n,\varepsilon}}} \subset G_\varepsilon \cap \overset{\circ}{F}_{n,\varepsilon} = \emptyset$$

donc par le Théorème de Baire (voir le point 3 de la Remarque 3.21) dans l'espace X supposé complet, G_ε est d'intérieur vide, i.e. Ω_ε est dense.

92. *Ceci montre que $x_0 \in X \mapsto \omega_f(x_0) \in \mathbb{R}$ est semi-continue supérieurement. Par contre cette application n'est pas continue en général, comme on peut le voir en considérant $f = \mathbb{1}_{\{0\}}$ pour laquelle $\omega_f = \mathbb{1}_{\{0\}}$.

— En réappliquant le théorème de Baire, on en déduit que $\bigcap_{N\in\mathbb{N}^*}\Omega_{1/N}$ est un ensemble dense car il est intersection dénombrable d'ouverts denses. On se donne enfin $x\in\bigcap_{N\in\mathbb{N}^*}\Omega_{1/N}$ et on va montrer que f est continue en x. Pour tout $N\in\mathbb{N}^*$, $x\in\Omega_{1/N}$ donc il existe $n_0\in\mathbb{N}$ et $\delta>0$ tel que $B(x,\delta)\subset F_{n_0,1/N}$ et donc

$$\forall y\in B(x,\delta),\forall(p,q)\geq n_0,\ \ d(f_p(y),f_q(y))\leq 1/N$$

ce qui par passage à la limite $q\to\infty$ et en prenant $p=n$ donne

$$\forall y\in B(x,\delta),\ d(f_{n_0}(y),f(y))\leq 1/N.$$

Or f_n est continue en x donc il existe $\delta'>0$ tel que

$$\forall y\in B(x,\delta'),\ d(f_{n_0}(y),f_{n_0}(y))\leq 1/N$$

d'où finalement

$$\forall y\in B(x,\inf\{\delta,\delta'\}),\ d(f(x),f(y))\leq d(f(x),f_{n_0}(x))+d(f_{n_0}(x),f_{n_0}(y))+d(f_{n_0}(y),f(y))\leq 3/N,$$

ce qui conclut la démonstration.

□

Corollaire 3.18. *Si $f:\mathbb{R}\to\mathbb{R}$ est une fonction dérivable, alors f' est continue sur un sous-ensemble dense de $\mathbb{R}$.*

Démonstration. La fonction f' est limite simple de la suite de fonctions

$$\left(x\in\mathbb{R}\mapsto\frac{f(x+\frac{1}{n})-f(x)}{\frac{1}{n}}\right)_{n\in\mathbb{N}^*}$$

qui sont continues sur $\mathbb{R}$ qui est complet, le Théorème 3.17 s'applique donc. □

Exercice 3.19 (Continuité séparée)**.** Soit (X,Y,Z) trois espaces métriques, avec X et Y complet. On se donne $f:X\times Y\to\mathbb{Z}$ telle que

$$\forall x\in X, y\mapsto f(x,y) \text{ est continue, et } \forall y\in X, x\mapsto f(x,y) \text{ est continue}$$

c'est-à-dire séparément continue en chaque variable.

1. Pour $\varepsilon>0$ on pose $O_\varepsilon=\{(x,y)\in X\times Y,\omega_f(x,y)<\varepsilon\}$. On veut montrer que pour tout $r>0$ et tout $(x_0,y_0)\in X\times Y$, l'ensemble

$$D_\varepsilon\cap\big(B(x_0,r)\times B(y_0,r)\big)$$

est non vide :

   (a) Montrer qu'il existe $\alpha\in]0,r[$ tel que

$$\forall x\in B(x_0,r),\ \ d(f(x,y_0),f(x_0,y_0))<\varepsilon/4.$$

   (b) On pose pour $\beta>0$,

$$F_\beta=\Big\{x\in B(x_0,\alpha),\forall y\in B(y_0,\beta),\ d(f(x,y),f(x,y_0))<\varepsilon/4\Big\}.$$

   Montrer qu'il existe $\beta_0\in]0,r[$ tel que F_{β_0} est d'intérieur non vide.

   (c) Conclure.

2. En déduire que f est continue sur une partie dense de $X\times Y$.

3.2 Théorème de Banach-Steinhaus et applications

Voici une autre application importante du Théorème de Baire dans l'étude des applications linéaires continues, qui aura elle-même ses propres applications.

Théorème 3.20 (Théorème de Banach-Steinhaus, version 1)**.** *Soient E un espace de Banach, F un espace vectoriel normé, et $(T_i)_{i\in I}$ une famille*[93] *d'application linéaires continues de E dans F. On suppose cette famille ponctuellement bornée, c'est-à-dire*

$$\forall x \in E, \quad (\|T_i(x)\|_F)_{i\in I} \text{ est bornée.}$$

Alors la famille $(T_i)_{i\in I}$ est bornée dans $\mathcal{L}(E,F)$, c'est-à-dire

$$(\|T_i\|_{\mathcal{L}(E,F)})_{i\in I} \text{ est bornée.}$$

Remarque 3.21. Il s'agit d'un résultat d'uniformité, et d'ailleurs en anglais ce résultat est appelé "principe de la borne uniforme"[94]. En effet l'hypothèse peut se réécrire :

$$\forall x \in E, \exists M_x \in \mathbb{R}_+, \quad \forall i \in I, \quad \|T_i(x)\|_F \leq M_x,$$

et la conclusion affirme

$$\exists M \in \mathbb{R}_+, \quad \forall i \in I, \; \forall x \in E, \quad \|T_i(x)\|_F \leq M\|x\|_E.$$

Notons que cette uniformité n'a pas été obtenue par un argument de compacité, comme cela avait été le cas au paragraphe 1.5.4 du Chapitre III.

On peut en fait améliorer l'énoncé de la façon suivante (avec essentiellement la même preuve) :

Théorème 3.22 (Théorème de Banach-Steinhaus, version 2)**.** *Soient E et F deux espaces de Banach, et $(T_i)_{i\in I}$ une famille de $\mathcal{L}(E,F)$. Alors*[95]

- *soit $(T_i)_{i\in I}$ est bornée dans $\mathcal{L}(E,F)$,*
- *soit l'ensemble $\Big\{x \in E, (T_i(x))_{i\in I} \text{ est non bornée dans } F\Big\}$ est dense dans E.*

Dans la preuve suivante, nous montrons les Théorèmes 3.20 et 3.22.

Démonstration. Commençons par prouver le Théorème 3.20 : on pose pour tout $n \in \mathbb{N}$,

$$F_n = \{x \in E, \quad \forall i \in I, \|T_i(x)\|_F \leq n\},$$

qui est un fermé (c'est l'intersection sur $i \in I$ des fermés $\{x \in E, \|T_i(x)\| \leq n\}$) et par hypothèse de borne ponctuelle, $\bigcup_n F_n = E$ qui est donc d'intérieur non vide, donc par le Théorème de Baire dans l'espace E supposé complet (voir la remarque 3.21), il existe $n_0 \in \mathbb{N}$ tel que F_{n_0} est d'intérieur non vide, c'est-à-dire qu'il existe $x_0 \in E$ et $r_0 > 0$ tel que $B(x_0, r_0) \subset F_{n_0}$, ce qui signifie

$$\forall x \in B(x_0, r_0), \quad \forall i \in I, \quad \|T_i(x)\|_F \leq n_0$$

d'où

$$\forall x \in B(0,1), \quad \forall i \in I, \quad \|T_i(x)\|_F = \left\|\frac{T_i(x_0 + rx) - T_i(x_0)}{r}\right\|_F \leq \frac{2n_0}{r},$$

d'où le résultat.

Pour montrer le Théorème 3.22, on suppose que l'ensemble $\Big\{x \in E, (T_i(x))_{i\in I} \text{ est non bornée dans } F\Big\}$ n'est pas dense dans E, ce qui revient à dire que $\Big\{x \in E, (T_i(x))_{i\in I} \text{ est bornée dans } F\Big\}$ est d'intérieur non vide. La suite de la preuve est exactement la même en constatant que ce dernier ensemble est égal à $\bigcup_n F_n$ avec les notations précédentes. □

93. On ne suppose rien sur le cardinal de cette famille, elle peut donc être non dénombrable.
94. Principle of uniform boundedness.
95. Ces deux alternatives sont de plus incompatibles, donc une et une seule d'entre elles sera satisfaite.

Applications : En première application du Théorème 3.20 de Banach-Steinhaus, on voit que $\mathcal{L}(E,F)$ est fermé pour la convergence simple :

Proposition 3.23. *Soit E un espace de Banach, et F un evn. Si $(T_n)_{n\in\mathbb{N}}$ est une suite d'applications linéaires continues de E dans F qui converge simplement vers $T : E \to F$, alors T est continue*[96].

Démonstration. Par hypothèse, pour tout $x \in E$, la suite $(T_n(x))$ est convergente dans F, donc elle est bornée. Par complétude de E on peut appliquer le Théorème 3.20 qui donne l'existence de $M \in \mathbb{R}_+$ tel que

$$\forall n \in \mathbb{N}, \forall x \in E, \quad \|T_n(x)\| \leq M\|x\|$$

et on peut passer à la limite $n \to \infty$ pour en déduire que T est continue. □

Remarque 3.24. Attention, la proposition 3.23 ne dit pas que la convergence de $(T_n)_{n\in\mathbb{N}}$ vers T est valable dans l'espace $\mathcal{L}(E,F)$. On peut par exemple considérer $E = c_0(\mathbb{N})$ (espace des suites rélles de limite nulle, qui est un espace de Banach pour la norme $\|\cdot\|_\infty$) et $\forall n \in \mathbb{N}, \forall x \in E, T_n(x) = x_n$. Alors (T_n) est bien une suite de $\mathcal{L}(E,\mathbb{R})$, et elle converge simplement vers 0, mais comme $\forall n \in \mathbb{N}, \|T_n\| = 1$, la convergence de T_n vers 0 n'est pas valable pour la norme de $\mathcal{L}(E,\mathbb{R})$.

Remarque 3.25. Si E n'est pas supposé complet dans la proposition 3.23, le résultat est faux en général. On considère $E = c_{00}(\mathbb{N})$ l'espace des suites à support fini munit de $\|\cdot\|_\infty$ et

$$\forall n \in \mathbb{N},\ \forall x \in E,\ \ T_n(x) := \sum_{k=0}^{n} x_k.$$

Pour tout $n \in \mathbb{N}$, $T_n \in \mathcal{L}(E,\mathbb{R})$ car $\|T_n(x)\| \leq (n+1)\|x\|_\infty$ pour tout $x \in E$, et (T_n) converge simplement vers $T : x \in E \mapsto \sum_{k\in\mathbb{N}} x_k$. Mais T n'est pas continue, car avec $x^{(N)} = \mathbb{1}_{[\![0,N]\!]}$ on obtient

$$\frac{T(x^{(N)})}{\|x^{(N)}\|_\infty} = N+1 \xrightarrow[N\to\infty]{} +\infty.$$

Proposition 3.26. *Soit E_1, E_2 et F trois espaces vectoriels normés avec E_1 complet, et $b : E_1 \times E_2 \to F$ une application bilinéaire. On suppose :*

$$\forall x_1 \in E_1,\ \ y \in E_2 \mapsto b(x_1,y) \text{ est continue}, \quad \text{et} \quad \forall x_2 \in E_2,\ \ y \in E_2 \mapsto b(y,x_2) \text{ est continue.}$$

Alors b est continue.

Remarque 3.27. L'énoncé affirme qu'une application bilinéaire qui est continue séparément en chaque variable, est continue en son couple de variable (lorsque l'espace E_1 est complet). Ce type de résultat ne découle pas de la définition bien sûr, voir l'exemple 3.28.

Démonstration. Pour $y \in E_2$, on pose $\varphi_y = b(\cdot,y)$. Par continuité de b en la première variable, $\varphi_y \in \mathcal{L}(E_1,F)$ pour tout $y \in E_2$. On va appliquer le théorème 3.20 de Banach-Steinhaus à la famille $\{\varphi_y, y \in E_2 \text{ tel que } \|y\| = 1\}$. En effet, par continuité de b en la seconde vairable, pour tout $x \in E_1$ il existe $M_x \in \mathbb{R}_+$ tel que $\forall y \in E_2$, $\|b(x,y)\|_F \leq M_x\|y\|$, ce qui implique

$$\forall x \in E_1, \quad (\|\varphi_y(x)\|)_{\{y,\|y\|=1\}} \text{ est bornée.}$$

Comme l'espace E_1 est supposé complet, on peut appliquer le Théorème 3.20, et en déduire qu'il existe $M \in \mathbb{R}_+$ tel que

$$\forall x \in E_1, \forall y \in E_2 \text{ tel que } \|y\| = 1, \quad \|b(x,y)\| \leq M\|x\|.$$

Par homogénéité, cela implique

$$\forall x \in E_1, \forall y \in E_2, \quad \|b(x,y)\| \leq M\|x\|\|y\|,$$

ce qui conclut la démonstration. □

96. On a aussi l'inégalité $\|T\| \leq \limsup_{n\to\infty} \|T_n\|$. Pour le voir, on écrit

$$\forall n \in \mathbb{N}, \forall x \in E,\ \|T_n(x)\| \leq \|T_n\|\|x\| \leq (\sup_{p\geq n} \|T_p\|)\|x\|$$

et on fait tendre n vers $+\infty$.

Exemple 3.28. Bien sûr, le résultat reste valable si E_2 est complet au lieu de E_1. Par contre, on va donner ici un exemple pour montrer que le résultat est faux quand ni E_1 ni E_2 n'est complet. On pose $E = c_{00}(\mathbb{N}) = \{u : \mathbb{N} \to \mathbb{R}$ à support fini$\}$ qu'on munit de la norme $\|\cdot\|_\infty$, et

$$b : (u,v) \in E \times E \longmapsto \sum_{n\in\mathbb{N}} u_n v_n.$$

Alors si u est fixé,

$$\forall v \in E, \quad |b(u,v)| \leq \left(\sum_{n\in\mathbb{N}} |v_n|\right) \|u\|_\infty$$

qui montre que $b(u,\cdot)$ est continue. De même on a $b(\cdot,v)$ continue pour tout $v \in E$. Néanmoins, b n'est pas continue sur $E \times E$, comme on le voit en considérant $u^{(N)} = \mathbb{1}_{[\![0,N]\!]}$ qui donne

$$\frac{b(u^{(N)},u^{(N)})}{\|u^{(N)}\|_\infty^2} = N+1 \xrightarrow[N\to+\infty]{} +\infty.$$

3.3 Théorème de l'application ouverte et applications

Théorème 3.29 (Theorème de l'application ouverte)**.** *Soient E et F deux espaces de Banach, et T une application linéaire continue et surjective de E vers F. Alors T est ouverte, c'est-à-dire que l'image par T de tout ouvert de E est un ouvert de F.*

Démonstration. **Étape 1 :** montrons qu'il suffit de prouver qu'il existe $c > 0$ tel que

$$T(B_E(0,1)) \supset B_F(0,c). \tag{3.2}$$

En effet si tel est le cas, on se donne U un ouvert de E, et $y_0 = T(x_0)$ avec $x_0 \in U$. Comme U est ouvert, il existe $r > 0$ tel que $B_E(x_0,r) \subset U$. Montrons alors que $B_F(y_0,rc) \subset T(U)$: si $y \in B_F(y_0,rc)$, alors $\frac{y-y_0}{r} \in B_F(0,c)$ donc par (3.2) il existe $z \in B_E(0,1)$ tel que $\frac{y-y_0}{r} = T(z)$, d'où $y = T(x_0+rz) \in T(U)$ puisque $x_0 + rz \in B_E(x_0,r) \subset U$.

Étape 2 : On montre que sous les hypothèses du Théorème 3.29, il existe $c > 0$ tel que

$$\overline{T(B_E(0,1))} \supset B_F(0,2c). \tag{3.3}$$

Pour cela, on pose pour tout $n \in \mathbb{N}$, $F_n = \overline{T(B_E(0,n))}$, qui est une famille dénombrable de fermés de F. Comme T est surjective, on a $E = \bigcup_n F_n$, donc par le théorème de Baire 3.2 (qui s'applique car F est supposé complet), il existe $n_0 \in \mathbb{N}$ tel que F_{n_0} est d'intérieur non vide, donc il existe $r > 0$ et $y_0 \in F$ tel que $B_F(y_0,r) \subset F_{n_0}$.

Ceci implique en fait $B_F(0,r) \subset F_{2n_0}$: en effet si $z \in B_F(0,r)$, alors $(y_0, y_0+z) \in F_{n_0}^2$ donc il existe deux suites $(z_p = T(x_p), \widetilde{z}_p = T(\widetilde{x}_p))$ de $T(B_E(0,n_0))$ qui convergent respectivement vers y_0 et y_0+z, ce qui donne que z est limite de $T(\widetilde{x}_p - x_p)$ et donc $z \in F_{2n_0}$.

Par linéarité de T, on a également $F_{2n_0} = 2n_0 F_1$, donc finalement en posant $c = r/(2n_0)$ on obtient bien $B_F(0,2c) \subset F_1$.

Étape 3 : Montrons maintenant que (3.3) implique (3.2) pour conclure la preuve du Théorème. Soit $y \in B_F(0,c)$. Comme $2y \in B_F(0,2c) \subset \overline{T(B_E(0,1))}$ on a

$$\forall \varepsilon > 0, \exists x \in B_E(0,1), \quad \|2y - T(x)\| < \varepsilon.$$

- On applique ceci avec $\varepsilon = c$ et on obtient l'existence de $x_1 \in B_E(0,1)$ tel que $\|y - T(x_1/2)\| < c/2$.
- On reproduit l'argument en considérant $4(y - T(x_1/2)) \in B(0,2c)$, ce qui donne l'existence de $x_2 \in B(0,1)$ tel que
$$\|y - T(x_1/2) - T(x_2/4)\| < c/4.$$
- En itérant le procédé, on construit une suite $(x_n)_{n\geq 1}$ de $B_E(0,1)$ telle que
$$\left\| y - \sum_{k=1}^n T(x_k/2^k) \right\| < c/2^n.$$

Pour conclure, on constate que $\sum_{k\in\mathbb{N}^*}\|T(x_k/2^k)\| \le \|T\|\sum_k 1/2^k < +\infty$, donc par complétude de E, la proposition 2.18 du Chapitre III montre que la série $\sum_k x_k/2^k$ converge vers $S \in E$. Par continuité de T, on a bien $y = T(S)$, et de plus $\|S\| \le \sum_{k\in\mathbb{N}}\|x_k\|/2^k < 1$, d'où finalement $y \in T(B_E(0,1))$. □

Corollaire 3.30 (Théorème d'isomorphisme de Banach)**.** *Soient E et F deux espaces de Banach, et T une application linéaire continue et bijective de E vers F. Alors T^{-1} est continue, c'est-à-dire qu'il existe $c > 0$ tel que*

$$\forall x \in E, \quad \|T(x)\|_F \ge c\|x\|_E.$$

Démonstration. On peut appliquer le Théorème 3.29, qui nous dit que l'image $T(B(0,1))$ est un ouvert, et comme elle contient 0, il existe $c > 0$ tel que $B(0,c) \subset T(B(0,1))$, ce qui revient par bijectivité de T à dire que $T^{-1}(B(0,c)) \subset B(0,1)$, soit

$$\forall y \in B(0,c), \ \ \|T^{-1}(y)\|_E \le 1, \qquad \text{d'où } \forall y \in F\backslash\{0\}, \ \ \|T^{-1}(y)\|_E = \left\|T^{-1}\left(\frac{cy}{2\|y\|_F}\right)\right\|_E \frac{2\|y\|_F}{c} \le \frac{2}{c}\|y\|_F.$$

□

Exemple 3.31. Soit E est un espace vectoriel muni de deux normes N_1 et N_2, et tel que l'espace E est complet pour chacune de ces normes. S'il existe $C > 0$ tel que $\forall x \in E, N_1(x) \le CN_2(x)$, alors par le corollaire 3.30 les deux normes N_1 et N_2 sont équivalentes. En effet on applique le résultat à l'application $\mathrm{Id} : (E, N_2) \to (E, N_1)$ qui est linéaire, bijective et continue, et donc son inverse est continu, ce qui donne l'existence de $c > 0$ telle que $\forall x \in E, N_1(x) \ge cN_2(x)$.

Exercice 3.32 (Non surjectivité des coefficients de Fourier sur $L^1(\mathbb{T})$ [97])**.** On considère

$$\begin{array}{cccc} \mathcal{F} & L^1(\mathbb{T}) & \longrightarrow & c_0(\mathbb{Z}) \\ & f & \longmapsto & \widehat{f} = \left(\widehat{f}(n)\right)_{n\in\mathbb{Z}} \end{array}$$

l'opérateur de coefficients de Fourier, c'est-à-dire $\forall f \in L^1(\mathbb{T}), \forall n \in \mathbb{Z}, \ \widehat{f}(n) = \int_{\mathbb{T}} f(t)e^{-int}\frac{dt}{2\pi}$.

1. Montrer que $\mathcal{F}$ est linéaire continue et de norme 1 [98].
2. On rappelle que le noyau de Dirichlet s'écrit :

$$\forall N \in \mathbb{N}, \forall x \in \mathbb{R}, \ \ D_N(x) = \sum_{n=-N}^{N} e^{inx}, \quad \forall x \in \mathbb{R} \setminus 2\pi\mathbb{Z}, \ \ D_N(x) = \frac{\sin((N+1/2)x)}{\sin(x/2)}.$$

   Calculer $\|D_N\|_{L^1(\mathbb{T})}$ et $\|\mathcal{F}(D_N)\|_\infty$ pour $N \in \mathbb{N}$.
3. En déduire que $\mathcal{F}$ n'est pas surjective.

Exercice 3.33 (Non surjectivité de la transformation de Fourier sur $L^1(\mathbb{R})$ [99])**.** On considère la transformation de Fourier sur $\mathcal{F} : L^1(\mathbb{R}) \to C^0_{\to 0}(\mathbb{R})$, et on pose

$$\forall n \in \mathbb{N}, f_n = \widehat{\mathbb{1}_{[-1,1]}}\,\widehat{\mathbb{1}_{[-n,n]}}$$

1. Montrer que pour tout $n \in \mathbb{N}$, $f_n \in L^1(\mathbb{R})$.
2. Montrer que $(\|\widehat{f_n}\|_\infty)_{n\in\mathbb{N}}$ est bornée, mais que $(\|f_n\|_{L^1})_{n\in\mathbb{N}}$ n'est pas bornée.
3. En déduire que $\mathcal{F}$ n'est pas surjective.

97. On peut aussi montrer cette propriété sans le théorème d'isomorphisme de Banach, voir par exemple [Kat04, page 24].
98. Rappel : la norme sur $L^1(\mathbb{T})$ est $\|f\|_{L^1(\mathbb{T})} = \int_{\mathbb{T}} |f(t)|\frac{dt}{2\pi}$.
99. On peut aussi montrer cette propriété sans le théorème d'isomorphisme de Banach, voir par exemple [Amr08, Ex 3.5 page 171].

3.4 Théorème du graphe fermé et applications

Théorème 3.34 (Théorème du graphe fermé)**.** *Soient* $(E, \|\cdot\|_E)$ *et* $(F, \|\cdot\|_F)$ *deux espaces de Banach, et* T *une application linéaire de* E *vers* F*. Alors*

$$T \text{ est continue} \quad \Leftrightarrow \quad \text{le graphe de } T \text{ est fermé}$$

où le graphe de T *est l'ensemble* $\{(x, T(x)) \in E \times F\}$*.*

Remarque 3.35. Par définition, le graphe de T est fermé si pour toute suite $\forall (x_n)_{n\in\mathbb{N}} \in E^{\mathbb{N}}$, si on suppose que (x_n) converge vers x et $(T(x_n))_{n\in\mathbb{N}}$ converge vers y, alors il existe $x' \in E$ tel que $y = T(x')$.

En d'autres termes, dans le cadre d'application du théorème, au lieu de prouver la continuité en prouvant que pour tout $x_n \to x$ on a $T(x_n) \to T(x)$, il suffit de supposer $x_n \to x$ *et* $T(x_n) \to y$ et de prouver $y = T(x)$, c'est-à-dire qu'on peut supposer la convergence de $(T(x_n))$ et seulement vérifier que la limite est la bonne.

Démonstration. $\Rightarrow$: ce sens est facile et n'utilise pas les hypothèses de complétude : en effet si T est continu, et si (x_n) converge vers x et $T(x_n)$ converge vers y, alors nécessairement $y = T(x)$ et donc le graphe de T est fermé.

$\Leftarrow$: on suppose le graphe de T fermé. On va considérer $N : x \in E \mapsto \|x\|_E + \|T(x)\|_F$, qui est également une norme sur E. L'espace (E, N) est encore complet : en effet si $(x_n)_{n\in\mathbb{N}}$ est de Cauchy pour la norme N, alors elle est de Cauchy également pour la norme $\|\cdot\|_E$, donc par complétude de $(E, \|\cdot\|_E)$, il existe $x \in E$ tel que $x_n \to x$ pour la norme $\|\cdot\|_E$. Il reste à montrer que $x_n \to x$ pour la norme N. Pour cela, on voit que $(T(x_n))$ est aussi de Cauchy dans F, donc par complétude de F, il existe $y \in F$ tel que $T(x_n) \to y$ dans F. Mais comme le graphe de T est fermé (voir la remarque 3.35), on a $y = T(x)$, et donc x_n converge bien vers x pour la norme N.

D'autre part, $N \geq \|\cdot\|_E$, donc par l'exemple (3.31), il existe $C > 0$ tel que $C\|\cdot\|_E \geq N$, d'où a fortiori $\forall x \in E$, $\|T(x)\|_F \leq C\|x\|_E$. □

Exemple 3.36. Il n'est pas difficile de construire une application linéaire $T : E \to F$ qui a un graphe fermé mais qui n'est pas continue : on pose

$$\begin{array}{rccc} T: & (C^1([0,1]), \|\cdot\|_\infty) & \longrightarrow & (C^0([0,1]), \|\cdot\|_\infty) \\ & f & \longmapsto & f' \end{array}$$

Le graphe de T est fermé car si on suppose $f_n \to f$ et $f'_n \to g$ uniformément, alors la proposition 4.13 du Chapitre I montre que f est de classe C^1 et $f' = g$. Néanmoins, l'application T n'est pas continue comme le montre le calcul de $\|f_n\|_\infty$ et $\|T(f_n)\|_\infty$ pour $f_n : x \mapsto \sin(nx)$ et $n \in \mathbb{N}$.
Ici le Théorème 3.34 ne s'applique pas car $(C^1([0,1]), \|\cdot\|_\infty)$ n'est pas complet.

Applications du théorème du graphe fermé

Exercice 3.37. Soit E un espace de Banach, et F, G deux sev fermés de E tels que $E = F \oplus G$. Montrer que les projections sur F parallèlement à G et sur G parallèlement à F sont continues.[100]

Exercice 3.38. [101] Soit F un sev fermé de $(C^0([0,1]), \|\cdot\|_\infty)$ qui est tel que $F \subset C^1([0,1])$.

1. Montrer qu'il existe $C \in \mathbb{R}_+$ tel que $\forall f \in F$, $\|f'\|_\infty \leq C\|f\|_\infty$.
2. En déduire que F est de dimension finie[102].

100. Notons que le caractère fermé de F et G sont des conditions nécessaires pour avoir la continuité de ces projections, puisqu'ils sont les noyaux de ces projections.
101. Cet exercice est un grand classique (voir par exemple 2013 question IV.2) : il est "très accessible" (aucun calcul) mais il utilise plusieurs théorèmes, donc le jury peut vérifier à quel point le candidat sait utiliser les théorèmes qu'il a potentiellement mis dans sa leçon. Notons qu'un résultat similaire adapté aux espaces L^p et souvent appelé "lemme de Grothendieck" peut également trouver sa place à l'agrégation : il affirme que si F est un sev fermé de $L^1([0,1])$ tel que $F \subset L^\infty([0,1])$ alors $\dim(F) < \infty$, voir [Rud95].
102. On pourra trouver deux méthodes, l'une utilisant le théorème d'Ascoli, l'autre "à la main" obtenant une estimation explicite de $\dim(F)$ en fonction de C.

Exercice 3.39 (Théorème de Hellinger–Toeplitz)**.** Soit $(H, \langle\cdot,\cdot\rangle$ un espace de Hilbert, et $T : H \to H$ tel que

$$\forall (x,y) \in H^2, \ \ \langle T(x), y\rangle = \langle x, T(y)\rangle.$$

1. Soit $(x_n) \in H^{\mathbb{N}}$ et $(x,y) \in H^2$ tels que $x_n \to x$ et $T(x_n) \to y$. Montrer

$$\forall z \in H, \ \ \langle z, y\rangle = \langle z, T(x)\rangle.$$

2. Conclure que T est continu.

On termine ce paragraphe avec le résultat suivant qui améliore l'exemple 3.31 dans le cas de $E = C^0([0,1])$:

Exercice 3.40. Soit N une norme sur $C^0([0,1])$ tel que

— $(C^0([0,1]), N)$ est complet,

— pour toute suite (f_n) de $C^0([0,1])$ qui converge pour la norme N vers une limite f, on a

$$f_n \text{ converge simplement vers } f \text{ sur } [0,1].$$

Montrer que N est équivalente à $\|\cdot\|_\infty$.

3.5 Existence de contre-exemples par le Théorème de Baire (ou ses conséquences)

Le théorème de Baire est aussi un moyen de prouver de façon non constructive l'existence d'objets pathologiques :

Théorème 3.41. *L'ensemble des fonctions continues sur* $[0,1]$ *qui ne sont dérivables en aucun point de* $[0,1]$ *est dense dans l'espace* $(C^0([0,1]), \|\cdot\|_\infty)$.

Remarque 3.42. Dans cet énoncé, c'est surtout le fait que l'ensemble des fonctions continues nulle part dérivable est non vide qui est étonnant : le fait que c'est en fait dense n'est pas tellement plus fort (voir néanmoins l'exercice 3.43). En effet, si on connait $f_0 : [0,1] \to \mathbb{R}$ une fonction continue nulle part dérivable [103], alors il n'est pas difficile de voir que l'ensemble $f_0 + \mathbb{R}[X] = \{f_0 + P, P \in \mathbb{R}[X]\}$ est constitué de fonctions continues nulle part dérivables, et est dense dans $C^0([0,1])$.

Démonstration. [104] On pose $E = C^0([0,1])$ et pour tout $n \in \mathbb{N}$,

$$F_n := \{f \in E, \ \ \exists x \in [0,1], \forall y \in [0,1], \ \ |f(y) - f(x)| \leq n|y - x|.$$

Étape 1 : Montrons que F_n est fermé : soit (f_p) une suite de F_n qui converge uniformément vers $f_\infty \in E$. Alors pour tout $p \in \mathbb{N}$ il existe x_p tel que

$$\forall y \in [0,1], \ \ |f_p(y) - f_p(x_p)| \leq n|y - x_p|. \tag{3.4}$$

Par compacité de $[0,1]$, il existe φ une extraction et $x_\infty \in [0,1]$ tels que $x_{\varphi(p)} \to x_\infty$. La convergence uniforme de f_p permet de dire que $f_{\varphi(p)}(x_{\varphi(p)}) \to f_\infty(x_\infty)$, et donc en passant à la limite dans (3.4) (pour les valeurs extraites $\varphi(p)$), on obtient

$$\forall y \in [0,1], \ \ |f_\infty(y) - f_\infty(x_\infty)| \leq n|y - x_\infty|,$$

et donc $f_\infty \in F_n$ ce qui montre le caractère fermé de F_n.

Étape 2 : Montrons maintenant que F_n est d'intérieur vide : on se donne $f \in F_n$ et $\varepsilon > 0$. Par le théorème de Weierstrass, il existe un polynôme [105] P tel que $\|f - P\|_\infty < \varepsilon$. On définit maintenant

103. Notons qu'il est possible de construire explicitement de telles fonctions. Il faut donc garder en tête que la preuve donnée ici du Théorème 3.41 est non constructive, et qu'il en existe des preuves constructives.

104. On suit la preuve de [QZ13]. On peut trouver une variante dans [Gou08b].

105. En fait, ce n'est pas important que ce soit un polynôme ; mais on veut approcher f par une fonction lipschitzienne, et c'est le cas d'un polynôme. L'importance d'approcher f par une fonction lipschitzienne est ensuite de la perturber par une fonction très oscillante (g_N) qui la fera sortir de F_n, alors que si on avait perturbé f directement, on ne pourrait pas assurer que sa perturbation par g_N est très oscillante, vu qu'on ne contrôle pas les oscillations de f.

$g_N : [0,1] \to \mathbb{R}$ une fonction continue et affine par morceaux et telle que sur chaque intervalle $\left[\frac{k}{N}, \frac{k+1}{N}\right]$ avec $k \in [\![0, N-1]\!]$ on a $g_N(\frac{k}{N}) = 0$, $g_N(\frac{k+1/2}{N}) = \varepsilon$ et $g_N(\frac{k+1}{N}) = 0$. Alors $P + g_N \in B(f, 2\varepsilon)$ et

$$\begin{aligned}\forall (x,y) \in [0,1],\ \ |(P+g_N)(y) - (P+g_N)(x)| \ &\geq \ |g_N(y) - g_N(x)| - |P(y) - P(x)| \\ &\geq \ |g_N(y) - g_N(x)| - \|P'\|_\infty |y - x|.\end{aligned}$$

Pour tout $x \in [0,1]$ fixé, il appartient à un intervalle de la forme $\left[\frac{k}{N}, \frac{k+1/2}{N}\right]$ ou $\left[\frac{k+1/2}{N}, \frac{k+1}{N}\right]$ (pour un certain $k \in [\![0, N-1]\!]$) sur lequel g_N est affine de pente $2N\varepsilon$, et donc on peut trouver $y \neq x$ dans le même intervalle de sorte qu'on a

$$\forall x \in [0,1], \exists y \in [0,1],\ |g_N(y) - g_N(x)| - \|P'\|_\infty |y - x| = (2N\varepsilon - \|P'\|_\infty)\,|y - x|$$

qui est strictement positif dès que $2N\varepsilon - \|P'\|_\infty > 0$, ce qui est le cas pour N suffisamment grand ($\varepsilon > 0$ et P ne bougent pas). Ainsi $P + g_N \notin F_n$, ce qui conclut la preuve que F_n est d'intérieur vide.

Le théorème de Baire s'applique dans l'espace E qui est complet, et on en conclut que $\bigcup_n F_n$ est d'intérieur vide, et que son complémentaire est dense dans E.

Conclusion : Si $f \in E$ est une fonction continue dérivable en un point $x_0 \in [0,1]$, alors la fonction

$$y \mapsto \begin{cases} \frac{f(y) - f(x_0)}{y - x_0} & \text{si } y \neq x_0 \\ f'(x_0) & \text{si } y = x_0 \end{cases}$$

est continue donc bornée sur $[0,1]$, ce qui montre que $f \in \bigcup_n F_n$. Ainsi l'ensemble des fonctions continues nulle part dérivable est inclus dans le complémentaire de $\bigcup_n F_n$, et est donc dense dans E. □

Exercice 3.43. [*] On pose $\mathcal{A} \subset C^0([0,1])$ l'ensemble des fonctions continues nulle part dérivables. La preuve du théorème 3.41 donnée ici montre en fait que $\mathcal{A}$ est un G_δ dense dans $C^0([0,1])$, ce qui est mieux que juste de savoir que c'est un ensemble dense. On voit ici une application :

1. Montrer que si $f \in C^0([0,1])$ alors $\{f - g, g \in \mathcal{A}\}$ est un G_δ dense dans $C^0([0,1])$.
2. Montrer que toute fonction de $C^0([0,1])$ est somme de deux fonctions continues nulle part dérivables.

Théorème 3.44. *L'ensemble des fonctions $f : \mathbb{R} \to \mathbb{C}$ continues et 2π-périodiques pour lesquelles la série de Fourier ne converge pas simplement vers f est dense dans $(C^0(\mathbb{T}), \|\cdot\|_\infty)$.*

Remarque 3.45. Comme pour la remarque 3.42, c'est surtout le fait qu'il existe une fonction continues dont la série de Fourier ne converge pas simplement qui est intéressant. On peut en effet prouver le théorème à partir du moment où on sait que l'ensemble des fonctions continues dont la série de Fourier ne converge pas simplement est non vide. A nouveau, il est intéressant de noter qu'on peut construire explicitement une telle fonction [106] et ainsi obtenir une preuve constructive du Théorème 3.44.

Démonstration. On pose $E = C^0(\mathbb{T})$ et pour tout $N \in \mathbb{N}$,

$$\forall f \in E,\ \ T_N(f) = \sum_{n=-N}^{N} \widehat{f}(n)$$

qui est la valeur en 0 de la somme partielle de la série de Fourier de f, et qui est une forme linéaire sur E. On a vu à la Remarque 1.14 du Chapitre VIII que [107] [108]

$$\forall f \in E,\ \ T_N(f) = D_N * f(0) = \frac{1}{2\pi}\int_{\mathbb{T}} f(t) D_N(t)dt, \qquad \text{où } \forall x \in \mathbb{R}, D_N(x) = \sum_{n=-N}^{N} e^{int}.$$

106. Voir par exemple [Gou08b, page 264]
107. Attention on rappelle que dans le cadre des fonctions 2π-périodique, on considère la convolution

$$f * g(x) = \frac{1}{2\pi}\int_{\mathbb{T}} f(t) g(x - t)dt.$$

108. Logiquement on aurait dû écrire $\int_{\mathbb{T}} f(t) D_N(-t)dt$ mais D_N est paire.

Ceci montre que T_N est continue et

$$\|T_N\| \leq \frac{1}{2\pi}\int_{\mathbb{T}} |D_N(t)|dt.$$

En considérant les fonctions $(f_\varepsilon = \frac{D_N}{|D_N|+\varepsilon})_{\varepsilon>0}$ qui sont bien des éléments de E, on en déduit en fait $\|T_N\| = \frac{1}{2\pi}\int_{\mathbb{T}} |D_N(t)|dt$.

D'autre part, on a vu au Chapitre VIII que

$$\forall x \in \mathbb{R}\setminus 2\pi\mathbb{Z},\ \ D_N(x) = \frac{\sin\left((N+1/2)x\right)}{\sin(x/2)}$$

et donc avec l'inégalité classique $|\sin(y)| \leq |y|$ (valable sur $\mathbb{R}$) on obtient

$$\int_{\mathbb{T}} |D_N(t)|dt \geq \int_{\mathbb{T}} \left|\frac{\sin\left((N+1/2)t\right)}{t/2}\right| dt = 2\int_{-(N+1/2)\pi}^{(N+1/2)\pi} \left|\frac{\sin(y)}{y}\right| dy$$

et comme la fonction $y \mapsto \frac{\sin(y)}{y}$ n'est pas intégrable sur $\mathbb{R}$, $\|T_N\| \xrightarrow[N\to\infty]{} +\infty$. Ainsi le Théorème 3.22 s'applique et donc il existe un ensemble dense dans E de fonctions f telles que $(T_N(f))_{N\in\mathbb{N}}$ n'est pas bornée, et donc en particulier la série de Fourier de ces fonctions ne converge pas en 0. □

On a vu au Chapitre I que lorsqu'une fonction $f : [0,1] \to \mathbb{R}$ est de classe C^1 [109], alors la méthode des rectangles à gauche a une erreur en $O(1/n)$. Plus précisément, si on pose

$$\forall f \in C^0([0,1]),\ \forall n \in \mathbb{N},\ \ R_n(f) = \frac{1}{n}\sum_{k=0}^{n-1} f\left(\frac{k}{n}\right)$$

alors

$$\forall f \in C^1([0,1]),\ \forall n \in \mathbb{N},\ \ \left|\int_0^1 f(t)dt - R_n(f)\right| \leq \frac{\|f'\|_\infty}{2n}.$$

Le résultat suivant montre qu'on ne peut pas avoir une convergence en $O(1/n)$ pour toute fonction continue (même si on sait qu'il y a convergence pour toute fonction continue, voir la proposition 5.16 au Chapitre I) :

Théorème 3.46. *L'ensemble des fonctions continues sur* $[0,1]$ *et telles que la suite*

$$n\left|\int_0^1 f(t)dt - R_n(f)\right|$$

n'est pas bornée, est dense dans $C^0([0,1])$.

Démonstration. Soit $n \in \mathbb{N}$. On pose $E = C^0([0,1])$ et

$$\forall f \in C^0([0,1]),\ \ T_n(f) = n\left[\int_0^1 f(t)dt - R_n(f)\right].$$

L'application T_n est une forme linéaire continue sur E puisque

$$\forall f \in C^0([0,1]),\ \ |T_n(f)| \leq 2n\|f\|_\infty.$$

On en déduit que $\|T_n\|_{E'} \leq 2n$. On se donne maintenant $\varepsilon > 0$, et on considère la fonction f_ε : $[0,1] \to \mathbb{R}$ continue, qui vaut 1 sur $[0,1]$ sauf sur les intervalles $([k/n-\varepsilon, k/n+\varepsilon])_{k\in[\![0,n-1]\!]}$ où pour tout $k \in [\![0,n-1]\!]$ elle est affine sur $[k/n-\varepsilon, k/n]$ et sur $[k/n, k/n+\varepsilon]$ et telle que $f(k/n) = -1$. On obtient :

$$\int_0^1 f_\varepsilon(t)dt = 1-(n-1)\frac{2\varepsilon.2}{2} - \frac{1}{2}\frac{2\varepsilon.2}{2} = 1+\varepsilon-2n\varepsilon,\quad \frac{1}{n}\sum_{k=0}^{n-1} f(k/n) = -1.$$

Ainsi $\|T_n\| \geq \frac{|T_n(f_\varepsilon)|}{\|f_n\|_\infty} = n\Big[2-(2n-1)\varepsilon\Big]$ et en faisant tendre ε vers 0 on obtient $\|T_n\| = 2n$.

La famille (T_n) n'est donc pas bornée dans E' or E est complet donc le Théorème 3.22 de Banach-Steinhaus s'applique et affirme l'existence d'un ensemble dense dans E de fonctions f telles que $(T_n(f))_{n\in\mathbb{N}}$ n'est pas bornée. □

109. En fait on peut obtenir une telle estimation si f est seulement supposée Lipschitzienne.

Chapitre X

Equations différentielles

La théorie des équations différentielles intervient relativement rarement de manière centrale dans les sujets d'écrits de l'agrégation externe, mais elle intervient quand même régulièrement dans certaines parties des sujets : en effet les équations différentielles apparaissent souvent naturellement dans le cadre du programme, soit de manière calculatoire, soit de manière théorique (voir par exemple les parties I des sujets 2008, 2012, 2017 et 2021). On notera quand même que les équations différentielles sont relativement prépondérantes dans les sujets 2010 et 1998, ainsi que dans le sujet 2020 du problème d'analyse de l'agrégation docteur.

Dans la liste des leçons de l'oral, les équations différentielles sont au cœur des leçons **220-221** ; mais elles peuvent également être valorisées de façon transverse dans de nombreux sujets : en analyse fonctionnelle[1], en algèbre linéaire, en calcul différentiel, en analyse numérique, ou plus simplement dans l'utilisation de l'analyse d'une variable réelle.

On distinguera autant que possible deux aspects assez différents de la théorie : d'une part l'aspect calculatoire, qui consiste à savoir résoudre certains types d'équations différentielles. Cette aspect n'est évidemment pas à négliger, puisque le jury pourrait justement s'offenser de voir un candidat patauger dans la résolution d'une équation linéaire simple, sujet que ce dernier serait très susceptible d'enseigner en tant qu'agrégé. D'autre part, l'aspect qualitatif, qui consiste à analyser et comprendre les solutions d'une équation différentielle qu'on ne peut pas résoudre explicitement[2]. Ce second sujet est certainement plus ambitieux et peut amener à des développements plus ou moins sophistiqués.

En terme de références, on peut citer le classique [HW15] qui donne une présentation originale avec de nombreux dessins, mais qui demande probablement un gros investissement. Les livres [Dem16] et [BG10] sont certainement plus abordables ; j'ai aussi un excellent souvenir de la lecture de la seconde partie de [GT98], même si c'est probablement mieux de le réserver pour une "seconde" lecture sur les équations différentielles. On pourra également consulter les polycopiés [Ton, Boyb].

1. Dans le cadre du programme, le théorème de Cauchy-Lipschitz est probablement la quintescence de "l'application de l'analyse fonctionnelle" à une situation "concrète" ; en effet il répond à une question très naturelle (et qui intervient très fréquemment en mathématiques, mais aussi dans toutes les sciences), à savoir qu'il existe une unique solution à un certain type d'équation fonctionnelle, et sa démonstration repose sur le théorème de point fixe et donc la notion d'espace complet.

2. Remarquons qu'il n'est pas simple de comprendre ce que signifie "ne peut pas être résolu explicitement", mais qu'on peut le formaliser rigoureusement : c'est l'objet de la théorie de Galois différentielle, qui au même titre que la théorie de Galois "standard" explique qu'on ne peut pas résoudre "explicitement" certaines équations polynomiales, explique qu'on ne peut pas résoudre "explicitement" certaines équations différentielles. Le cas le plus célèbre d'une telle équation est

$$y''(t) = -ty'(t)$$

dont la résolution serait équivalente à l'identification d'une primitive de la fonction $x \mapsto e^{-t^2/2}$, et le théorème de Liouville affirme qu'une telle résolution/identification est impossible "explicitement". Ces considérations sortent assez largement du programme, même si on peut trouver une approche élémentaire dans le sujet ENS Lyon-Cachan 1995. On peut également consulter ce lien avec une jolie introduction à la théorie de Galois différentielle (et sa comparaison avec la théorie de Galois pour les équations polynomiales) et qui détaille le cas de l'équation $y' = t - y^2$.

1 Théorie de Cauchy-Lipschitz

1.1 Vocabulaire

Définition 1.1. — On appelle équation différentielle[3][4] d'ordre $n \in \mathbb{N}^*$ une équation de la forme

$$y^{(n)} = f(t, y, \dots, y^{(n-1)}) \tag{1.1}$$

où $f : I \times U_0 \times \dots U_{n-1} \to \mathbb{K}^d$[5] est une application avec I un intervalle de $\mathbb{R}$, et $U_0, \dots, U_{n-1}$ sont des ouverts de $\mathbb{K}^d$.

— On appelle solution de (1.1) la donnée d'un couple (J, y) où J est un intervalle[6] inclus dans I et $y : J \to \mathbb{K}^d$ est de classe C^n[7] et telle que pour tout $t \in J$ et tout $k \in [\![0, n]\!]$, $y^{(k)}(t) \in U_k$, et

$$\forall t \in J, \ \ y^{(n)}(t) = f(t, y(t), \dots, y^{(n-1)}(t)). \tag{1.2}$$

Plus précisément :

1. on dira que la solution (J, y) est maximale s'il n'existe pas $(\widetilde{J}, \widetilde{y})$ solution de (1.8) telle que
$$\widetilde{J} \supset J, \ \widetilde{J} \neq J, \ \forall t \in J, \widetilde{y}(t) = y(t),$$
autrement dit si la solution (J, y) ne peut pas être prolongée en une solution définie sur un intervalle plus grand que J,
2. on dit que la solution (J, y) est globale si $J = I$.

Exemple 1.2. Si $f : I \times U \to \mathbb{K}^d$, où I est un intervalle de $\mathbb{R}$ et U un ouvert de $\mathbb{K}^d$,

$$y' = f(t, y) \tag{1.3}$$

est une équation différentielle d'ordre 1, et une solution est la donnée d'un couple (J, y) où J est un intervalle inclus dans I et $y : J \to U$ est de classe C^1 et telle que

$$\forall t \in J, \ \ y'(t) = f(t, y(t)).$$

Exercice 1.3 (Régularité des solutions)**.** On considère (J, y) une solution de l'équation (1.3) et on suppose f de classe C^k avec $k \in \mathbb{N}^* \cup \{+\infty\}$. Montrer que y est de classe C^{k+1}.

---

3. On précise souvent équation différentielle ordinaire (EDO pour les intimes, et ODE pour les anglophones), pour signifier que la fonction inconnue y ne dépend que d'une seule variable (qu'on note le plus souvent t en référence à la variable temporelle en physique, même si en pratique il arrivera que cette variable modélise autre chose que le temps), et par opposition aux équations aux dérivées partielles pour lesquelles les fonctions inconnues dépendent d'au moins deux variables.

4. Certains appellent (1.1) une équation différentielle sous forme "résolue" (pas la meilleure terminologie à mon avis), par opposition à une équation différentielle "générale" qui serait sous la forme

$$F(t, y, \dots, y^{(n-1)}, y^{(n)}) = 0.$$

Aucun énoncé n'est vraiment adapté à une forme si générale d'équation, et quand on est face à une telle situation, on s'efforce de commencer par la mettre sous la forme (1.1), ce qui nécessite parfois une certaine vigilance.

5. Comme d'habitude, $\mathbb{K}$ désigne $\mathbb{R}$ ou $\mathbb{C}$. Notons qu'on autorise $\mathbb{K} = \mathbb{C}$ par commodité, mais qu'on pourrait se restreindre à $\mathbb{K} = \mathbb{R}$ et considérer que $\mathbb{C}^d$ est isomorphe à $\mathbb{R}^{2d}$; d'ailleurs le programme mentionne seulement le cas $\mathbb{R}^d$. Ceci n'est pas toujours commode néanmoins, notamment quand on abordera les équations linéaires où il pourra être naturel de se plonger dans $\mathbb{C}$ pour diagonaliser ou trigonaliser les matrices.

6. Précisons qu'en général on demande que J ne soit pas restreint à un singleton, sinon ça n'a pas beaucoup de sens (on ne peut même pas définir la dérivée d'une fonction définie seulement sur un singleton...). Aussi, il arrive que par convention on demande que J soit un intervalle ouvert ; c'est un peu restrictif mais ça simplifie un peu la présentation. Il est intéressant de noter a posteriori que ce n'est pas si restrictif que ça, puisqu'on montrera que les intervalles maximaux sont toujours ouverts, voir le corollaire 1.25.

7. On pourrait trouver plus logique de demander seulement y n-fois dérivables. Si la fonction f est continue, ce qui sera toujours le cas en pratique, cela ne fait pas de différence : en effet si on suppose y n-fois dérivables, alors les fonctions $y, \dots, y^{(n-1)}$ sont continues, et donc par composition (1.1) montre que $y^{(n)}$ est continue.

Remarque 1.4. — On pourrait plus généralemet considérer E un espace de Banach, à la place de $\mathbb{K}^d$. Mais sauf si vous avez des applications concrètes à l'esprit, on déconseille de considérer ce cas. Il peut être intéressant de noter néanmoins que certains aspects restent inchangés dans ce cadre (par exemple la théorie de Cauchy-Lipschitz) mais que d'autres ne sont plus valables (comme le théorème de Péano).

— La définition d'une solution paraît plus compliquée qu'elle ne l'est en réalité ; il faut surtout retenir qu'une solution doit toujours être associée à l'ensemble de définition sur lequel elle est définie. Toutes les conditions qui sont faites sont là juste pour que (1.2) ait un sens.

— L'écriture (1.1) est abusive : en effet on fait prendre à f à la fois le paramètre t qui est réel, et les paramètres $y, \dots, y^{(n-1)}$ qui sont des fonctions. C'est un abus classique dans la théorie des équations différentielles, on l'utilisera donc, mais on met en garde le lecteur sur les erreurs qu'il peut amener[8].

— On ne distinguera pas vraiment ici la notion d'équation différentielle et de système différentiel ; en fait on peut voir (1.1) comme une équation, mais aussi comme un système d'équation si $E = \mathbb{K}^d$ avec $d \geq 2$: dans ce cas et avec $n = 1$, on écrit $y = (y_1, \dots, y_d)$ et $f = (f_1, \dots, f_d)$, et (1.1) et (1.2) deviennent

$$\begin{cases} y_1' = f_1(t, y_1, y_2, \dots, y_d) \\ \vdots \\ y_d' = f_d(t, y_1, y_2, \dots, y_d) \end{cases} \quad \text{i.e.} \quad \forall t \in J, \begin{cases} y_1'(t) = f_1(t, y_1(t), y_2(t), \dots, y_d(t)) \\ \vdots \\ y_d'(t) = f_d(t, y_1(t), y_2(t), \dots, y_d(t)) \end{cases}$$

On dira par contre que l'équation différentielle est scalaire si $d = 1$, c'est-à-dire $E = \mathbb{K}$.

Remarque 1.5. Au regard de la définition, une solution globale est maximale. Le contraire n'est pas vrai en général et c'est d'ailleurs une question fréquente de savoir si une solution maximale est globale. Lorsqu'une résolution explicite est possible, il n'est pas trop difficile de décider : considérons par exemple l'équation différentielle

$$y' = y^2 \tag{1.4}$$

pour laquelle $I = \mathbb{R}$ et $d = 1$. L'ensemble des solutions maximales sont

$$(\mathbb{R}, 0), \quad \left(t \in]-\infty, \alpha[\mapsto \frac{1}{\alpha - t}\right), \quad \left(t \in]\alpha, \infty[\mapsto \frac{1}{\alpha - t}\right)$$

où α est un réel quelconque ; seule la première solution (la fonction nulle) est globale ; toutes les autres sont maximales (ayant une limite infinie en la borne finie de leur intervalle de définition, on ne peut pas prolonger ces fonctions de manière continue) mais ne sont pas globales.

Intuitivement, le lecteur peut retenir que la relation (1.4) implique que y' va augmenter "trop" rapidement quand t augmente (à cause du carré), et ainsi que y va "exploser". On reviendra sur le caractère général de cette situation pour des solutions maximales non globales.

Remarque 1.6 (Réduction à des équations d'ordre 1)**.** Dans cette remarque, on explique comment toute équation différentielle d'ordre $n \geq 2$ peut en fait être ramenée à une équation différentielle d'ordre 1 en modifiant la valeur de d : on se donne $I, d, U_0, \dots U_{n-1}$ et $f : I \times U_0 \times \dots U_{n-1}$ comme dans la définition 1.1.

Si (J, y) est une solution de

$$y^{(n)} = f(t, y, \dots, y^{(n-1)}) \tag{1.5}$$

alors on pose

$$\forall t \in J,\ Y(t) = \begin{pmatrix} y(t) \\ \vdots \\ y^{(n-1)}(t) \end{pmatrix} \in U_0 \times \dots U_{n-1} \subset (\mathbb{K}^d)^n \tag{1.6}$$

8. Notamment, quand on parlera de la régularité de f, on parlera bien de la régularité de l'application $f : (t, v_0, \dots, v_{n-1}) \mapsto f(t, v_0, \dots, v_{n-1})$ qui est une fonction définie sur un sous-ensemble de $\mathbb{R} \times (\mathbb{K}^d)^n$, ce qui n'est pas du tout la même chose que la régularité de l'application $t \mapsto f(t, y(t), \dots y^{(n-1)}(t))$ où $y : J \to E$ est une fonction de classe C^n.

qui est une application $Y : J \to U_0 \times \ldots U_{n-1}$ de classe C^1 et qui satisfait

$$\forall t \in J, \;\; Y'(t) = g(t, Y(t))$$

où

$$g : \begin{array}{ccc} I \times U_0 \times \ldots U_{n-1} & \longrightarrow & E^n \\ (t, v_0, \ldots v_{n-1}) & \longmapsto & (v_1, \ldots, v_{n-1}, f(t, v_0, \ldots v_{n-1})) \end{array}.$$

Réciproquement, toute solution (J, Y) de l'équation

$$Y' = g(t, Y) \tag{1.7}$$

est en fait de la forme (1.6) pour une certaine fonction y qui sera solution de (1.5). En conclusion les équations (1.5) et (1.7) sont équivalentes.

Pour cette raison, la plupart des définitions et énoncés donnés dans la suite seront formulés seulement pour les équations d'ordre 1 ; il faut savoir appliquer la méthode décrite ci-dessus pour pouvoir appliquer ces définitions/résultats aux équations d'ordre n.

Remarque 1.7 (*)**.** Toute solution peut être prolongée en une solution maximale [9]. Plus précisément, si $y : J \to E$ est une solution de (1.3), alors il existe $(\widetilde{J}, \widetilde{y})$ solution maximale de (1.3) et telle que

$$J \subset \widetilde{J}, \quad \text{et} \quad \forall t \in J, \;\; \widetilde{y}(t) = y(t).$$

On renvoit par exemple à [Dem16, Page 128] pour une démonstration.

Définition 1.8 (Problème de Cauchy)**.** On appelle problème de Cauchy une équation différentielle à laquelle on adjoint une condition initiale :

— plus précisément, étant donnés I un intervalle de $\mathbb{R}$, U un ouvert de $\mathbb{K}^d$, $f : I \times U \to \mathbb{K}^d$, et $(t_0, y_0) \in I \times U$, on appelle problème de Cauchy en t_0 le problème

$$\left\{ \begin{array}{rcl} y' & = & f(t, y) \\ y(t_0) & = & y_0 \end{array} \right. . \tag{1.8}$$

— On appelle solution du problème de Cauchy (1.8) la donnée d'un couple (J, y) où J est un intervalle inclus dans I avec $t_0 \in J$, et

$$\left\{ \begin{array}{rcl} y'(t) & = & f(t, y(t)), \;\; \forall t \in J \\ y(t_0) & = & y_0 \end{array} \right. .$$

On dira que la solution (J, y) est une solution locale du problème de Cauchy si J est un intervalle tel que $t_0 \in \mathring{J}$. Comme à la définition 1.1, on pourra aussi parler de solution maximale et globale du problème de Cauchy.

Remarque 1.9. Comme on l'a précisé à la remarque 1.6, il faut savoir faire la gymnastique d'adapter la définition précédente au cas d'une équation différentielle d'ordre n ; avec les notations de la définition 1.1, un problème de Cauchy est de la forme

$$\left\{ \begin{array}{rcl} y^{(n)} & = & f(t, y, \ldots, y^{(n-1)}) \\ y(t_0) & = & y_0 \\ & \vdots & \\ y^{(n-1)}(t_0) & = & y_{n-1} \end{array} \right. .$$

où $(y_0, \ldots, y_{n-1}) \in U_0 \times \ldots \times U_{n-1}$ est donné.

9. On démontrera ce résultat dans le cadre de la théorie de Cauchy-Lipschitz, en prouvant en fait existence et unicité, sous les hypothèses ad hoc sur f ; mais il est intéressant de voir que sans hypothèse, toute solution (J, y) peut être étendue en une solution maximale (mais pas de manière unique a priori).

1.2 Théorème de Cauchy-Lipschitz local

Voici LE théorème de ce chapitre ; il faut bien sûr bien comprendre l'énoncé et savoir l'appliquer (ce qui est moins souvent le cas que ce qu'on imagine). Notons également que les idées de la démonstration sont intéressantes en soit, et on conseille de connaître au moins le point de départ qui consiste à remplacer l'équation différentielle par sa version intégrale, et de voir celle-ci comme une équation de point fixe. Notons également qu'on parle de "la" démonstration, mais il existe plusieurs stratégies : on se concentre ici sur la méthode par théorème de point fixe. Nous renvoyons au paragraphe 3 pour une autre approche.

Théorème 1.10 (Cauchy-Lipschitz local)**.** *Soit U un ouvert de $\mathbb{K}^d$, I un intervalle ouvert de $\mathbb{R}$, $f : I \times U \to \mathbb{K}^d$ une fonction. On suppose*

1. *f est continue sur $I \times U$,*
2. *f est localement lipschitzienne en la seconde variable [10], c'est-à-dire que pour tout $(t_1, y_1) \in I \times U$, il existe ω un voisinage ouvert de (t_1, y_1) dans $I \times U$ et une constante L telle que*

$$\forall (t, y), (t, \hat{y}) \in \omega^2, \ \ \|f(t, y) - f(t, \hat{y})\| \leq L\|y - \hat{y}\|. \tag{1.9}$$

Alors pour tout $(t_0, y_0) \in I \times U$, il existe un intervalle ouvert J contenant t_0 pour lequel le problème (1.8) admet une solution y définie sur J, et cette solution est unique.

Remarque 1.11. La première confusion fréquente est d'omettre l'hypothèse de continuité de f en pensant que la seconde hypothèse l'implique ; il n'en est rien car on a demandé le caractère lipschitzien de f seulement en la seconde variable, ce qui se traduit par le fait que dans (1.9) on a bien (t, y) et $(t, \hat{y})$ et non (t, y) et $(\hat{t}, \hat{y})$.

Remarque 1.12. Comme d'habitude [11], on peut montrer que f est localement lipschitzienne en la seconde variable au sens précédent si et seulement si pour tout compact $K \subset I \times U$ il existe $L_K \in \mathbb{R}$ telle que

$$\forall (t, y), (t, \hat{y}) \in K^2, \ \ \|f(t, y) - f(t, \hat{y})\| \leq L_K\|y - \hat{y}\|.$$

Exemple 1.13. Si on suppose $f : I \times U \to \mathbb{K}^d$ de classe C^1 [12], alors elle satisfait les hypothèses du théorème 1.10 : en effet, d'une part elle est continue, et d'autre part par le théorème des accroissements finis, on a

$$\forall \big((t, y), (t, \hat{y})\big) \in (J \times U)^2, \ [y, \hat{y}] \subset U \ \Rightarrow \ \|f(t, y) - f(t, \hat{y})\| \leq \left(\sup_{t \in J, z \in [y, \hat{y}]} \|\partial_2 f(t, z)\| \right) \|y - \hat{y}\|.$$

Donc si $(t_0, y_0) \in I \times U$, comme I et U sont ouverts, il existe $\varepsilon > 0$ tel que $\omega =]t_0 - \varepsilon, t_0 + \varepsilon[\times B(y_0, \varepsilon)$ est tel que $\overline{\omega}$ est inclus dans $I \times U$; comme c'est un compact et que $\partial_2 f$ est continue, et que de plus $B(y_0, \varepsilon)$ est convexe, on a

$$\forall \big((t, y), (t, \hat{y})\big) \in \omega^2, \ \ \sup_{t \in J, z \in [y, \hat{y}]} \|\partial_2 f(t, z)\| \leq \sup_{(t,z) \in \omega} \|\partial_2 f(t, z)\| < +\infty.$$

10. Évidemment, parler de seconde variable nécessite d'avoir bien ordonné les variables (autrement dit écrire (t, y) et non (y, t)), ce qui est complètement arbitraire ; on peut préférer retenir "en la variable d'état". Aussi, si l'équation est autonome ($y' = f(y)$) on peut se demander ce que signifie la dépendance en "la seconde variable" : il faut alors voir f comme la fonction de deux variables $(t, y) \mapsto f(y)$. Encore une fois, le plus important n'est pas le choix des mots mais de savoir ce que ceux-ci sous-entendent.

11. Faites le en exercice ! On peut soit faire une preuve via la définition séquentielle de la compacité, ou alors par la définition par recouvrement de la compacité.

12. Attention ! Ici réside l'erreur la plus fréquente dans l'application de ce théorème (voir aussi la note 8) : quand on demande à un étudiant de justifier l'application du théorème de Cauchy-Lipschitz dans ce cadre, on entend très fréquemment quelque chose du genre : "comme y est de classe C^1 et f aussi, alors $t \mapsto f(t, y(t))$ est de classe C^1 donc est localement lipschitzienne". Cette argumentation est non seulement erronée, mais elle n'a même pas vraiment de sens :
— en premier lieu en effet, elle fait référence à une fonction y dont on ignore encore l'existence, puisque c'est l'objet du théorème de montrer qu'une telle fonction existe,
— en second lieu, elle témoigne de l'incapacité de faire la différence entre la régularité de f dont les variables sont $(t, y) \in \mathbb{R} \times \mathbb{K}^d$, et la régularité de $t \mapsto f(t, y(t))$ où y est elle-même une fonction. Ces deux questions ont du sens, mais dans le cadre du théorème de Cauchy-Lipschitz, c'est bien de la première dont il s'agit, c'est-à-dire qu'on voit f comme une fonction des deux variables $(t, y) \in \mathbb{R} \times \mathbb{K}^d$ et on étudie sa régularité.

Exemple 1.14 (Contre-exemple à l'unicité)**.** On ne va pas pouvoir produire un contre-exemple à l'existence, car le théorème de Peano affirme que sous la seule hypothèse de continuité de f, il existe une solution à (1.8). Néanmoins, on peut fournir des contre-exemples à la propriété d'unicité : si on considère le problème de Cauchy

$$\left\{\begin{array}{lll} y' & = & \sqrt{|y|} \\ y(0) & = & 0 \end{array}\right. \tag{1.10}$$

on peut constater que les fonctions

$$\Big[\forall t \in \mathbb{R}, \quad y(t)=0\Big], \ \left[\forall t \in \mathbb{R}, \quad y_\alpha(t) = \left\{\begin{array}{ll} -\frac{1}{2}(t-\alpha)^2 & \text{si } t \leq \alpha \\ 0 & \text{sinon} \end{array}\right.\right], \ \left[\forall t \in \mathbb{R}, \quad y_\beta(t) = \left\{\begin{array}{ll} 0 & \text{si } t \leq \beta \\ \frac{1}{2}(t-\beta)^2 & \text{sinon} \end{array}\right.\right]$$

où $\alpha \leq 0$, $\beta \geq 0$ ou même encore

$$\forall t \in \mathbb{R}, \quad y_{(\alpha,\beta)}(t) = \left\{\begin{array}{ll} -\frac{1}{2}(t-\alpha)^2 & \text{si } t \leq \alpha \\ 0 & \text{si } \alpha \leq t \leq \beta \\ \frac{1}{2}(t-\beta)^2 & \text{si } t \geq \beta \end{array}\right.$$

où $\alpha \leq 0 \leq \beta$, sont toutes des solutions distinctes de (1.10).

Exercice 1.15. Montrer que le problème de Cauchy

$$\left\{\begin{array}{lll} y' & = & y^{1/3} \\ y(0) & = & 0 \end{array}\right.$$

admet une infinité de solutions.

Démonstration du théorème de Cauchy-Lipschitz :

Étape préliminaire : Mise sous forme intégrale (à connaitre) : supposons que (J, y) est une solution de (1.8). Alors en intégrant l'égalité $y' = f(t, y)$ entre t_0 et $t \in J$, on obtient

$$\forall t \in J, y(t) = y_0 + \int_{t_0}^{t} f(s, y(s))ds. \tag{1.11}$$

Réciproquement, soit J est un intervalle inclus dans I et contenant t_0, et $y : J \to U$ une fonction continue [13] qui satisfait (1.11). Alors comme f est continue, la fonction $s \mapsto f(s, y(s))$ est continue sur J, et le théorème fondamental de l'analyse (théorème 5.12 au Chapitre I) montre que $t \mapsto \int_{t_0}^{t} f(s, y(s))ds$ est de classe C^1 sur J. Ainsi y est de classe C^1, et en dérivant (1.11), on obtient $y'(t) = f(t, y(t))$ pour tout $t \in J$. De plus en prenant $t = t_0$ dans (1.11), on conclut que y est solution de (1.8).

Au final, on a montré que

$$\left(y \in C^1(J, U) \text{ est solution de } \left\{\begin{array}{lll} y' & = & f(t,y) \\ y(t_0) & = & y_0 \end{array}\right.\right) \iff \Big(y \in C^0(J, U) \text{ est solution de } y = \phi(y)\Big)$$

où on a posé

$$\begin{array}{llll} \phi : & C^0(J, U) & \longrightarrow & C^0(J, \mathbb{K}^d) \\ & y & \longmapsto & \left[t \mapsto y_0 + \displaystyle\int_{t_0}^{t} f(s, y(s))ds\right]. \end{array}$$

Application du théorème de point fixe : [14] pour $\alpha > 0$ on pose $J_\alpha = [t_0 - \alpha, t_0 + \alpha]$ et pour $r > 0$ on pose $\overline{B}_r := \overline{B}(y_0, r)$. On commence par remarquer qu'il existe α_0 tel que $J_{\alpha_0} \subset I$ et r_0 tel que $\overline{B}_{r_0} \subset U$, puisque I et U sont supposés ouverts. Enfin on se donne $\alpha \leq \alpha_0$.

13. Il n'est pas anodin de supposer la fonction y seulement continue ; c'est l'équation (1.11) qui "portera" la régularité de y, et non une hypothèse faite sur y directement. Cela permet d'appliquer le théorème de point fixe à un espace de fonctions continues.

14. Il y a plusieurs obstacles a priori à l'application du théorème de point fixe. En premier lieu, il n'est pas clair que l'espace $C^0(J, U)$ est stable, c'est-à-dire que dans l'espace d'arrivée de ϕ, les fonctions soient bien à valeurs dans U. Aussi, comme U est ouvert, l'espace n'est pas complet. Mais même si $U = \mathbb{K}^d$, auquel cas les deux problèmes précédents s'envolent, la fonction ϕ n'est pas contractante sans hypothèse supplémentaire. On va profité du caractère local pour mieux estimer $\|\phi(y_1) - \phi(y_2)\|_\infty$.

1. L'espace
$$E_\alpha = C^0(J_\alpha, \overline{B}_{r_0}) = \{y : [t_0-\alpha, t_0+\alpha] \to \mathbb{K}^d,\ \ \forall t \in [t_0-\alpha, t_0+\alpha], \|y(t)-y_0\| \leq r_0\}$$
est inclus dans $C^0(J_\alpha, U)$ et si on le munit de la métrique
$$\forall (y_1, y_2) \in E_\alpha,\ d(y_1, y_2) = \|y_1 - y_2\|_\infty,$$
il est complet [15] : en effet, si $(y_n)_{n\in\mathbb{N}}$ est une suite de Cauchy de E_α, alors par complétude de $(C^0(J_\alpha, \mathbb{K}^d), \|\cdot\|_\infty)$ il existe $y_\infty \in C^0(J_\alpha, \mathbb{K}^d)$ telle que $y_n \xrightarrow[n\to+\infty]{\|\cdot\|_\infty} y_\infty$. De plus, la condition
$$\forall n \in \mathbb{N}, \forall t \in [t_0-\alpha, t_0+\alpha], \|y_n(t) - y_0\| \leq r_0$$
passe à la limite $n \to +\infty$ donc $y_\infty \in E_\alpha$ [16].
2. Avec la remarque 1.12, on note L_0 la constante associée au compact $J_{\alpha_0} \times \overline{B}_{r_0} \subset U$, pour laquelle on a
$$\forall t \in J_{\alpha_0}, \forall (y, \hat{y}) \in \overline{B}_{r_0}^2,\ \ \|f(t,y) - f(t,\hat{y})\| \leq L_0\|y-\hat{y}\|.$$
3. Posons $M = \sup_{J_{\alpha_0}\times\overline{B}_{r_0}} \|f\|$, qui est fini car f est continue et $J_{\alpha_0}\times\overline{B}_{r_0}$ est compact. Soit $y \in E_\alpha$. Alors
$$\forall t \in J_\alpha, \|\phi(y)(t) - y_0\| = \left\|\int_{t_0}^t f(s, y(s))ds\right\| \leq \int_{t_0}^t \|f(s,y(s))\|\, ds \leq M|t-t_0|. \tag{1.12}$$
Ainsi l'espace E_α est stable si $M\alpha \leq r_0$.
4. Soit $(y_1, y_2) \in E_\alpha^2$. Alors
$$\begin{aligned}\forall t \in J_\alpha, \|\phi(y_2)(t) - \phi(y_1)(t)\| &= \left\|\int_{t_0}^t \Big[f(s,y_2(s)) - f(s,y_1(s))\Big]ds\right\| \\ &\leq \int_{t_0}^t \|f(s,y_2(s)) - f(s,y_1(s))\|\, ds \\ &\leq \int_{t_0}^t L_0\|y_2(s) - y_1(s)\|ds \leq \alpha L_0 \|y_2 - y_1\|_\infty.\end{aligned}$$
Il y a donc contractance de la fonction ϕ si $\alpha L_0 < 1$.

Pour conclure, on choisit donc α suffisamment petit pour avoir $\alpha M \leq r_0$ et $\alpha L_0 < 1$, et alors par le théorème de point fixe de Picard (théorème 1.81 au Chapitre III), il existe une unique fonction $y \in E_\alpha$ telle que $\phi(y) = y$, et par l'étape précédente, y est bien solution locale de (1.8).

Unicité : L'étape précédente a permis de montrer l'unicité de la solution dans l'espace E_α, mais il pourrait y avoir des solutions qui ne sont pas dans E_α. Nous allons montrer que si α est tel que [17]
$$M\alpha < r_0 \text{ et } \alpha L_0 < 1$$
alors ce n'est pas possible, autrement dit que toute solution du problème de Cauchy est dans E_α : soit donc $y \in C^0(J_\alpha, U)$ solution de $\phi(y) = y$, et supposons par l'absurde que $y \notin E_\alpha$. Il existe donc $t_1 \in J_\alpha$ tel que $\|y(t_1) - y_0\| > r_0$. Pour fixer les idées on suppose que $t_1 \geq t_0$, l'autre cas se traitant de façon similaire. Alors on pose
$$t_2 = \inf\{t \in [t_0, t_0+\alpha],\ \|y(t) - y_0\| > r_0\}.$$

15. Attention, on ne peut pas dire qu'on munit E_α d'une norme, puisque E_α n'est pas un espace vectoriel.
16. En fait cette dernière phrase montre que l'ensemble E_α est fermé dans $C^0(J_\alpha, \mathbb{K}^d)$. Or ce dernier étant complet pour la métrique induite par la norme infinie, E_α est donc complet, comme on vient de le redémontrer ici dans ce cas particulier.
17. On a repris les hypothèses de l'étape précédente et on a juste renforcé un tout petit peu l'une des hypothèses en remplaçant une inégalité large par une inégalité stricte.

Cet infimum est bien défini car l'ensemble considéré est non vide (il contient t_1) et minoré par t_0. Par continuité de y en t_0, on déduit que $t_2 > t_0$, et par continuité en t_2 que $\|y(t_2) - y_0\| = r_0$. De plus, par le caractère minimal de t_2 on a

$$\forall t \in [t_0, t_2],\ \|y(t) - y_0\| \le r_0.$$

En reprenant le calcul (1.12), qui est légitime grâce à l'estimation précédente, on obtient

$$\forall t \in [t_0, t_2], \quad \|y(t) - y_0\| \le M|t - t_0| \le M\alpha,$$

et donc en particulier $\|y(t_2) - y_0\| \le M\alpha < r_0$, ce qui constitue une contradiction.

Pour conclure, il reste à remarquer qu'on a travaillé avec un intervalle fermé J_α, mais que pour avoir un intervalle ouvert, il suffit de se restreindre à $]t_0 - \alpha, t_0 + \alpha[$ qui est ouvert. □

1.3 Sur la propriété d'unicité ; lemme de Gronwall

Nous avons vu au théorème 1.10 une propriété d'unicité dite locale, c'est-à-dire que l'énoncé affirmait l'existence d'un intervalle J sur lequel il y a unicité de la solution, mais on souhaiterait avoir un résultat plus flexible :

Proposition 1.16 (Propriété d'unicité forte). *Sous les hypothèses du théorème 1.10 de Cauchy-Lipschitz, si $y_1 \in C^1(J_1, U)$ et $y_2 \in C^1(J_2, U)$ sont deux solutions de (1.3), et s'il existe $t_0 \in J_1 \cap J_2$ tel que $y_1(t_0) = y_2(t_0)$* [18]*, alors $y_1(t) = y_2(t)$ pour tout $t \in J_1 \cap J_2$.*

On va donner deux démonstrations, la première repose sur l'utilisation de l'unicité locale prouvée dans le théorème 1.10. La seconde repose sur un outil important en soit, à savoir le lemme de Gronwall, qui est le cas particulier de ce qu'on appellera un principe de comparaison et qui permet d'obtenir des estimations sur les solutions d'une EDO, ce qui est en pratique utile quand on ne peut pas résoudre cette dernière explicitement.

Première démonstration de la proposition 1.16 : pour simplifier la présentation, on suppose J_1 et J_2 ouverts[19]. On pose $J = J_1 \cap J_2$ et J^* l'ensemble des points de J sur lesquels y_1 et y_2 coïncident. Alors par continuité de y_1 et y_2, l'ensemble J^* est un sous-ensemble fermé de J. De plus, si $t_1 \in J^*$, alors y_1 et y_2 sont toutes les deux solutions locales du problème de Cauchy

$$\begin{cases} y' &= f(t, y) \\ y(t_1) &= y_1(t_1) \end{cases}$$

donc par l'unicité dans le théorème 1.10, il existe un voisinage de t_1 sur lequel y_1 et y_2 coïncident ; autrement dit, J^* est un ouvert de J.

Mais comme $y_1(t_0) = y_0 = y_2(t_0)$, J^* est non vide, donc par connexité de J, $J^* = J$, ce qui conclut la preuve. □

Lemme 1.17 (Lemme de Gronwall). *Soit $J = [t_0, B[\subset \mathbb{R}$, et $a : J \to \mathbb{R}$ continue.*

— Forme différentielle *: on suppose $u : J \to \mathbb{R}$ de classe C^1 et telle que*

$$\forall t \in J,\ u'(t) \le a(t)u(t). \tag{1.13}$$

Alors

$$\forall t \in J,\ u(t) \le u(t_0) \exp\left(\int_{t_0}^{t} a(s)ds\right).$$

18. Autrement dit, y_1 et y_2 sont deux solutions du même problème de Cauchy (1.8), mais possiblement sur des intervalles différents.

19. Dans certaines littérature, on suppose toujours les ensembles de définition ouverts, ce qui est restrictif mais simplifie parfois la présentation. Ici pour traiter le cas général, il faudrait distinguer plusieurs cas ; on n'a pas pris cette peine, d'autant qu'on fournit une autre démonstration sans hypothèse sur le caractère ouvert de J_1, J_2.

— Forme intégrale *: on suppose* a **positive**, $u : J \to \mathbb{R}$ *continue et* $K \in \mathbb{R}$ *tels que*

$$\forall t \in J,\ u(t) \leq K + \int_{t_0}^{t} a(s)u(s)ds. \tag{1.14}$$

Alors

$$\forall t \in J,\ u(t) \leq K \exp\left(\int_{t_0}^{t} a(s)ds\right).$$

Remarque 1.18. Attention, le lemme de Gronwall s'énonce pour des temps supérieurs à t_0. On pourrait donner des énoncés pour des temps inférieurs, mais on conseille plutôt de “changer le sens du temps” en considérant la fonction $t \mapsto u(-t)$ et en lui appliquant le lemme de Gronwall classique.

Exemple 1.19. Si $u(t_0) = 0$ dans la première version ou $K = 0$ dans la seconde, on obtient $u \leq 0$.

Remarque 1.20. Les deux versions du lemme ont leurs avantages et inconvénients : la première version demande u de classe C^1, la seconde requière la positivité de a. Remarquons aussi que si on suppose u de classe C^1 et satisfait (1.13), alors u satisfait (1.14) avec $K = u(t_0)$.

Le fait que la seconde ne requière que la continuité n'est pas anodin car il arrivera qu'on l'applique à $t \mapsto \|y(t)\|$ où y est solution d'une EDO.[20]

Démonstration. — Soit y de classe C^1 qui satisfait (1.13) : la fonction $\phi : t \in J \mapsto y(t) \exp\left(-\int_{t_0}^{t} a(s)ds\right)$ est C^1 et satisfait

$$\forall t \in J,\ \phi'(t) = \left[y'(t) - a(t)y(t)\right] \exp\left(-\int_{t_0}^{t} a(s)ds\right) \leq 0$$

donc ϕ est décroissante, d'où $\forall t \in J,\ \ \phi(t) \leq \phi(t_0) = y(t_0)$, ce qui donne le résultat.

— Soit y continue qui satisfait (1.14). On pose

$$\forall t \in J,\ \ z(t) = K + \int_{t_0}^{t} a(s)y(s)ds,$$

et l'hypothèse devient $y \leq z$ sur J. Alors z est de classe C^1 par le théorème 5.12 du Chapitre I, et

$$\forall t \in J,\ \ z'(t) = a(t)y(t) \leq a(t)z(t)$$

par (1.14) et positivité de a. Par la partie précédente, on en déduit

$$\forall t \in J,\ \ z(t) \leq z(t_0) \exp\left(\int_{t_0}^{t} a(s)ds\right),$$

d'où le résultat puisque $y \leq z$ et $z(t_0) = K$.

□

Deuxième preuve de la proposition 1.16 : on pose $J = J_1 \cap J_2$ et

$$\widehat{J} = \{t \in J,\ \forall s \in [t_0, t],\ \ y_1(s) = y_2(s)\}$$

et où $[t_0, t]$ est remplacé par $[t, t_0]$ si $t < t_0$. On souhaite montrer que $\widehat{J} = J$.

Par construction, $\widehat{J}$ est un intervalle de J et contient t_0. On pose $t_+ = \sup \widehat{J}$ et on suppose par l'absurde que $t_+ < \sup J$. Alors $y_2 = y_1$ sur $[t_0, t_+[$ donc sur $[t_0, t_+]$ par continuité, et de plus, $t_+ \in \mathring{J}$.

On introduit L une constante de lipschitz en la seconde variable pour f, valable pour $t \in]t_+ - \varepsilon, t_+ + \varepsilon[$ et $(y_1, y_2) \in \overline{B}(y_1(t_+), \varepsilon)$. Par continuité de y_1 et y_2, il existe $\alpha < \varepsilon$ tel que

$$\forall t \in [t_+, t_+ + \alpha[, \forall i \in \{1, 2\},\ \ y_i(t) \in B(y_1(t_+), \varepsilon).$$

20. Pour cette raison, il est, me semble-t-il, plus fréquent de trouver la seconde version dans la littérature ; je préfère personnellement la première qui me semble plus facile à retenir, notamment dans sa version généralisée. Mais en pratique, il n'est pas rare de redémontrer ce lemme à chaque fois (dans la version qui nous arrange), et pour cette raison on conseille de retenir l'astuce du début de sa démonstration (poser la bonne fonction ; le reste est naturel).

En écrivant les équations intégrales satisfaites par y_1 et y_2, on obtient

$$\forall t \in [t_+, t_+ + \alpha[, \quad \|y_2(t) - y_1(t)\| \leq \int_{t_+}^{t} \|f(s, y_1(s)) - f(s, y_2(s))\| \leq L \int_{t_+}^{t} \|y_2(s) - y_1(s)\| ds.$$

On peut appliquer le lemme de Gronwall dans sa version intégrale à la fonction $u : t \mapsto \|y_2(t) - y_1(t)\|$ avec la fonction $a \equiv L$ et $K = 0$, ce qui donne $u \leq 0$ sur $[t_+, t_+ + \alpha[$, c'est-à-dire $y_1(t) = y_2(t)$ pour tout $t \in [t_+, t_+ + \delta[$. Ceci contredit la maximalité de t_+, et donc finalement $t_+ = \sup J$.

De la même manière on montre que $\inf \widehat{J} = \inf J$, et donc $\widehat{J} = J$. □

Le résultat n'est pas à proprement parler au programme, mais il met un peu de lumière dans le lemme de Gronwall dont il est une généralisation.

Proposition 1.21 (Principes de comparaison)**.** *Soit $J = [t_0, B[$, et $f : [t_0, B[\times \mathbb{R} \to \mathbb{R}$ de classe C^1.*

— Forme différentielle *: on suppose $u, v : J \to \mathbb{R}$ de classe C^1 et telle que*

$$u(t_0) \leq v(t_0), \quad \text{et} \quad \forall t \in J,\ u'(t) \leq f(t, u(t)), \quad v'(t) \geq f(t, v(t)) \tag{1.15}$$

Alors

$$\forall t \in J,\ u(t) \leq v(t).$$

— Forme intégrale *: on suppose f croissante en la seconde variable, $u : J \to \mathbb{R}$ continue et $K \in \mathbb{R}$ tels que*

$$\forall t \in J,\ u(t) \leq K + \int_{t_0}^{t} f(s, u(s)) ds. \tag{1.16}$$

Alors $\forall t \in J,\ u(t) \leq v(t)$. où v est solution de

$$\left\{ \begin{array}{rcl} v' & = & f(t, v) \\ v(t_0) & = & K \end{array} \right. .$$

Exemple 1.22. Le lemme 1.17 correspond au cas où $f : (t, u) \in J \times \mathbb{R} \mapsto a(t)u$. Le lemme de Gronwall est donc un principe de comparaison linéaire.

Remarque 1.23. Les fonctions u et v qui satisfont (1.15) sont appelées respectivement sous-solution et sur-solution de l'équation $y' = f(t, y)$.

Exemple 1.24. Mise en pratique : si on se donne y une solution de $y' = f(t, y)$ que l'on ne sait pas calculer explicitement, on peut introduire une fonction $g \geq f$ plus "commode" que f. Alors $y' \leq g(t, y)$, et si on introduit z la solution de $z' = g(t, z)$ et $z(t_0) = y(t_0)$ [21], on déduit de la proposition 1.21 l'estimation $y \leq z$, qui s'avère utile notamment si on sait calculer explicitement z. Si on veut une estimation de y par en-dessous, on devra trouver $h \leq f$ qui soit commode.

Démonstration. — On pose

$$\forall t \in J, \quad a(t) := \left\{ \begin{array}{ll} \dfrac{f(t, v(t)) - f(t, u(t))}{v(t) - u(t)} & \text{si } u(t) \neq v(t) \\ \partial_2 f(t, u(t)) & \text{si } u(t) = v(t) \end{array} \right.$$

Admettons un instant que a est continue : alors $w = u - v$ satisfait

$$\forall t \in J, \quad w'(t) = u'(t) - v'(t) \leq f(t, u(t)) - f(t, v(t)) = a(t)w(t)$$

avec $w(t_0) \leq 0$, donc par le lemme de Gronwall, pour tout $t \in J$,

$$w(t) \leq w(t_0) \exp\left(\int_{t_0}^{t} a(s) ds \right) \leq 0,$$

d'où le résultat.

21. On pourrait faire d'autres choix pour la condition initiale, mais celui-ci donnera la meilleure estimation.

Montrons la continuité de a. Soit $t^* \in J$: dans le cas $u(t^*) \neq v(t^*)$, alors dans un voisinage de t^* on a encore $u \neq v$ et alors la définition de a est bien continue. On se place dans le cas plus difficile $u(t^*) = v(t^*)$. Alors pour t voisin de t^* tel que $u(t) \neq v(t)$, le théorème des accroissements finis montre l'existence de $c(t)$ compris entre $u(t)$ et $v(t)$ tel que

$$a(t) = \frac{f(t, v(t)) - f(t, u(t))}{v(t) - u(t)} = \partial_2 f(t, c(t))$$

et dans le cas où $u(t) = v(t)$ on a la même égalité en posant $c(t) = u(t)$. Par encadrement on a

$$c(t) \xrightarrow[t \to t^*]{} u(t^*)$$

et donc par continuité de $\partial_2 f$ on a

$$a(t) = \partial_2 f(t, c(t)) \xrightarrow[t \to t^*]{} \partial_2 f(t, u(t^*)) = a(t^*)$$

d'où la continuité de a en t^*.

— On pose $U(t) = K + \int_{t_0}^{t} f(s, u(s))ds$. L'hypothèse est donc $u \leq U$. Alors

$$\forall t \in J,\ U'(t) = f(t, u(t)) \leq f(t, U(t))$$

en utilisant la croissance de f en la seconde variable. Ainsi U est une sous-solution de $y' = f(t, y)$ et comme v est une solution (donc une sur-solution) et que $U(t_0) = K = v(t_0)$, la version différentielle du principe de comparaison donne $U \leq v$ sur J d'où le résultat puisque $u \leq U$.

□

1.4 A propos du temps d'existence

1.4.1 Solutions maximales

Le théorème 1.10 donne l'existence d'une solution locale au problème de Cauchy. Voici une version plus globale du théorème de Cauchy-Lipschitz

Corollaire 1.25 (Théorème de Cauchy-Lipschitz pour les solutions maximales)**.** *Sous les hypothèses du théorème 1.10 de Cauchy-Lipschitz,*

1. *il existe une unique solution maximale à l'équation* (1.8)*. De plus, son intervalle de définition est ouvert dans I.*
2. *les solutions de* (1.8) *sont les restrictions de la solution maximale à des sous-intervalles : si on note (J_{max}, y_{max}) l'unique solution maximale de* (1.8)*, alors (J, y) est solution de* (1.8) *si et seulement si J est un sous-intervalle de J_{max} qui contient t_0, et $y(t) = y_{max}(t)$ pour tout $t \in J$.*

Remarque 1.26. — Sous les hypothèses du théorème 1.10 de Cauchy-Lipschitz, on en déduit que deux solutions maximales de (1.3) sont soient identiques, soit ne s'intersectent pas : en effet si (J_1, y_1) et (J_2, y_2) sont deux solutions maximales de (1.3), alors s'il existe $t_0 \in J_1 \cap J_2$ tel que $y_1(t_0) = y_2(t_0)$ alors y_1 et y_2 sont deux solutions maximales du problème de Cauchy (1.8) avec condition initiale en t_0 et donc par la propriété d'unicité, $J_1 = J_2$ et $y_1 \equiv y_2$.

— si de plus $d = 1$ et $\mathbb{K} = \mathbb{R}$, on peut même en déduire que si deux solutions maximales ne sont pas identiques, alors l'une est strictement supérieure à l'autre : en effet si (J_1, y_1) et (J_2, y_2) sont deux solutions de (1.3) et si $t_0 \in J_1 \cap J_2$ est tel que $y_1(t_0) < y_2(t_0)$, alors nécessairement $y_1(t) < y_2(t)$ pour tout $t \in J_1 \cap J_2$. En effet si par l'absurde tel n'était pas le cas, alors par continuité on trouverait t_1 tel que $y_1(t_1) = y_2(t_1)$ et par le point précédent on aurait $J_1 = J_2$ et $y_1 \equiv y_2$, ce qui contredit l'hypothèse faite en t_0.

Démonstration. — Considérons l'ensemble S de toutes les solutions (J, y) du problème de Cauchy (1.8). Cet ensemble est non vide d'après le théorème 1.10 de Cauchy-Lipschitz local. On peut alors poser

$$J_{max} = \bigcup_{(J,y) \in S} J,$$

et la fonction $y_{max} : J_{max} \to U$ par

$$\forall t \in J_{max}, \;\; y_{max}(t) = y(t) \text{ pour tout } (J, y) \in S \text{ telle que } t \in J.$$

Cette fonction est bien définie, d'une part car si $t \in J_{max}$ alors par définition il existe $(J, y) \in S$ tel que $t \in J$, et d'autre part parce que si $(\widehat{J, \widehat{y}}) \in S$ est une autre solution telle que $t \in J$ alors par la propriété d'unicité 1.16 on a $y(t) = \widehat{y}(t)$, il n'y a donc pas d'ambiguité sur la valeur de $y_{max}(t)$. De plus y_{max} est encore une solution de (1.8).

— Par définition, les solutions de (1.8) sont bien les restrictions de (J_{max}, y_{max}) à des sous-intervalles contenant t_0. Cela implique aussi l'unicité de (J_{max}, y_{max}).

— Il reste à conclure que J_{max} est ouvert : si par l'absurde ce n'est pas le cas, alors J_{max} contient une de ses bornes : supposons par exemple $\beta := \sup J_{max} \in J_{max}$, le cas $\inf J_{max} \in J_{max}$ étant similaire. Mais alors par définition des solutions, $(\beta, y(\beta)) \in I \times U$ et on peut appliquer le théorème 1.10 de Cauchy-Lipschitz local et en déduire qu'il existe une solution locale au problème

$$\begin{cases} y' &= f(t, y) \\ y(B) &= y_{max}(\beta) \end{cases}$$

qu'on note y_β et qui est définie sur un intervalle $]\beta - \alpha, \beta + \alpha[$. On peut alors définir

$$y(t) = \begin{cases} y_{max}(t) & \text{si } t \in J_{max} \\ y_\beta(t) & \text{si } t \in [\beta, \beta + \alpha[\end{cases}$$

qui est encore une solution de (1.8) définie sur un intervalle strictement plus grand que J_{max}, ce qui contredit la maximalité de J_{max}.

□

Dans les énoncés suivants, nous allons voir que le phénomène observé à la remarque 1.5 est général, c'est-à-dire que pour que dans le cadre des hypothèses du théorème de Cauchy-Lipschitz, pour qu'une solution ne puisse être prolongée en une solution sur un intervalle plus grand (c'est-à-dire pour qu'elle soit maximale), il faut en fait qu'elle "explose" : commençons par un cas particulier faible, plus facile à comprendre et à démontrer et qui suffit dans de nombreux cas.

Théorème 1.27 (Alternative d'explosion, version faible). *On se place sous les hypothèses du théorème 1.10 de Cauchy-Lipschitz, et on suppose que $U = \mathbb{K}^d$, c'est-à-dire que f est définie sur $I \times \mathbb{K}^d$. Si $y \in C^1(J, U)$ est une solution maximale de (1.3), on note b la borne maximale de I et β la borne maximale de J. Alors deux situations sont possibles :*

1. *ou bien $\beta = b$,*[22]
2. *ou bien $\beta < b$ et*

$$\limsup_{t \to \beta^-} \|y(t)\| = +\infty.$$

De la même manière, si $\inf J > \inf I$, alors

$$\limsup_{t \to (\inf J)^+} \|y(t)\| = +\infty.$$

Démonstration. On se place dans le cas $\beta < b$ et on suppose par l'absurde que $\limsup_{t \to \beta^-} \|y(t)\| < +\infty$. Ainsi y est bornée au voisinage de β^-. Comme on a la relation $y' = f(t, y)$ et qu'en tant que fonction continue la fonction f est bornée sur les compacts de $I \times \mathbb{K}^d$ on en déduit que y' est bornée sur un voisinage de β^-. Par le théorème des accroissements finis et le critère de Cauchy (voir l'exercice 1.90 au Chapitre III), cela implique que y admet une limite réelle quand t tend vers β^-, qu'on note y_β.

Le couple $(\beta, y_\beta) \in I \times \mathbb{K}^d$ est une donnée admissible pour définir un problème de Cauchy, et d'après le théorème de Cauchy-Lipschitz local, il existe une solution z de $y' = f(t, y)$ définie sur un voisinage de β et telle que $z(\beta) = y_\beta$. Mais alors on peut comme à la fin de la preuve du corollaire 1.25 on peut recoller la solution y et la solution z pour obtenir une solution sur un intervalle strictement plus grand que J, ce qui contredit la maximalité de J. □

22. C'est-à-dire que la solution est "maximale à droite", c'est-à-dire pour les temps qui vont dans le direction de la borne supérieure de I.

On va maintenant énoncer un cas plus général où f est définie sur un ouvert de $\mathbb{K}^d$, et également plus fort que le théorème 1.27, voir l'exemple 1.30.

Théorème 1.28 (Théorème des bouts/sortie de tout compact)**.** *Sous les hypothèses du théorème 1.10 de Cauchy-Lipschitz, si $y \in C^1(J, U)$ est une solution maximale de (1.3), on note b la borne maximale de I et β la borne maximale de J. Alors deux situations sont possibles :*

1. *ou bien $\beta = b$,*
2. *ou bien $\beta < b$ et y sort de tout compact de U, c'est-à-dire que pour tout $K \subset U$ compact, il existe $\varepsilon > 0$ tel que*
$$y(t) \in U \setminus K, \quad \forall t \in [\beta - \varepsilon, \beta[.$$

De la même manière, si $\inf J > \inf I$, alors y sort de tout compact lorsque $t \to (\inf J)^+$.

Exemple 1.29. On a déjà vu un phénomène d'explosion à la remarque 1.5, voyons ici à quoi peut ressembler une "sortie de tout compact" dans le cas où $U \neq \mathbb{K}^d$. Considérons l'équation

$$y' = -\frac{1}{2y}.$$

Ici on a $I = \mathbb{R}$, $d = 1$ et $U = \mathbb{R}^*$ avec $f : (t, y) \in \mathbb{R} \times \mathbb{R}^* \mapsto 1/y$. On peut calculer que les solutions de cette équation sont de la forme

$$t \in]-\infty, \alpha[\mapsto \sqrt{\alpha - t}, \qquad t \in]-\infty, \alpha[\mapsto -\sqrt{\alpha - t}.$$

Ces solutions sont maximales mais non globales : en fait, lorsque t s'approche du point limite d'existence α, on a $y(t)$ qui se rapproche de 0 qui n'est plus dans l'ensemble de définition de f. La propriété de sortir de tout compact de $\mathbb{R}^*$ consiste ici à se rapprocher de 0.

Exemple 1.30. Le théorème 1.28 est plus fort que le théorème 1.27 même dans le cas $U = \mathbb{K}^d$. En effet il permet d'affirmer que si $\beta < b$ alors on a

$$\lim_{t \to \beta^-} \|y(t)\| = +\infty$$

c'est-à-dire qu'on remplace une limite supérieure par une limite ; on peut appeler ce résultat l'alternative d'explosion, version forte.

Démonstration. Plaçons dans le cas $\beta < b$, et supposons par l'absurde qu'il existe K compact de U et une suite $(t_n)_{n\in\mathbb{N}}$ telle que

$$t_n \xrightarrow[n \to +\infty]{} \beta, \quad \text{ et } \quad \forall n \in \mathbb{N},\ y(t_n) \in K.$$

Par compacité de K, il existe φ une extraction et $y_\beta \in K$ tels que $y(t_{\varphi(n)}) \xrightarrow[n \to +\infty]{} y_\infty$ [23]. Pour simplifier les notations, nous notons $(y_n)_{n\in\mathbb{N}}$ à la place de $(y_{\varphi(n)})_{n\in\mathbb{N}}$.

Nous allons montrer qu'on a en fait la propriété plus forte que $y(t) \xrightarrow[t \to \beta]{} y_\beta$. Si tel est le cas, alors $(\beta, y_\beta) \in I \times K$ (car K est fermé et $\beta < b$), et donc est une condition de Cauchy admissible, donc par le théorème 1.10 il existe z solution de $z' = f(t, z)$ définie sur un voisinage de β et telle que $z(\beta) = y_\beta$. En recollant y et z on construit une solution sur un intervalle plus grand que J, ce qui contredit la maximalité de J.

Il reste donc à montrer $y(t) \xrightarrow[t \to \beta]{} y_\beta$:

— Comme $(\beta, y_\beta) \in I \times U$ et $I \times U$ est ouvert, il existe r_0 tel que $\overline{B}_{r_0} := [\beta - r_0, \beta + r_0] \times \overline{B}(y_\beta, r_0) \subset I \times U$; par compacité de $\overline{B}_{r_0}$ et continuité de f, on peut considérer $M = \sup_{\overline{B}_{r_0}} \|f\| \in \mathbb{R}$.

— Comme $t_n \xrightarrow[n \to +\infty]{} \beta$ et $y(t_n) \xrightarrow[n \to +\infty]{} y_\infty$, il existe $n_0 \in \mathbb{N}$ tel que $t_{n_0} \in [\beta - r_0, \beta[$ et $\|y(t_{n_0}) - y_\beta\| \leq \frac{r_0}{2}$. Enfin on peut également supposer $M(t_{n_0} - \beta) < r_0/2$ puisque $M(t_n - \beta) \xrightarrow[n \to +\infty]{} 0$.

23. Autrement dit y_β est un point d'accumulation de $(y(t))_{t\in[\beta-\varepsilon,\beta[}$ pour $\varepsilon > 0$

— Montrons que pour tout $t \in [t_{n_0}, \beta[$, $\|y(t) - \beta\| \leq r_0$. Par l'absurde, si tel n'est pas le cas, on pose
$$\tau = \inf\{t \in [t_{n_0}, \beta[,\ \|y(t) - \beta\| > r_0\}$$
qui est bien défini et tel que $\tau > t_{n_0}$ par continuité de y en t_{n_0}. On a alors par définition
$$\forall t \in [t_{n_0}, \tau],\ \|y(t) - \beta\| \leq r_0, \quad \|y(\tau) - \beta\| = r_0.$$
Ainsi par l'équation $y' = f(t, y)$ et le fait que $(t, y(t)) \in \overline{B}_{r_0}$ pour tout $t \in [t_{n_0}, \tau]$, on a
$$\|y(t_{n_0}) - y(\tau)\| \leq \int_{t_{n_0}}^{\tau} \|y'(t)\| dt \leq M(\tau - t_{n_0}) < \frac{r_0}{2}.$$
Ainsi,
$$\|y(\tau) - y_\beta\| \leq \|y(\tau) - y(t_{n_0})\| + \|y(t_{n_0}) - y_\beta\| < \frac{r_0}{2} + \frac{r_0}{2} = r_0$$
ce qui fournit une contradiction.

— Ainsi pour tout $t \in [t_{n_0}, \beta[$, $(t, y(t)) \in \overline{B}_{r_0}$ donc $y' = f(t, y)$ est bornée par M et donc par l'exercice 1.90 au Chapitre III [24], y admet une limite réelle en β, qui ne peut être que y_β, ce qui conclut la démonstration.

□

Remarque 1.31. On peut présenter une approche assez différente de la preuve du théorème 1.28, qui repose sur le résultat suivant qui est intéressant en soit : étant donné un compact $K \subset I \times U$, on peut trouver $\tau > 0$ tel que pour tout $(t_0, y_0) \in K$, il existe une solution du problème de Cauchy (1.8) associé sur $[t_0 - \tau, t_0 + \tau]$. La nouveauté étant que le temps d'existence de la solution ne dépend que de K et non de (t_0, y_0). Je n'ai pas trouvé ce résultat énoncé ainsi dans la littérature, mais on peut trouver une version faible dans [BG10, Lemme 5.8].

À partir de ce résultat (ou de variante), on peut donner une autre démonstration du théorème 1.28, voir [BG10, Théorème 5.7], [QZ13, Théorème III.3].

1.4.2 Solutions globales

Comme on l'a évoqué précédemment, une des questions naturelles est de savoir si une solution maximale et globale ou non, c'est-à-dire si la [25] solution maximale du problème de Cauchy (1.8) sera définie sur l'intervalle I tout entier ou non.

Commençons par un exemple simple qui est un corollaire immédiat du théorème 1.27 :

Exemple 1.32. Si $f : I \times \mathbb{K}^d \to \mathbb{K}^d$ est bornée, alors les solutions maximales de $y' = f(t, y)$ sont globales : en effet si (J, y) est une solution maximale, on choisit t_0 quelconque dans J et on a
$$\forall t \in J, t \geq t_0, \quad \|y(t)\| \leq \|y(t_0)\| + \|f\|_\infty (t - t_0).$$
En conséquence, la fonction y ne peut pas exploser en temps fini, donc par le théorème 1.27 on a $\sup J = \sup I$. De même $\inf J = \inf I$ d'où la globalité de la solution.

Théorème 1.33 (Condition d'existence global 1 ; cas globalement lipschitzien)**.** *Si $f : I \times \mathbb{K}^d \to \mathbb{K}^d$ est continue et globalement lipschitzienne par rapport à la seconde variable, c'est-à-dire qu'il existe $k \in \mathbb{R}$ tel que*
$$\forall t \in I,\ \forall (y, \hat{y}) \in (\mathbb{K}^d)^2, \quad \|f(t, y) - f(t, \hat{y})\| \leq k\|y - \hat{y}\| \tag{1.17}$$
alors pour tout $(t_0, y_0) \in I \times \mathbb{K}^d$, il existe une unique solution globale à (1.8).

24. Voici une autre façon de mener cet argument : comme y' est bornée, elle est intégrable sur $[t_{n_0}, \beta[$ et donc l'intégrale impropre $\int_{t_{n_0}}^{\beta} y'(t)dt$ est convergente, donc y admet une limite en β.

25. La plupart du temps, on parle d'une équation différentielle en sous-entendant qu'on satisfait les hypothèses du théorème de Cauchy-Lipschitz (ces dernières étant relativement faibles), et donc on s'autorise à parler de “la” solution, cette dernière existant et étant unique.

Remarque 1.34. On peut en fait enlever le caractère uniforme en t, en remplaçant (1.17) par

$$\forall t \in I,\ \forall (y,\hat{y}) \in U^2,\ \ \|f(t,y)-f(t,\hat{y})\| \leq k(t)\|y-\hat{y}\| \tag{1.18}$$

où $k : I \to \mathbb{R}$ est continue et positive. Vérifiez que les preuves ci-dessous fonctionnent encore !

Démonstration. **Première démonstration : avec le corollaire 1.25 et le théorème 1.27 :** le corollaire 1.25 donne l'existence d'une unique solution maximale (J,y). Il reste à montrer que celle-ci est globale. On choisit $t_0 \in J$ quelconque ; alors d'après la version intégrale de l'équation différentielle, on a

$$\forall t \in J, t \geq t_0,\ \|y(t)\| \leq \|y(t_0)\| + \int_{t_0}^t \|f(s,y(s))\|ds \leq \|y(t_0\| + \int_{t_0}^t \Big[\|f(s,0)\| + k\|y(s)\|\Big]ds.$$

Supposons que $\beta = \sup J < \sup I$. Alors on obtient la majoration

$$\forall t \in [t_0,\beta[,\ \|y(t)\| \leq \left(\|y(t_0)\| + \int_{t_0}^\beta \|f(s,0)\|ds\right) + k\int_{t_0}^t \|y(s)\|ds.$$

Par le lemme de Gronwall-version intégrale appliquée à $u = \|y\|$, on obtient

$$\forall t \in [t_0,\beta[,\ \|y(t)\| \leq \left(\|y(t_0)\| + \int_{t_0}^\beta \|f(s,0)\|ds\right)\exp(k(t-t_0)),$$

donc la fonction y est bornée, ce qui contredit le théorème 1.27. Ainsi $\sup J = \sup I$ et de même pour la borne inférieure.

Deuxième démonstration : preuve directe, en reprenant les idées de la preuve du théorème 1.10 : remarquons en effet qu'on peut démontrer directement l'existence et l'unicité d'une solution globale : pour cela on reprend la démonstration du théorème 1.10 mais grâce au caractère globalement lipschitz de f, nous n'aurons pas besoin de restreindre le temps d'existence. Comme cette preuve est bien référencée dans [Rou03, Exercice 60] [26], on donne seulement le schéma de preuve :

— On définit

$$\begin{array}{rccc} \phi : & C^0(I,\mathbb{K}^d) & \longrightarrow & C^0(I,\mathbb{K}^d) \\ & y & \longmapsto & \left[t \mapsto y_0 + \displaystyle\int_{t_0}^t f(s,y(s))ds\right]. \end{array}$$

et on cherche un point fixe de ϕ.

— On ne peut pas travailler avec I intervalle ouvert directement, on suppose donc pour le moment que I est compact. Dans ce cas l'espace $(C^0(I,\mathbb{K}^d), \|\cdot\|_\infty)$ est un espace de Banach. De plus par la propriété de Lipschitz de ϕ, on montre

$$\forall (y_1,y_2) \in C^0(I,\mathbb{K}^d)^2,\ \ \|\phi(y_2)-\phi(y_1)\|_\infty \leq k|I|\|y_2-y_1\|_\infty$$

où $|I|$ désigne la longueur de I, qui est finie puisqu'on suppose I compact.

— Cette étape est différente de la preuve du théorème 1.10 : en effet, au lieu de jouer sur la longueur de l'intervalle d'existence (on veut qu'il y ait existence pour tout temps) on va itérer ϕ et constater que cela fait diminuer sa constante de Lipschitz, au point qu'une itération de ϕ soit contractante : en effet par récurrence on montre [27]

$$\forall (y_1,y_2) \in C^0(I,\mathbb{K}^d)^2, \forall p \in \mathbb{N}\ \ \|\phi^p(y_2)-\phi^p(y_1)\|_\infty \leq \frac{(k|I|)^p}{p!}\|y_2-y_1\|_\infty$$

26. Notez qu'on trouve une autre preuve (aussi basée sur le théorème de point fixe) dans [Boyb, 2.6].
27. Attention, il ne suffit pas d'itérer $\|\phi(y_2)-\phi(y_1)\|_\infty \leq k|I|\|y_2-y_1\|_\infty$, il faut itérer la version précisée :

$$\forall (y_1,y_2) \in C^0(I,\mathbb{K}^d)^2, \forall t \in I,\ \ \|\phi(y_2)(t)-\phi(y_1)(t)\| \leq k|t-t_0|\|y_2-y_1\|_\infty.$$

— Pour $p \in \mathbb{N}^*$ assez grand, on a $\frac{(k|I|)^p}{p!} < 1$, donc par le théorème de point fixe, ϕ^p admet un unique point fixe, et par suite ϕ également (voir l'exercice 1.85 au Chapitre III), ce qui conclut la preuve dans le cas où I est compact. [28]

— Revenons au cas où I est quelconque : alors on peut écrire $I = \cup_{i\in\mathbb{N}} I_i$ où $(I_i)_{i\in\mathbb{N}}$ est une suite exhaustive de compacts. Par l'étape précédente, on peut noter y_i la solution du problème de Cauchy sur I_i, pour $i \in \mathbb{N}$. Du fait de la propriété d'unicité sur I_i, on a pour tout $j \geq i$, $(y_j)_{|I_i} = y_i$, donc on peut définir y ainsi : pour tout $t \in I$, il existe $i \in \mathbb{N}$ tel que $t \in I_i$, et alors on pose $y(t) = y_i(t)$, ce qui ne dépend pas de la valeur de i. La fonction ainsi construite est donc solution du problème de Cauchy sur I.

Quant à l'unicité, de la même manière si z est solution sur I, elle est solution sur chaque I_i pour $i \in \mathbb{N}$, donc par unicité sur I_i elle doit coïncider avec y_i sur I_i, et ce pour tout $i \in \mathbb{N}$, et donc par suite elle coïncide avec y sur I.

□

Remarque 1.35. Remarquons qu'on peut déduire du théorème 1.33 le théorème 1.10 de Cauchy-Lipschitz local (évidemment il n'y a donc de sens que d'utiliser la seconde preuve du théorème 1.33 si on veut utiliser cette stratégie) : en effet si $f : I\times U \to \mathbb{K}^d$ est continue et localement lipschitzienne en la seconde variable et $(t_0, y_0) \in I \times U$, on choisit $r_0 > 0$ tel que $\overline{B}_{2r_0} = [t_0 - 2r_0, t_0 + 2r_0] \times \overline{B}(y_0, 2r_0) \subset U$ et f est lipschitzienne en la seconde variable sur $\overline{B}_{2r_0}$. On introduit ensuite $\rho : \mathbb{K}^d \to \mathbb{R}$ de classe C^∞ et telle que $\rho = 1$ sur $\overline{B}_{r_0}$, $0 \leq \rho \leq 1$, et $\mathrm{Supp}(\rho) \subset \overline{B}_{2r_0}$ [29]. La fonction $\widetilde{f} := f\rho$ est alors globalement lipschitzienne et on peut lui appliquer le théorème 1.33, qui donne l'existence d'une solution globale (I, y) au problème de Cauchy pour $\widetilde{f}$; par continuité de y, il existe $\alpha > 0$ tel que $(t, y(t)) \in B_{r_0}$ pour tout $t \in]t_0 - \alpha, t_0 + \alpha[$, et comme $f = \widetilde{f}$ sur B_{r_0}, y est donc solution au problème de Cauchy pour f sur cet intervalle.

Théorème 1.36 (Condition d'existence global 2 ; cas sous-linéaire)**.** *Si f est continue, localement lipschitzienne par rapport à la seconde variable, et sous-linéaire en la seconde variable uniformément en la première variable, c'est-à-dire qu'il existe $(A, B) \in \mathbb{R}^2$ tels que*

$$\forall (t, y) \in I \times U, \quad \|f(t, y)\| \leq A\|y\| + B \tag{1.19}$$

alors toutes les solutions maximales de (1.3) sont globales.

Démonstration. On suit la même stratégie que la première preuve du théorème 1.33 : si (J, y) est une solution maximale et $t_0 \in J$ (quelconque), on a

$$\forall t \in J, t \geq t_0,\ \|y(t)\| \leq \|y(t_0)\| + \int_{t_0}^{t} \|f(s, y(s))\| ds \leq \|y(t_0\| + \int_{t_0}^{t} \Big[A\|y(s)\| + B\Big] ds.$$

Si $\beta := \sup J < \sup I$ alors on en déduit

$$\forall t \in J, t \geq t_0,\ \|y(t)\| \leq \Big[\|y(t_0)\| + B(\beta - t_0)\Big] + A \int_{t_0}^{t} \|y(s)\| ds,$$

donc par le lemme de Gronwall, y est bornée sur $[t_0, \beta[$, ce qui contredit le théorème 1.27. □

Remarque 1.37. Comme à la remarque 1.34, on peut en fait enlever le caractère uniforme en t, en remplaçant (1.19) par

$$\forall (t, y) \in I \times U, \quad \|f(t, y)\| \leq A(t)\|y\| + B(t)$$

où $A, B : I \to \mathbb{R}$ sont continues. D'ailleurs, dans cette version, le théorème 1.36 est plus fort que le théorème 1.33 (même que la version plus générale de la remarque 1.34) car l'hypothèse (1.18) implique

$$\forall (t, y) \in I \times U, \quad \|f(t, y)\| \leq k(t)\|y\| + \|f(t, 0)\|.$$

D'ailleurs, si on suit la démonstration (la première) du théorème 1.33 on réalise que c'est cette conséquence de (1.18) qu'on a utilisée.

28. Comme remarqué dans [Rou03, page 174], on peut en fait démontrer directement l'existence et l'unicité d'un point fixe de ϕ, par une méthode qui est une variante de la preuve du théorème de point fixe.

29. Une telle fonction existe ; c'est la construction de fonctions plateaux.

2 Méthodes de résolution explicite

Commençons par une remarque générale de méthodologie : nous allons présenter ici des méthodes calculatoires pour trouver la solution du problème de Cauchy (1.8). Ces méthodes ne sont pas toujours faciles à justifier rigoureusement : il peut arriver par exemple de diviser par une expression qui dépend de la solution y pour mener les calculs, sans que l'on sache que cette expression ne s'annule pas. On peut généralement rendre ces calculs rigoureux avec un peu d'attention, mais remarquons également l'"astuce" suivante, rarement bien comprise et pourant parfaitement rigoureuse : si le calcul amène à une fonction explicite $t \in J \mapsto y(t)$, alors plutôt que de justifier les calculs qui ont mené à cette formule, on peut tout à fait vérifier à la main que cette fonction est effectivement solution de (1.8). Auquel cas, si bien sûr on est dans le cadre d'application du théorème 1.10 de Cauchy-Lipschitz, l'unicité de la solution permet de conclure.

Remarquons aussi qu'on peut préférer trouver l'ensemble des solutions de (1.3) plutôt que de résoudre un problème de Cauchy ; dans ce cas, la remarque précédente s'applique, si on prend la peine de vérifier que l'ensemble des solutions qu'on a obtenues décrivent tous les problèmes de Cauchy (1.8) possibles, c'est-à-dire que les solutions obtenues passent par toutes les valeurs possibles de (t_0, y_0). Un dessin des solutions obtenues aidera grandement à se convaincre.

Exemple 2.1. À titre d'exemple, considérons l'équation

$$y' = -y^2.$$

On peut vérifier à la main que la fonction $t \in]-1, +\infty[\mapsto \frac{1}{t+1}$ est solution du problème de Cauchy associé à la condition $y(0) = 1$. Elle est de plus maximale, puisqu'elle ne peut pas être prolongée en -1. Le théorème de Cauchy-Lipschitz s'applique ici car $(t, y) \mapsto -y^2$ est une fonction de classe C^1, et donc cette solution est l'unique solution du problème de Cauchy

$$\begin{cases} y' &= -y^2 \\ y(0) &= 1 \end{cases}$$

De la même manière, les fonctions $t \in]-\alpha, +\infty[\mapsto \frac{1}{t+\alpha}$ pour $\alpha > 0$ sont solutions et décrivent les conditions de Cauchy $y(0) = \frac{1}{\alpha}$ et donc décrivent toutes les conditions telles que $y(0) > 0$ (car la fonction $\alpha > 0 \mapsto \frac{1}{\alpha}$ a pour image $]0, +\infty[$). Les fonctions $t \in]-\infty, \beta[\mapsto \frac{1}{t-\beta}$ pour $\beta > 0$ décrivent quant à elles les solutions associées aux conditions de Cauchy telles que $y(0) < 0$. Il reste à ajouter la solution $t \in \mathbb{R} \mapsto 0$ correspondant à la condition $y(0) = 0$, et on a ainsi décrit toutes les solutions de $y' = -y^2$ qui sont définies en 0.

En reprenant les formules

$$(\mathbb{R}, 0), \quad t \in]-\alpha, +\infty[\mapsto \frac{1}{t+\alpha} \text{ pour } \alpha \in \mathbb{R}, \quad t \in]-\infty, \beta[\mapsto \frac{1}{t-\beta} \text{ pour } \beta \in \mathbb{R}$$

on obtient toutes les solutions de $y' = -y^2$.

2.1 Équations scalaires à variables séparées

On se place dans le cas $d = 1$, et on fait l'hypothèse que $f : I \times U \to \mathbb{R}$ est de la forme $(t, y) \mapsto g(t)h(y)$ où $g : I \to \mathbb{R}$ et $h : U \to \mathbb{R}$ sont des fonctions continues. L'équation s'écrit alors

$$y' = g(t)h(y). \tag{2.1}$$

L'idée générale est d'écrire

$$\frac{y'}{h(y)} = g(t)$$

et d'intégrer. Mais il faut être un peu plus vigilant si on veut pouvoir justifier la division : il y a deux stratégies pour justifier proprement :

— on se place dans le cadre des hypothèses de Cauchy-Lipschitz, et on utilise que les solutions de (2.1) ne peuvent pas se croiser ;

— on prend une solution (J, y), on suppose que $h(y)$ ne s'annule pas en un $t_0 \in J$, et on effectue le calcul sur $\widehat{J}$ la composante connexe de $\{t \in J, h(y(t)) \neq 0\}$ qui contient t_0 ; après calcul, on constate (ou non) que la formule obtenue est telle que $h(y)$ ne peut s'annuler, et donc que $\widehat{J} = J$.

Précisons la situation en prenant la stratégie qui consiste à utiliser le théorème de Cauchy-Lipschitz :

1. On note $(z_i)_i$ l'ensemble des solutions dans U de l'équation $h(z) = 0$ [30]. Alors les fonctions constantes $y \equiv z_i$ sont solutions de (2.1). Une conséquence importante est que si $y_0 \in]z_i, z_{i+1}[$, alors la solution y du problème de Cauchy doit rester à valeurs dans $]z_i, z_{i+1}[$. En effet si par l'absurde elle n'y restait pas, par continuité il existerait t_1 tel que $y(t_1) = z_i$ ou z_{i+1}, mais alors par unicité dans le théorème de Cauchy-Lipschitz, on aurait $y \equiv z_i$ ou z_{i+1}, ce qui est une contradiction.
2. On suppose $y_0 \in]z_i, z_{i+1}[$. Alors en notant y la solution du problème de Cauchy associé, $h(y)$ ne s'annule pas et on peut écrire après intégration :

$$\int_{t_0}^{t} \frac{y'(s)}{h(y(s))} = \int_{t_0}^{t} g(s)ds$$

   On introduit G une primitive de g, et $\widetilde{H}$ une primitive de la fonction $1/h$. Alors l'égalité précédente se réécrit après changement de variable :

$$\widetilde{H}(y(t)) = G(t) + \lambda$$

   où λ est une constante [31]. Comme $\widetilde{H}' = 1/h$ est de signe constant sur $]y_i, y_{i+1}[$, la fonction $\widetilde{H}$ est une bijection sur son image, et on peut finalement écrire :

$$(y(t)) = \widetilde{H}^{-1}\Big[G(t) + \lambda\Big].$$

Exemple 2.2. Soit $a : I \to \mathbb{R}$ continue. Alors l'ensemble des solutions maximales de

$$y' = ay$$

est

$$\forall t \in I,\ y(t) = \lambda \exp\left(\int_{t_0}^{t} a(s)ds\right)$$

où $(\lambda, t_0) \in \mathbb{K} \times I$. Remarquons que l'ensemble des solutions (maximales) est un $(\mathbb{K})-$espace vectoriel de dimension 1, engendré par $t \in I \mapsto \exp\left(\int_{t_0}^{t} a(s)ds\right)$.

L'unique solution maximale du problème de Cauchy en $(t_0, y_0) \in I \times \mathbb{K}$ est

$$\forall t \in I,\ y(t) = y_0 \exp\left(\int_{t_0}^{t} a(s)ds\right).$$

Ceci est considéré comme un résultat de cours, et on n'est pas obligé de reproduire la stratégie précédente. Remarquons que les solutions sont en fait globales (ce qu'on aurait pu savoir avec la remarque 1.37, même s'il est beaucoup plus simple de le constater par le calcul).

2.2 Équations scalaires d'ordre 1 ; méthode de variation de la constante

On se donne $a, b : I \to \mathbb{R}$ deux fonctions continues, et on souhaite résoudre

$$y' = ay + b, \tag{2.2}$$

qui est une généralisation de l'exemple 2.2 $y' = ay$ qui est appelée équation homogène associée à (2.2) et correspond au cas $b = 0$.

30. On suppose que cet ensemble est discret.
31. Parfois on n'a pas besoin de garder trace de (t_0, y_0), parfois c'est mieux ; ici $\lambda = -G(t_0) + \widetilde{H}(y_0)$.

Proposition 2.3. *1. Les solutions maximales de* (2.2) *sont*

$$\forall t \in I,\ y(t) = \lambda \exp\left(\int_{t_0}^{t} a(s)ds\right) + \int_{t_0}^{t} b(s) \exp\left(\int_{s}^{t} a(u)du\right) ds, \tag{2.3}$$
$$= \lambda e^{A(t)} + e^{A(t)} \int_{t_0}^{t} b(s)e^{-A(s)}ds$$

où $\lambda \in \mathbb{K}$, $t_0 \in I$ et $A : t \in I \mapsto \int_{t_0}^{t} a(s)ds$.

2. Le problème de Cauchy en $(t_0, y_0) \in I \times \mathbb{K}$ associé à (2.2) *admet une unique solution maximale, donné par la formule* (2.3) *avec $\lambda = y_0$.*

Remarque 2.4. La formule (2.3) s'appelle la formule de Duhamel ; on peut bien sûr l'apprendre par cœur, mais en général on redémontre le résultat précédent avec ce qu'on appelle la méthode de variation de la constante (voir la preuve ci-dessous).

Retenons les deux conséquences suivantes :

— Si y_p est une solution particulière de (2.2), alors l'ensemble des solutions de (2.2) s'écrit

$$y = y_p + \lambda e^{A(t)}$$

où $\lambda \in \mathbb{K}$ et où on reconnait comme deuxième morceau les solutions de l'exemple 2.2. On dit que la solution générale de (2.2) est la somme d'une solution particulière, et des solutions de l'équation homogène associée $y' = ay$. Cette remarque est importante à retenir, et on la démontre ci-dessous.

— L'ensemble des solutions de (2.2) forme un espace affine de dimension 1 et dirigé par l'ensemble des solutions de l'équation homogène.

Remarque 2.5. À nouveau on remarque que les solutions sont globales.

Démonstration. Rappelons que si on admet le théorème de Cauchy-Lipschitz, alors une stratégie pourrait consister à simplement constater que les formules écrites sont bien solutions. Néanmoins, nous allons présenter une preuve (une méthode) qui permet de se passer du théorème de Cauchy-Lipschitz, ainsi que de retenir la formule (2.3). On se base par contre sur l'exemple 2.2.

On se donne $t_0 \in I$.

1. Rappelons que les solutions de l'équation homogène associée s'écrit $t \in I \mapsto \lambda e^{A(t)}$ où $A : t \in I \mapsto \int_{t_0}^{t} a(s)ds$.
2. Recherche d'une solution particulière (méthode de variation de la constante) : nous allons chercher une solution de (2.2) de la forme $y_p : t \in I \mapsto \mu(t)e^{A(t)}$ où $\mu : I \to \mathbb{R}$ est une fonction de classe C^1. On constate en effet que

$$y_p' = ay_p + b \quad \Leftrightarrow \quad \mu' e^A + \mu A' e^A = \mu a e^A + b$$
$$\Leftrightarrow \quad \mu' = be^{-A}$$

et donc en posant $\mu : t \in I \mapsto \int_{t_0}^{t} b(s)e^{-A(s)}ds$ on obtient que $y_p = \mu e^A$ est solution de (2.2).
3. Identification de l'ensemble des solutions. On écrit maintenant que

$$y \text{ est solution de (2.2)} \ \Leftrightarrow\ z = (y - y_p) \text{ est solution de } z' = az \ \Leftrightarrow\ y = y_p + \lambda e^A$$

où $\lambda \in \mathbb{K}$. Ceci donne bien la formule (2.3).
4. En prenant la formule précédente et ajoutant la condition $y(t_0) = y_0$ on obtient $\lambda = y_0$.

□

Remarque 2.6. Il peut arriver qu'on reconnaisse très facilement une solution particulière de (2.2) ; dans ce cas, on peut se passer de la méthode de la variation de la constante et se contenter de la fin de la méthode décrite ci-dessus.

Exercice 2.7. On suppose que $(a, b) \in \mathbb{R}^2$ et on s'intéresse à l'équation

$$y' = ay + b. \tag{2.4}$$

1. Identifiez une solution particulière de l'équation [32], et en déduire l'ensemble des solutions.
2. Appliquez la méthode de la variation de la constante pour retrouver l'ensemble des solutions de (2.4).

3 Equations linéaires

3.1 Théorème de Cauchy-Lipschitz linéaire

On se donne $A : I \to \mathcal{M}_d(\mathbb{K})$ et $B : I \to \mathbb{K}^d$ deux fonctions continues, et on s'intéresse dans ce paragraphe aux équations de la forme [33]

$$Y' = A(t)Y + B(t), \tag{3.1}$$

c'est-à-dire qu'on suppose $f : (t, Y) \in I \times \mathbb{K}^d \mapsto A(t)Y + B(t)$, ce qui se réécrit

$$\left\{ \begin{array}{ccl} y_1' & = & a_{11}(t)y_1 + a_{12}(t)y_2 + \ldots + a_{1d}(t)y_d + b_1(t) \\ & \vdots & \\ y_d' & = & a_{d1}(t)y_1 + a_{d2}(t)y_2 + \ldots + a_{dd}(t)y_d + b_d(t) \end{array} \right.$$

où $A = (a_{ij})_{1\le i,j\le d}$, $B = (b_i)_{1\le i\le d}$ et $Y = (y_i)_{1\le i\le d}$.

On parle d'un système linéaire d'équations différentielles. Remarquons qu'il s'agit d'une généralisation au cas vectoriel de (2.2).

Théorème 3.1 (Théorème de Cauchy-Lipschitz linéaire)**.** *Soit I un intervalle ouvert, et $A : I \to \mathcal{M}_d(\mathbb{K})$ et $B : I \to \mathbb{K}^d$ deux fonctions continues. Alors*

1. *pour tout $(t_0, Y_0) \in I \times \mathbb{K}^d$, le problème de Cauchy*

$$\left\{ \begin{array}{ccl} Y' & = & A(t)Y + B(t) \\ Y(t_0) & = & Y_0 \end{array} \right.$$

admet une unique solution globale.

2. *l'ensemble S_H des solutions de $Y' = A(t)Y$ (correspondant au cas $B \equiv 0$) est un $(\mathbb{K})-$espace vectoriel de dimension d ; plus précisément, si $(Y_i)_{1\le i\le d}$ sont d solutions de $Y' = A(t)Y$, on a les équivalences :*

$$\begin{aligned} (Y_i)_{1\le i\le d} \text{ est une base de } S_H \quad &\Leftrightarrow \quad (Y_i(t_0))_{1\le i\le d} \text{ est une base de } \mathbb{K}^d, \\ &\Leftrightarrow \quad \forall t \in I, (Y_i(t))_{1\le i\le d} \text{ est une base de } \mathbb{K}^d. \end{aligned} \tag{3.2}$$

3. *on obtient la forme générale des solutions $Y' = A(t)Y + B(t)$ en sommant une solution particulière de $Y' = A(t)Y + B(t)$ avec la forme générale des solutions de $Y' = A(t)Y$; autrement dit si S_{NH} est l'ensemble des solutions de $Y' = A(t)Y + B(t)$, et $Y_p \in S_{NH}$, alors*

$$S_{NH} = Y_p + S_H = \Big\{ Y : I \to \mathbb{K}^d,\ \exists X \in S_H,\ Y = Y_p + X \Big\}.$$

Dit encore autrement, si $(Y_i)_{1\le i\le d}$ est une base de S_H et $Y_p \in S_{NH}$, alors toutes les solutions de S_{NH} s'écrivent

$$\sum_{i=1}^{d} \alpha_i Y_i + Y_p \quad \text{où} \quad \alpha \in \mathbb{K}^d.$$

Finalement, cela revient à dire que S_{NH} est un espace affine de dimension d et dirigé par S_H.

32. Si vous ne trouvez pas, faites la question suivante, puis revenez en arrière pour voir la solution particulière(-ment simple) qu'on suggérait de trouver !

33. On change ici les notations du paragraphe 1 en réservant la notation y pour des fontions scalaires, et Y pour des fonctions à valeurs vectorielles.

Remarque 3.2. Ce résultat généralise en quelque sorte la proposition 2.3, mais on n'a pas accès à une formule qui généraliserait (2.2), sauf dans le cas où A est constant (indépendant de t) comme on le voit au paragraphe suivant. Ça ne nous empêche pas de montrer les différentes propriétés qui avaient été obtenues comme conséquence de la formule (2.2).

Démonstration. 1. On commence par remarquer que la fonction $f : (t, Y) \in I \times \mathbb{K}^d \mapsto A(t)Y + B(t)$ satisfait les hypothèses du théorème 1.33 de Cauchy-Lipschitz global (dans la version plus générale de la remarque 1.34) [34] : en effet f est continue et

$$\forall \Big((t, Y), (t, \widehat{Y})\Big) \in (I \times \mathbb{K}^d)^2, \ \ \|f(t, \widehat{Y}) - f(t, Y)\| \leq \|A(t)\|_{\mathcal{M}_d(\mathbb{K})} \|\widehat{Y} - Y\|,$$

où $\|\cdot\|$ est une norme sur $\mathbb{K}^d$ et $\|\cdot\|_{\mathcal{M}_d(\mathbb{K})}$ est sa norme subordonnée sur $\mathcal{M}_d(\mathbb{K})$.

2. Le fait que S_H est un espace vectoriel se voit facilement en voyant qu'il s'agit d'un s.e.v. de l'ensemble des fonctions de I dans $\mathbb{K}^d$. Pour $t \in I$, on définit l'application $\delta_t : Y \in S_H \mapsto Y(t) \in \mathbb{K}^d$. On vérifie avec la définition que ces applications sont linéaires. Comme pour $t_0 \in I$ donné, le point précédent montre que pour tout $Y_0 \in \mathbb{K}^d$, il existe une unique fonction $Y \in S_H$ telle que $\delta_{t_0}(Y) = Y_0$, la fonction δ_{t_0} est bijective (un isomorphisme) ; on en déduit $\dim(S_H) = \dim(\mathbb{K}^d) = d$, ainsi que les équivalences (3.2) (on déduit en effet la première équivalence avec l'application δ_{t_0} qui envoie une base de S_H sur une base de $\mathbb{K}^d$, et comme cela ne dépend pas du choix de t_0, on en déduit l'équivalence suivante).
3. Si $Y_p \in S_{NH}$ alors en écrivant les équations on obtient facilement

$$Y \in S_{NH} \ \Leftrightarrow \ Y - Y_p \in S_H.$$

□

3.2 Le cas $A(t) \equiv A$

On a dit à la remarque 3.2 que l'équation (3.1) n'avait pas de formule générale de résolution si $d \geq 2$; néanmoins dans le cas A constant, on peut généraliser la proposition 2.3 :

Proposition 3.3. *Soit $A \in \mathcal{M}_d(\mathbb{K})$ et $B : I \to \mathbb{K}^d$ continue.*

1. *Les solutions maximales de*

$$Y' = AY + B(t) \tag{3.3}$$

sont

$$\begin{aligned} \forall t \in I, \ Y(t) \ &= \ e^{A(t-t_0)}\left(\lambda + \int_{t_0}^{t} e^{(t_0-s)A} B(s)ds\right), \\ &= \ e^{A(t-t_0)}\lambda + \int_{t_0}^{t} e^{A(t-s)} B(s)ds, \end{aligned} \tag{3.4}$$

où $\lambda \in \mathbb{K}^d$, $t_0 \in I$.

2. *Le problème de Cauchy en $(t_0, Y_0) \in I \times \mathbb{K}^d$ associé à (3.3) admet une unique solution maximale, donnée par la formule (2.3) avec $\lambda = Y_0$.*

Remarque 3.4. La formule (3.4) s'appelle à nouveau formule de Duhamel.

34. Remarquons qu'il existe dans la littérature des preuves du théorème de Cauchy-Lipschitz linéaire (voir par exemple [Dan07, Théorème 26.1]) ; c'est d'ailleurs fréquemment proposé par les étudiants dans la leçon **221**, mais il me semble que la démonstration en question est en fait la même que la seconde preuve du théorème 1.33 (je devrais dire "les" démonstrations, puisqu'il y a plusieurs variantes ; néanmoins, il me semble que ma remarque vaut pour toutes les démonstrations ; je précise que je peux me tromper, n'hésitez pas à me contacter si vous connaissez une preuve du théorème de Cauchy-Lipschitz linéaire qui ne s'adapte clairement pas au cas globalement lipschitzien). Pour cette raison, disons que je ne suis pas fan de cette idée, la méthode n'ayant en fait rien de vraiment linéaire. Disons qu'au minimum, il faut s'être posé la question et avoir conscience que l'on montre un cas particulier d'un résultat plus général (sans pour autant avoir une preuve plus simple).

Exemple 3.5. Si B est définie sur $\mathbb{R}$, on peut choisir $t_0 = 0$ et retenir la formule

$$Y(t) = e^{At}\left(\lambda + \int_0^t e^{-sA}B(s)ds\right) = e^{At}\lambda + \int_0^t e^{A(t-s)}B(s)ds.$$

Exemple 3.6. Le cas $B \equiv 0$ est très important à retenir ; d'ailleurs, comme expliqué dans la preuve ci-dessous, c'est ce cas qu'on conseille de retenir par cœur, le reste étant obtenu par la méthode de variation de la constante. Il a pour conséquence que si $(x_i)_{1\leq i\leq d}$ est une base de $\mathbb{K}^d$, alors

$$\forall i \in [\![1, d]\!], \quad Y_i : t \in \mathbb{R} \mapsto\ e^{tA}x_i$$

forme une base de S_H.

Démonstration. On commence par le cas homogène en supposant $B \equiv 0$ (auquel cas on peut considérer $I = \mathbb{R}$) : soit $(t_0, Y_0) \in \mathbb{R} \times \mathbb{K}^d$. D'après le théorème 3.1, le problème de Cauchy pour $Y' = AY$ en (t_0, Y_0) admet une unique solution. Il suffit donc de vérifier que la fonction $Y : t \in \mathbb{R} \mapsto e^{A(t-t_0)}Y_0$ est solution. On pose

$$\varphi : t \in \mathbb{R} \mapsto e^{tA} := \sum_{n\in\mathbb{N}} f_n(t) \in \mathcal{M}_d(\mathbb{K}), \qquad \text{où} \quad \forall n \in \mathbb{N},\ f_n(t) = \frac{(tA)^n}{n!}.$$

Pour tout $n \in \mathbb{N}$, $f_n : \mathbb{R} \to \mathcal{M}_d(\mathbb{K})$ est de classe C^1 sur $\mathbb{R}$ et

$$f_0' = 0, \qquad \text{et} \qquad \forall n \in \mathbb{N}^*, \ \ \forall t \in \mathbb{R}, \ \ f_n'(t) = \frac{t^{n-1}A^n}{(n-1)!}$$

La série $\sum_{n\in\mathbb{N}} f_n$ converge ponctuellement, et pour tout $T \in \mathbb{R}_+$,

$$\sum_{n=0}^{N} \|f_n'\|_{\infty,[-T,T]} \leq \sum_{n=1}^{N} \frac{T^{n-1}\|A\|^n}{(n-1)!} \leq \|A\| \sum_{n=1}^{+\infty} \frac{T^{n-1}\|A\|^{(n-1)}}{(n-1)!} = \|A\|e^{\|A\|T} < +\infty$$

donc par la proposition 4.13 au Chapitre 1, la fonction φ est de classe C^1 sur tout $[-T, T]$ avec $T \in \mathbb{R}_+$, donc φ est C^1 sur $\mathbb{R}$ et

$$\forall t \in \mathbb{R}, \quad \varphi'(t) = \sum_{n\in\mathbb{N}} f_n'(t) = A \sum_{n\in\mathbb{N}} \frac{t^{n-1}A^{n-1}}{(n-1)!} = Ae^{tA} = A\varphi(t).$$

En conséquence, comme $Y = \varphi(t - t_0)Y_0$, on a

$$\forall t \in \mathbb{R}, \ \ Y'(t) = \varphi'(t - t_0)Y_0 = A\varphi(t - t_0)Y_0 = AY(t).$$

Pour le cas général, on peut soit :

— constater que (3.4) fournit bien une solution de $Y' = AY + B(t)$, et conclure via le théorème 3.1,

— soit appliquer la méthode de variation de la constante : on cherche une solution particulière de la forme $Y_p : t \in I \mapsto e^{(t-t_0)A}\lambda(t)$ avec $\lambda : I \to \mathbb{K}$ de classe C^1. Alors

$$\begin{aligned} Y_p' = AY_p + B \quad &\Leftrightarrow \quad \forall t \in I, \ Ae^{(t-t_0)A}\lambda(t) + e^{(t-t_0)A}\lambda'(t) = Ae^{(t-t_0)A}\lambda(t) + B(t) \\ &\Leftrightarrow \quad \forall t \in I, \ \lambda'(t) = e^{-A(t-t_0)}B(t) \end{aligned}$$

Pour simplifier, on cherche à avoir $Y_p(t_0) = 0$, ce qui donne comme solution particulière

$$\forall t \in I, \ Y_p(t) = e^{(t-t_0)A}\lambda(t) \quad \text{avec } \lambda(t) = \int_{t_0}^t e^{-A(s-t_0)}B(s)ds$$

Et pour conclure, on voit que Y est solution du problème de Cauchy pour $Y' = AY + B(t)$ en (t_0, Y_0) si et seulement si $Z = Y - Y_p$ est solution du problème de Cauchy pour $Z' = AZ$ en $(t_0, 0)$, d'où

$$Y(t) = e^{(t-t_0)A}Y_0 + Y_p(t) = e^{(t-t_0)A}Y_0 + e^{(t-t_0)A}\int_{t_0}^t e^{-A(s-t_0)}B(s)ds,$$

d'où le résultat.

□

Exercice 3.7 (Autre preuve de la dérivée de φ)**.** Soit $A \in \mathcal{M}_d(\mathbb{K})$. On pose

$$\forall t \in \mathbb{R}, \quad \varphi(t) = e^{tA}.$$

1. Montrer que
$$\lim_{t \to 0} \left[\frac{\varphi(t) - \varphi(0)}{t} - A \right] = 0,$$
et en déduire que φ est dérivable en 0 et $\varphi'(0) = A$.
2. Utilisez la question précédente pour montrer que φ est dérivable en tout $t \in \mathbb{R}$ et $\varphi'(t) = Ae^{tA}$.[35]

Exemple 3.8. Il va sans dire que pour comprendre ces systèmes d'équations différentielles, on doit s'y connaître en exponentielle matricielle : on conseille de réviser en se posant les questions suivantes (n'hésitez pas à détailler des exemples, même dans le cas $d = 2$) :

1. A est diagonalisable sur $\mathbb{K}$: quelle est l'allure générale des solutions ?
2. on se place dans le cas $\mathbb{K} = \mathbb{R}$ mais A est diagonalisable sur $\mathbb{C}$; bien sûr, rien n'interdit de diagonaliser sur $\mathbb{C}$, mais si on veut rester cohérent, il faut chercher à voir comment se ramener à des fonctions à valeurs réelles : prenons l'exemple
$$\left\{ \begin{array}{rcl} y_1' & = & y_2 \\ y_2' & = & -y_1 \end{array} \right. \qquad \text{c'est-à-dire} \qquad \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}' = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{3.5}$$
   (a) première approche : on diagonalise dans $\mathbb{C}$ (c'est-à-dire qu'on utilise le cas précédent, qu'on décrit rapidement ici) : alors on obtient rapidement
   $$A = PDP^{-1} \quad \text{avec} \quad D = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$
   et donc
   $$\forall t \in \mathbb{R}, \quad e^{tA} = P \begin{pmatrix} e^{it} & 0 \\ 0 & e^{-it} \end{pmatrix} P^{-1}$$
   En conséquence, les solutions à valeurs complexes de (3.5) sont de la forme[36]
   $$\forall t \in \mathbb{R}, \qquad y_1(t) = \alpha e^{it} + \beta e^{-it}, \qquad y_2(t) = i\alpha e^{it} - i\beta e^{-it}$$
   où $(\alpha, \beta) \in \mathbb{C}$. Si on ajoute la condition que y_1, y_2 doivent être à valeurs réelles (il suffit de le faire pour y_1 puisque $y_2 = y_1'$), on obtient la condition (nécessaire et suffisante)
   $$\alpha = \overline{\beta} \quad \text{et} \quad \beta = \overline{\alpha},$$
   d'où
   $$\forall t \in \mathbb{R}, \; y_1(t) = \alpha e^{it} + \overline{\alpha} e^{-it} = 2\mathrm{Re}\left(\alpha e^{it}\right) = 2\mathrm{Re}(\alpha)\cos(t) - 2\mathrm{Im}(\alpha)\sin(t).$$
   Au final, les solutions à valeurs réelles de (3.5) sont
   $$\forall t \in \mathbb{R}, \quad y_1(t) = \lambda_1 \cos(t) + \lambda_2 \sin(t), \qquad y_2(t) = -\lambda_1 \sin(t) + \lambda_2 \cos(t),$$
   où $(\lambda_1, \lambda_2) \in \mathbb{R}^2$.

35. Pour rappel, on pourra utiliser que si A et B commutent, alors $e^{A+B} = e^A e^B$.

36. Remarquons qu'on peut facilement calculer $P = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$ et $P^{-1} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -i & 1 \end{pmatrix}$, ce qu'on pourrait utiliser, mais il important de noter qu'on se passe de cette information ici. En effet, on a simplement écrit que y_1 est combinaison linéaire de $t \mapsto e^{it}$ et $t \mapsto e^{-it}$ et on a écrit $y_2 = y_1'$.

(b) La deuxième approche consiste à observer que $J^2 = -I_2$, et donc on obtient [37]

$$\forall n \in \mathbb{N}, \quad J^{2n} = (-1)^n I_2 \quad J^{2n+1} = (-1)^n J,$$

d'où

$$\exp(tA) = \sum_{n\in\mathbb{N}} \frac{J^{2n}}{(2n)!} + \sum_{n\in\mathbb{N}} \frac{J^{2n+1}}{(2n+1)!} = \cos(t)I_2 + \sin(t)J = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix}$$

ce qui donne le même résultat en écrivant

$$\begin{pmatrix} y_1(t) \\ y_2(t) \end{pmatrix} = \exp(tA) \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}.$$

3. que faire dans le cas où la matrice est trigonalisable mais non diagonalisable ? À titre d'exemple, résolvez le système

$$\begin{cases} y_1' &= y_1 + y_2 \\ y_2' &= y_2 \end{cases}$$

En fait, on peut montrer plus généralement les deux résultats suivants (voir par exemple [Dan, page 15-16] [38]) :

Proposition 3.9. *1. Dans le cas $A \in \mathcal{M}_d(\mathbb{C})$, on note $(\lambda_1, \dots, \lambda_p) \in \mathbb{C}^p$ les valeurs propres distinctes de A, et $(\alpha_j)_{1\le j\le p}$ leurs multiplicités. Alors les composantes des solutions de $Y' = AY$ sont combinaisons linéaires des fonctions*

$$t \in \mathbb{R} \mapsto t^k e^{\lambda_j t}, \quad k \in [\![0, \alpha_j - 1]\!],\ j \in [\![1, p]\!].$$

2. Dans le cas $A \in \mathcal{M}_d(\mathbb{R})$, on note $(\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$ les valeurs propres réelles de A [39] de multiplicités $(\alpha_j)_{1\le j\le p}$ et

$$(\mu_k \pm i\nu_k)_{1\le k\le q}$$

les valeurs propres complexes conjuguées et non réelles de A, de multiplicités $(\beta_j)_{1\le j\le q}$. Alors les composantes des solutions réelles de $Y' = AY$ sont combinaisons linéaires réelles des fonctions

$$t \mapsto t^k e^{\lambda_j t} \quad k \in [\![0, \alpha_j - 1]\!],\ j \in [\![1, p]\!]$$

et

$$t \mapsto t^k e^{\mu_j t}\cos(\nu_j t) \text{ et } t \mapsto t^k e^{\mu_j t}\sin(\nu_j t) \quad k \in [\![0, \beta_j - 1]\!],\ j \in [\![1, q]\!]$$

La première partie peut se démontrer avec la décomposition de Jordan, la seconde se montre par exemple à partir de la première.

3.3 *Le cas $t \mapsto A(t)$ non constant

Dans le cas général où A n'est pas constant, on ne peut pas toujours résoudre explicitement (3.1). On peut néanmoins donner quelques éléments qui permettent d'étudier les solutions : en premier lieu, on peut généraliser l'idée de la méthode de variation de la constante qui dit qu'on peut exprimer les solutions de l'équation non homogène à partir des solutions de l'équation homogène :

Définition-Proposition 3.10 (Résolvante)**.** Pour tout $(s,t) \in \mathbb{R}^2$, on pose $R(t,s) \in \mathcal{M}_d(\mathbb{K})$ l'unique matrice telle que la solution de

$$\begin{cases} Y' &= A(t)Y \\ Y(s) &= Y_0 \end{cases} \tag{3.6}$$

pour $Y_0 \in \mathbb{K}^d$ soit $Y(t) = R(t,s)Y_0$. On appelle R la résolvante de l'équation $Y' = A(t)Y$.

37. Inutile de faire une récurrence !!!
38. Mais il manque un t^β dans l'énoncé.
39. Il se peut que $p = 0$, c'est-à-dire que A n'a pas de valeurs propres réelles. Rappelons quand même que si d est impair alors A a au moins une valeur propre réelle.

Exemple 3.11. Si $A(t) = A$ pour tout t, alors $R(t,s) = e^{A(t-s)}$ pour tout $(s,t) \in \mathbb{R}^2$.

Démonstration. Il faut justifier cette définition : par le théorème 3.1, pour tout $(s, Y_0) \in \mathbb{R} \times \mathbb{K}^d$, il existe une unique fonction $t \mapsto Y_{s,Y_0}(t)$ solution de (3.6) ; la dépendance en Y_0 est linéaire, donc à (s,t) fixés, on a $Y_0 \in \mathbb{K}^d \mapsto Y_{s,Y_0}(t) \in \mathbb{K}^d$ qui est une application linéaire d'où l'existence et l'unicité de la matrice $R(t,s)$. □

Même si on ne peut pas toujours calculer R, on peut donner les propriétés suivantes :

Proposition 3.12. *1. Pour tout $t \in \mathbb{R}$, $R(t,t) = I_d$.*

2. Pour tout (t,s,u), $R(t,u) = R(t,s)R(s,u)$.

3. Pour tout (t,s),
$$\partial_t R(t,s) = A(t)R(t,s), \quad \partial_s R(t,s) = -R(t,s)A(s).$$

Cela découle facilement des définitions.
Enfin, on peut généraliser les propositions 2.3 et 3.3 à ce cadre général :

Proposition 3.13. *Si R est la résolvante de l'équation $Y' = A(t)Y$, alors l'unique solution de*
$$\begin{cases} Y' &= A(t)Y + B(t) \\ Y(t_0) &= Y_0 \end{cases}$$
est
$$\forall t \in I,\ Y(t) = R(t,t_0)Y_0 + \int_{t_0}^{t} R(t,s)B(s)ds.$$

La preuve est la même que précédemment, on peut soit constater que la formule ci-dessus fournit une solution et invoquer l'unicité du théorème 3.1, ou on peut appliquer la méthode de variation de la constante pour trouver une solution particulière et conclure le raisonnement comme précédemment.
Afin d'étudier la fonction R, on peut utiliser les outils suivants :

Définition 3.14. Soit $(Y_i)_{1\le i\le d}$ des solutions indépendantes de $Y' = A(t)Y$. On pose
$$\forall t \in \mathbb{R}, \quad W(t) = (Y_1(t), \dots, Y_d(t))$$
qu'on appelle matrice wronskienne de $(Y_i)_{1\le i\le d}$, et
$$\forall t \in \mathbb{R}, \quad w(t) = \det(W(t))$$
qu'on appelle wronskien de $(Y_i)_{1\le i\le d}$.

On a alors le résultat suivant :

Proposition 3.15. *Avec les notations précédentes :*

1. On a la relation suivante :
$$\forall (t,s) \in \mathbb{R}^2, \quad R(t,s) = W(t)W(s)^{-1}$$

2. S'il existe $t_0 \in \mathbb{R}$ tel que $W(t_0)$ est inversible, alors $W(t)$ est inversible pour tout $t \in \mathbb{R}$.

3. Enfin on a la relation
$$w' = \mathrm{tr}(A)w, \quad \textit{ce qui donne} \quad w(t) = w(t_0)\exp\left(\int_{t_0}^{t} \mathrm{tr}(A(s))ds\right).$$

Démonstration. 1. Pour tout $(t,s) \in \mathbb{R}^2$ et $Y_0 \in \mathbb{K}^d$ on constate que
$$\frac{d}{dt}\left[W(t)W(s)^{-1}Y_0\right] = W'(t)W(s)^{-1}Y_0 = A(t)W(t)W(s)^{-1}Y_0, \quad \text{et} \quad W(s)W(s)^{-1}Y_0 = Y_0$$
donc par définition $(t,s) \mapsto W(t)W(s)^{-1}$ est bien la résolvante de $Y' = A(t)Y$.

2. C'est une conséquence de l'équivalence (3.2) au théorème 3.1.
3. Si w s'annule en un $t_0 \in \mathbb{R}$, alors par le point w est constant égal à 0, et donc satisfait bien $w' = \mathrm{tr}(A)w$. Sinon w ne s'annule pas, et donc $W(t)$ est une matrice inversible en tout $t \in \mathbb{R}$. En conséquence en utilisant la formule $d\det(A)(H) = \det(A)\mathrm{tr}(A^{-1}H)$ valable pour tout $A \in GL_d(\mathbb{K})$ et $H \in \mathcal{M}_d(\mathbb{K})$, on obtient par dérivation d'une composée : [40]
$$\begin{aligned} \forall t \in \mathbb{R}, \quad w'(t) &= d\det(W(t))(W'(t)) = \det(W(t))\mathrm{tr}(W(t)^{-1}W'(t)) \\ &= \det(W(t))\mathrm{tr}(W(t)^{-1}W(t)A(t)) = \mathrm{tr}(A(t))w(t). \end{aligned}$$

□

Voir par exemple [Gou08b] pour des exemples d'utilisation du wronskien.

3.4 Equations linéaires scalaires d'ordre n

On se donne $n \in \mathbb{N}^*$, $a_0, \ldots, a_{n-1} : I \to \mathbb{K}$ et $b : I \to \mathbb{K}$ des fonctions continues, et on considère l'équation scalaire d'ordre n
$$y^{(n)} = a_{n-1}(t)y^{(n-1)} + \ldots + a_0 y(t) + b(t). \tag{3.7}$$
Avec la méthode de la remarque 1.6, on peut réécrire cette équation sous la forme
$$Y' = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \ldots & 1 \\ a_0(t) & a_1(t) & \cdots & & a_{n-1}(t) \end{pmatrix} Y + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b(t) \end{pmatrix} \quad \text{où} \quad Y(t) = \begin{pmatrix} y(t) \\ y'(t) \\ \vdots \\ \vdots \\ y^{(n-1)}(t) \end{pmatrix}$$
On est donc ramené au cas d'une équation linéaire d'ordre 1 avec $t \mapsto A(t) \in \mathcal{M}_n(\mathbb{K})$ qui est une fonction à valeurs dans les matrices compagnons.

Remarque 3.16. Ainsi la théorie des paragraphes précédents s'applique :

— l'ensemble des solutions est un espace affine de dimension n,

— les solutions s'obtiennent en ajoutant aux solutions de l'équation homogène une solution particulière,

— pour tout $(y_0, \ldots, y_{n-1}) \in \mathbb{R}^n$ et $t_0 \in I$, il existe une unique solution maximale telle que
$$y(t_0) = y_0, y'(t_0) = y_1, \ldots, y^{(n-1)}(t_0) = y_{n-1}.$$

— si $(y_1, \ldots y_n)$ sont des solutions indépendantes, alors la matrice wronskienne est
$$\begin{pmatrix} y_1 & y_2 & \cdots & y_n \\ y_1' & y_2' & \cdots & y_n' \\ \vdots & & & \\ y_1^{(n-1)} & y_2^{(n-1)} & \ldots & y_n^{(n-1)} \end{pmatrix}$$
et l'équation du wronskien devient dans ce cas $w' = a_{n-1}(t)w$.

On suppose maintenant que les fonctions $(a_i)_{0 \le i \le n-1}$ sont constantes : la proposition 3.9 nous incite donc à chercher les valeurs propres de A qui sont les racines de
$$\chi_A = \det(XI_n - A) = X^n - a_{n-1}X^{n-1} - \ldots a_1 X - a_0.$$

40. Voir [Boyb, page 24] ou [Gou08b, page 368] pour une preuve différente.

Définition 3.17. L'équation

$$r^n - a_{n-1}r^n - \dots a_1 r - a_0 = 0$$

s'appelle l'équation caractéristique de l'équation (3.7)

Exemple 3.18. Décrivons un peu plus en détail le cas $n = 2$ (voir par exemple [Dan07, Chapitre 27]). L'équation s'écrit alors

$$y'' = a_1 y' + a_0 + b(t). \tag{3.8}$$

On suppose également que a_0, a_1 sont réels et qu'on cherche les solutions à valeurs réelles (i.e. $\mathbb{K} = \mathbb{R}$).

1. commençons par l'équation homogène où on suppose $b \equiv 0$. L'équation caractéristique s'écrit

$$r^2 = a_1 r + a_0. \tag{3.9}$$

   (a) Si (3.9) a deux racines réelles $r_1, r_2 \in \mathbb{R}^2$ distinctes (le discriminant est strictement positif), alors les solutions de $y'' = a_1 y' + a_0$ seront de la forme

$$t \mapsto \alpha e^{r_1 t} + \beta e^{r_2 t}, \quad (\alpha, \beta) \in \mathbb{R}^2$$

   Cela correspond au cas où la matrice A est diagonalisable dans $\mathbb{R}$.

   (b) Si (3.9) a deux racines complexes distinctes $\mu \pm i\nu$ (le discriminant est strictement négatif), alors les solutions de $y'' = a_1 y' + a_0 y$ seront de la forme

$$t \mapsto \alpha e^{\mu t}\cos(\nu t) + \beta e^{\mu t}\sin(\nu t) \quad (\alpha, \beta) \in \mathbb{R}^2.$$

   C'est le cas où la matrice A est diagonalisable dans $\mathbb{C}$ mais pas dans $\mathbb{R}$.

   (c) Si (3.9) a une racine réelle double $r_0 \in \mathbb{R}$ (c'est le cas où la matrice A n'est pas diagonalisable), alors les solutions de $y'' = a_1 y' + a_0$ seront de la forme

$$t \mapsto e^{r o t}(\alpha + \beta t), \quad (\alpha, \beta) \in \mathbb{R}^2.$$

2. Pour le cas non homogène, on peut soit appliquer la formule (3.4), soit adapter la méthode de variation de la constante à ce contexte [41] afin de trouver une solution particulière : si y_1, y_2 sont deux solutions de l'équation homogène, on cherche une solution particulière de (3.8) telle que [42]

$$\lambda(t)\begin{pmatrix} y_1 \\ y_1' \end{pmatrix} + \mu(t)\begin{pmatrix} y_2 \\ y_2' \end{pmatrix} \text{ soit solution de } Y' = \begin{pmatrix} 0 & 1 \\ a_0 & a_1 \end{pmatrix} Y + \begin{pmatrix} 0 \\ b(t) \end{pmatrix}$$

   ce qui amène aux égalités :

$$\begin{cases} \lambda' y_1 + \mu' y_2 = 0 \\ \lambda' y_1' + \mu' y_2' + b = 0 \end{cases}$$

   Ce système dont les inconnues sont (λ', μ') est résoluble car son déterminant est le wronskien de (y_1, y_2) qui est non nul puisque (y_1, y_2) sont indépendantes. Une fois calculée des primitives de λ et μ, on obtient une solution particulière de (3.8) à laquelle on ajoute les solutions de l'équation homogène pour trouver l'ensemble des solutions.

4 Equations différentielles autonomes

Dans ce paragraphe, on s'intéresse aux équations différentielles de la forme

$$y' = f(y) \tag{4.1}$$

où $f : U \to \mathbb{R}^d$ est de classe C^1 et U est un ouvert de $\mathbb{R}^d$; autrement dit, f ne dépend pas de t, mais seulement de la variable y. On dit qu'une telle équation est autonome [43]

41. Attention, il peut arriver qu'il soit plus simple de chercher une solution particulière "à vue de nez", notamment quand b est de la forme $P(t)e^{rt}$ où P est un polynôme.

42. Attention, il ne suffit pas de chercher une solution sous la forme $\lambda(t)y_1 + \mu(t)y_2$, qui ne donne pas suffisamment d'information.

43. Remarquons qu'une équation différentielle non autonome $y' = f(t, y)$ peut être rendue autonome en remplaçant l'inconnue y de dimension d par l'inconnue $\begin{pmatrix} y \\ \alpha \end{pmatrix}$ de dimension $d + 1$ et qui est solution de l'équation

$$\begin{pmatrix} y(t) \\ \alpha(t) \end{pmatrix}' = \begin{pmatrix} f(\alpha(t), y(t)) \\ 1 \end{pmatrix}$$

4.1 Remarques générales ; flot d'un champ de vecteur

Naturellement rien n'empêche de considérer $\widetilde{f}(t,y) = f(y)$ pour $(t,y) \in \mathbb{R} \times \mathbb{R}^d$ et donc la théorie du paragraphe 1 s'applique à ces équations avec $I = \mathbb{R}$. En particulier si f est localement lipschitzienne (par exemple si elle est de classe C^1), le problème de Cauchy associé à (4.1) admet une unique solution maximale.

Proposition 4.1. *On suppose f localement lipschitzienne. Soit $(t_0, y_0) \in \mathbb{R} \times U$ et $\tau \in \mathbb{R}$. Alors les solutions maximales (J, y) et (J_τ, y_τ) de (4.1) respectivement associées aux problèmes de Cauchy (t_0, y_0) et $(t_0 + \tau, y_0)$ sont liées par*

$$J_\tau = J + \tau, \quad \forall t \in J_\tau,\ y_\tau(t) = y(t - \tau).$$

Remarque 4.2. Ce résultat simple à prouver à des conséquences importantes :

— il suffit de supposer $t_0 = 0$: en effet si on décrit toutes les solutions de (4.1) avec $y(0) = y_0$ pour $y_0 \in U$ quelconque, alors on peut reconstruire toutes les solutions de (4.1) en translatant ces dernières.

— Si (J, y) est une solution de (4.1), on appelle trajectoire de la solution l'ensemble $\{y(t)\}_{t\in J}$. Un corollaire de la proposition 4.1 est que deux trajectoires de solutions de (4.1) sont soient identiques, soit ne s'intersectent pas. [44]

— Pour la raison précédente, plutôt que de dessiner les ensembles $\{(t, y(t))\}_{t\in J}$, difficiles à dessiner notamment si $d \geq 2$, on peut se contenter de dessiner les trajectoires $\{y(t)\}_{t\in J}$. Deux cas se présentent (on suppose comme souvent qu'on est dans le cadre d'application du théorème de Cauchy-Lipschitz) :

1. soit la trajectoire est réduite à un point, et la solution est constante : on parle de solution stationnaire, ou de point d'équilibre [45] de f ; pour les trouver il faut résoudre l'équation $f(y) = 0$.
2. soit la trajectoire est une courbe C^1 sans point singulier (car y' ne s'annule pas : il s'agit d'un cas particulier du théorème 4.6 adapté aux courbes paramétrées).

On peut donc dessiner ce qu'on appelle un portrait de phase, en traçant l'allure de différentes trajectoires ; ce n'est pas une description parfaite des solutions (on ne sait pas à un instant donné la "vitesse" de parcours le long de la courbe trajectoire), mais ça peut donner une bonne idée. Remarquons qu'on peut aussi mettre une flèche sur les trajectoires pour connaître le sens de parcours.

Preuve de la proposition 4.1 : Étant donné une solution (J, y) telle que $y(t_0) = y_0$, on pose $J_\tau = J + \tau$ et $z_\tau : t \in J_\tau \mapsto y(t - \tau)$. Alors z_τ est bien définie, est maximale, et est solution de

$$\forall t \in J_\tau, z_\tau'(t) = y'(t - \tau) = f(y(t - \tau)) = f(z_\tau(t)), \qquad \text{et} \quad z_\tau(t_0 + \tau) = y(t_0) = y_0$$

donc par unicité de la solution maximale (corollaire 1.25), $z_\tau = y_\tau$. □

Définition 4.3. On appelle champ de vecteur sur U toute fonction $f : U \to \mathbb{R}^d$ de classe C^1.
On dit que ce champ de vecteur est complet si les solutions de (4.1) sont globales, c'est-à-dire définies sur $\mathbb{R}$.

---

qui est une équation autonome, et si on ajoute la condition de Cauchy $(y(t_0), \alpha(t_0)) = (y_0, t_0)$, est équivalente à $y' = f(t, y)$ avec $y(t_0) = y_0$.

44. Ne pas confondre l'idée que deux trajectoires se croisent, avec l'idée que deux solutions se croisent ; si y_1, y_2 sont deux solutions, la première hypothèse suppose qu'il existe (t_0, t_1) tels que $y_1(t_0) = y_2(t_1)$, alors que la seconde suppose qu'il existe t_0 tel que $y_1(t_0) = y_2(t_0)$, ce qui est une hypothèse plus forte. À la remarque 1.26 on a montré que deux solutions de $y' = f(t, y)$ (équation non nécessairement autonome) ne peuvent pas se croiser ; maintenant on donne un résultat plus fort sur l'équation autonome $y' = f(y)$ qui dit que les trajectoires ne peuvent pas se croiser. Mais ce dernier résultat n'est pas valable pour les équations non autonomes.

45. Plus précisément, l'idée d'une solution stationnaire est valable pour une équation non autonome, alors que la notion de point d'équilibre (solution de $f(y) = 0$) ne vaut que pour les équations autonomes.

Définition 4.4 (Flot d'un champ de vecteur)**.** Soit $f : U \to \mathbb{R}^d$ un champ de vecteur complet [46]. On pose

$$\forall t \in \mathbb{R}, \ \ \forall y_0 \in U, \quad \phi(t, y_0) \text{ ou encore } \phi_t(y_0)$$

la valeur en t de l'unique solution de $y' = f(y)$ et telle que $y(0) = y_0$. On appelle ϕ ou $(\phi_t)_{t\in\mathbb{R}}$ le flot du champ de vecteur f.

Remarque 4.5. En effet, d'après la proposition 4.1, la trajectoire ne dépend que de la valeur de y_0 si on décide arbitrairement qu'il s'agit de la valeur en $t = 0$. De ce fait, on souhaite insister sur la dépendance en y_0, la dépendance en t étant possiblement mise en indice ; ainsi $\phi_t(y_0)$ est la position au temps t de la trajectoire partie de y_0.

On pourrait parler de flot pour une équation non autonome, mais alors il faudrait écrire $\phi(t, t_0, y_0)$ ou $\phi_t(t_0, y_0)$ ou $\phi_{t,t_0}(y_0)$.

Proposition 4.6 (Propriétés de base du flot)**.** *Soit $(\phi_t)_{t\in\mathbb{R}}$ le flot d'un champ de vecteur complet $f : U \to \mathbb{R}^d$: alors*

1. *$\phi_0 = Id_U$, autrement dit pour tout $y_0 \in U, \phi_0(y_0) = y_0$,*
2. *pour tout $(t, s) \in \mathbb{R}^2$,*
$$\phi_{t+s} = \phi_t \circ \phi_s = \phi_s \circ \phi_t, \ \ i.e. \quad \forall y_0 \in U, \ \ \phi_{t+s}(y_0) = \phi_t(\phi_s(y_0)) = \phi_s(\phi_t(y_0)),$$
3. *pour tout $t \in \mathbb{R}$, $\phi_t : U \to U$ est inversible et d'inverse ϕ_{-t},*
4. *pour tout $t \in \mathbb{R}$ et tout $y_0 \in U$,*
$$\frac{\partial}{\partial t}\phi_t(y_0) = f(\phi_t(y_0)).$$

Démonstration. Seul le second point mérite une démonstration, le troisième étant conséquence du second, et les deux autres venant des définitions. Pour montrer ce second point, on constate qu'à s fixé, les fonctions $y : t \mapsto \phi_{t+s}(y_0)$ et $z : t \mapsto \phi_t \circ \phi_s(y_0)$ satisfont le même problème de Cauchy en $t = 0$, et donc par unicité des solutions, sont égales, ce qui donne $\phi_{t+s} = \phi_t \circ \phi_s$. Ensuite en échangeant les rôles de s et t, on obtient $\phi_{s+t} = \phi_s \circ \phi_t$, ce qui permet de conclure puisque $t + s = s + t$. □

Théorème 4.7 (Continuité du flot)**.** *Soit $(\phi_t)_{t\in\mathbb{R}}$ le flot d'un champ de vecteur complet f. Alors pour tout $T \in \mathbb{R}_+$ et $y_0 \in U$, il existe $r_0 > 0$ et $K \in \mathbb{R}+$ tel que*

$$\forall y_1 \in B(y_0, r_0), \ \ \forall t \in [-T, T], \ \ \|\phi_t(y_1) - \phi_t(y_0)\| \le e^{K|t|}\|y_1 - y_0\|.$$

En particulier, (ϕ_t) est continue sur U, et même $(t, y_0) \mapsto (\phi_t(y_0))$ est continue.

Démonstration. On ne présente la preuve que dans le cas où f est globalement lipschitzienne, on renvoit à la littérature pour les cas généraux : dans ce cas on n'a pas besoin de supposer y_0 et y_1 proches, autrement dit on peut considérer $r_0 = +\infty$. Soit donc $(y_0, y_1) \in U^2$. L'équation intégrale donne l'estimation

$$\forall t \ge 0, \ \ \phi_t(y_1) - \phi_t(y_0) = y_1 - y_0 + \int_0^t f(\phi_s(y_1)) - f(\phi_s(y_0))ds$$

d'où en posant $u : t \mapsto \|\phi_t(y_1) - \phi_t(y_0)\|$,

$$\forall t \ge 0, \ \ u(t) \le u(0) + k\int_0^t u(s)ds$$

et donc par le lemme de Gronwall,

$$\forall t \ge 0, \ \ u(t) \le u(0)e^{kt} = \|y_1 - y_0\|e^{kt}.$$

On peut obtenir une estimation similaire pour les temps $t \le 0$ en étudiant $t \ge 0 \mapsto \phi_{-t}(y)$. □

46. On fait cette hypothèse pour simplifier.

Remarque 4.8. Si on ne suppose pas le champ de vecteur complet, pour chaque $y_0 \in U$, il existe un intervalle maximal J_{y_0} d'existence de la solution de $y' = f(y)$ telle que $y(0) = y_0$. Dans ce cas, il est naturel de définir le flot sur l'ensemble

$$(t, y_0) \in \omega := \cup_{y_0 \in U} I_{y_0} \times \{y_0\}.$$

On peut montrer que cet ensemble est ouvert et que $(t, y_0) \in \omega \mapsto \phi_t(y_0)$ est continue sur cet ouvert.

Remarque 4.9. Si le champ de vecteur f est de classe C^2, on peut même étudier la différentiabilité de ϕ_t en la variable y_0 : voir par exemple [BG10, Th. 5.13].

4.2 Stabilité

4.2.1 Vocabulaire

Définition 4.10. Soit y_0 une solution stationnaire de (4.1), c'est-à-dire $y_0 \in U$ et $f(y_0) = 0$.

1. on dit que y_0 est stable si pour tout voisinage ω_1 de y_0 dans U, il existe ω_2 voisinage de y_0 dans U tel que toute solution de (4.1) qui démarre dans ω_2 reste dans ω_1 pour tout temps $t \in \mathbb{R}_+$, c'est-à-dire
$$\forall \omega_1 \text{ voisinage de } y_0, \;\; \exists \omega_2 \text{ voisinage de } y_0 \text{ tel que } y \in \omega_2 \Rightarrow \forall t \in \mathbb{R}_+, \; \phi_t(y) \in \omega_1.$$
2. on dit que y_0 est attractif s'il existe ω un voisinage de y_0 dans U telle que toutes les solutions qui démarrent dans ω convergent vers y_0 quand $t \to +\infty$, c'est-à-dire
$$\forall \omega \text{ voisinage de } y_0, y \in \omega \Rightarrow \;\; \phi_t(y) \xrightarrow[t\to+\infty]{} y_0.$$
3. on dit que y_0 est asymptotique stable si y_0 est stable et attractif.
4. on dit que y_0 est instable si y_0 n'est pas stable.

Proposition 4.11 (Importance des points d'équilibre)**.** *On suppose $U = \mathbb{R}^d$ et $f : \mathbb{R}^d \to \mathbb{R}^d$ localement lipschitzienne. Soit (J, y) une solution maximale de (4.1). Si y admet une limite $y^* \in \mathbb{R}^d$ quand $t \to \sup(J)$* [47]*, alors y^* est un point d'équilibre.*

Démonstration. Soit (J, y) une solution maximale ayant une limite y^*. Par l'alternative d'explosion (théorème 1.27), $\sup J = +\infty$, et l'équation $y' = f(y)$ montre avec la continuité de f que $y'(t) \xrightarrow[t\to+\infty]{} g(y^*)$. Supposons par l'absurde que $g(y^*)$ est non nul, il existe $i_0 \in [\![1, d]\!]$ tel que $\alpha := g(y^*)_{i_0} \neq 0$. Supposons par exemple $\alpha > 0$, l'autre cas se traitant de façon similaire ; comme $y'_{i_0}(t) \xrightarrow[t\to+\infty]{} \alpha$, il existe $t_0 \in \mathbb{R}$ tel que pour tout $t \geq t_0$, $y'_i(t) \geq \alpha/2$. Alors par intégration on obtient

$$\forall t \geq t_0, \;\; y_{i_0}(t) = y_{i_0}(t_0) + \int_{t_0}^{t} y_{i_0}(s)ds \geq y_{i_0}(t_0) + \frac{\alpha}{2}(t - t_{i_0}) \xrightarrow[t\to+\infty]{} +\infty$$

ce qui contredit la convergence de y. Ainsi $g(y^*) = 0$. □

Remarque 4.12. Si une trajectoire $\{y(t)\}_{t\in J}$ converge, alors c'est vers un point d'équilibre. Mais attention, rien ne dit a priori que les trajectoires convergent : elles peuvent

— ne pas être bornée,

— être périodique,

— avoir un comportement plus complexe.

Exemple 4.13. On peut constater la proposition 4.11 sur l'exemple très simple $y' = y$ où $d = 1$. En effet les solutions convergent vers 0 en $-\infty$, qui est bien un point d'équilibre. Les solutions ne convergent pas en $+\infty$.

47. Le même résultat vaut pour $t \to \inf J$.

Exemple 4.14. On regarde le cas $d = 2$ et $f(x, y) = (-y, x)$. La solution $(\cos(t), \sin(t))$ est périodique, en particulier elle n'a pas de limite en $\pm\infty$. Le portrait de phase est fait de cercles concentriques de centre $(0, 0)$.

Remarque 4.15. Si $d = 1$, les équations autonomes sont un cas particulier des équations à variables séparées abordées au paragraphe 2.1. Dans ce cas on peut facilement montrer que les trajectoires ont soit des limites réelles, soit des limites infinies. Voir par exemple [Ton, paragraphe 2.3]. Pour vous exercer, tracez l'allure des trajectoires de l'équation logistique

$$y' = y(1 - y)$$

par deux méthodes radicalement différentes :

1. calculez explicitement les solutions,
2. étudiez les points d'équilibre, et étudiez les limites possibles des trajectoires suivant la condition initiale.

4.2.2 Cas des équations linéaires à coefficients constants

Commençons par étudier le cas des équations linéaires

$$Y' = AY.$$

On remarque que $Y \equiv 0$ est une solution stationnaire. On va étudier sa stabilité :

Proposition 4.16. *Soit $A \in \mathcal{M}_d(\mathbb{R})$. On pose*

$$\alpha = \max\Big\{\mathrm{Re}(\lambda), \quad \lambda \in \mathrm{Sp}(A)\Big\}$$

où $\mathrm{Sp}(A)$ désigne l'ensemble des valeurs propres complexe de A :

1. *si $\alpha < 0$ alors 0 est asymptotiquement stable,*
2. *si $\alpha > 0$ alors 0 est instable,*
3. *si $\alpha = 0$ alors 0 n'est pas attractif, et 0 est stable si et seulement si les multiplicités géométriques et algébriques des valeurs propres*[48] *de partie réelle nulle sont égales.*

Idées de preuve : Il s'agit d'étudier la limite de $e^{At}Y_0$ pour $t \to +\infty$:

1. si $\alpha < 0$, alors avec la décomposition de Dunford on arrive à montrer

$$\forall Y_0 \in \mathbb{R}^d, \quad \|e^{At}Y_0\| \leq e^{\alpha t}Q(|t|)\|Y_0\|$$

   où Q est un polynôme. Cela montre à la fois que 0 est stable et attractif, donc asymptotiquement stable.
2. si $\alpha > 0$, il y a des choix possibles pour Y_0 tels que $\|e^{At}Y_0\| \xrightarrow[t\to+\infty]{} +\infty$.
3. si $\alpha = 0$, il faut regarder les blocs de Jordan associés aux valeurs propres dont la partie réelle est nulle; si cette partie de la matrice est diagonalisable (ce qui correspond à l'égalité entre la multiplicité algébrique et la multiplicité géométrique), alors 0 est stable, sinon il y aura un terme polynomial dans l'exponentielle, et donc 0 n'est pas stable.

□

48. Pour $\lambda \in \mathrm{Sp}(A)$, la multiplicité géométrique de λ est la dimension de son espace propre $\dim(E_\lambda) = \dim(\mathrm{Ker}(A - \lambda I_d))$. Sa multiplicité algébrique est son ordre de multiplicité en tant que racine de $\chi_A = \det(A - XI_d)$, c'est-à-dire le plus grand exposant k tel que $(X - \lambda)^k$ divise χ_A. Dans la réduction de Jordan de la matrice A, la multiplicité géométrique est égale au nombre de blocs de Jordan associés à λ, et la multiplicité algébrique est égale à la somme des tailles de ces blocs.

4.2.3 Théorème de linéarisation

On revient au cas général
$$y' = f(y).$$
Grâce au calcul différentiel, nous allons lier dans certains cas la stabilité d'un équilibre y^* à celle du système linéarisé en y^*, à savoir
$$h' = df(y^*)(h),$$
où $df(y^*) \in \mathcal{L}(\mathbb{R}^d)$ est la différentielle de f en y^* et peut être assimilée à une matrice de $\mathcal{M}_d(\mathbb{R})$.

Théorème 4.17. *Soit y^* un point stationnaire pour f. On pose*
$$\alpha = \max \Big\{ \mathrm{Re}(\lambda), \quad \lambda \in \mathrm{Sp}(df(y^*)) \Big\}.$$

1. *Si $\alpha < 0$ alors y^* est asymptotiquement stable.*
2. *Si $\alpha > 0$ alors y^* est instable.*

Autrement dit, dans le cas $\alpha \neq 0$, la stabilité/l'instabilité de y^* pour $y' = f(y)$ est équivalente à celle de 0 pour $h' = df(y^*)(h)$. Voir par exemple [BG10, paragraphe 8.2] pour une démonstration, ou [Dan, page 59] pour une preuve du premier point.

Remarque 4.18. Dans le cas $\alpha = 0$, on ne peut a priori pas conclure par cette méthode.

4.2.4 Fonctions de Lyapunov

Définition 4.19. Soit y^* un point d'équilibre pour le champ de vecteur f. Une fonction $L : U \to \mathbb{R}$ différentiable est appelée fonction de Lyapunov pour y^* si

1. $L(y^*) = 0$ et $\forall y \in U \setminus \{y^*\}$, $L(y) > 0$ [49]
2. pour tout $y \in U$,
$$\langle \nabla L(y), f(y) \rangle \leq 0. \tag{4.2}$$

On l'appelle fonction de Lyapunov stricte si la condition 2 est remplacée par

2' pour tout $y \in U \setminus \{y^*\}$,
$$\langle \nabla L(y), f(y) \rangle < 0.$$

Remarque 4.20. La condition (4.2) signifie que si (J, y) est une solution de $y' = f(y)$, alors la valeur de la fonction L le long de la trajectoire est décroissante, car
$$\forall t \in J, \quad (L \circ y)'(t) = \langle \nabla L(y(t)), y'(t) \rangle = \langle \nabla L(y(t)), f(y(t)) \rangle \leq 0.$$

Proposition 4.21 (Stabilité sous condition de Lyapunov)**.** *Soit f un champ de vecteur complet, y^* un point d'équilibre.*

1. *S'il existe une fonction de Lyapunov pour y^*, alors y^* est stable.*
2. *S'il existe une fonction de Lyapunov stricte pour y^*, alors y^* est asymptotiquement stable.*
3. *S'il existe une fonction de Lyapunov pour y^* telle que $\langle \nabla L(y), f(y) \rangle = 0$, alors y^* est stable et non asymptotiquement stable.*

Voir par exemple [Ton, Prop 10.3.2].

Exemple 4.22 (Systèmes gradients)**.** Si $f : U \to \mathbb{R}^d$ est de la forme $-\nabla V$, c'est-à-dire qu'on regarde l'EDO
$$y' = -\nabla V(y)$$
pour V de classe C^2, alors on a
$$\forall x \in U, \langle \nabla V(x), f(x) \rangle = -\|\nabla V(x)\|^2 \leq 0$$
donc si de plus y^* est un minimum local de V, alors c'est un point d'équilibre, qui est de plus stable.

49. La valeur $L(y^*) = 0$ n'est pas importante, ce qui compte c'est que L soit minimale en y^*. En effet on peut remplace L par $L - L(y^*)$.

Exemple 4.23 (Systèmes Hamiltoniens). Si $d = 2$ et qu'on considère $H : \mathbb{R}^2 \to \mathbb{R}$ de classe C^2, et l'équation

$$\left\{ \begin{array}{rcl} y_1(t) & = & \partial_2 H(y_1, y_2) \\ y_2(t) & = & -\partial_1 H(y_1, y_2) \end{array} \right. .$$

Alors

1. la valeur de H est constante le long des trajectoires,
2. si y^* est un minimum local strict de H, alors y^* est stable et non asymptotiquement stable.

4.3 Exemple : le système de Lotka-Volterra

Voir ce lien

Bibliographie

[ACL99] S. Fermigier A. Chambert-Loir. *Analyse 2, exercices. Agrégation de mathématiques.* DUNOD, 1999.

[AF89] J.M. Arnaudiès and H. Fraysse. *Cours de mathématiques, tome 3 - compléments d'analyse.* Dunod, 1989.

[All12] G. Allaire. *Analyse numérique et optimisation.* Editions de l'école Polytechnique, 2e édition, 2012.

[AM04] E. Amar and E. Matheron. *Analyse complexe.* Cassini, 2004.

[Amr08] M. El Amrani. *Analyse de Fourier dans les espaces fonctionnels, niveau M1.* Ellipses, 2008.

[App15] W. Appel. *Probabilités pour les non-probabilistes.* H&K, 2015.

[Ave83] A. Avez. *Calcul différentiel.* Masson, 1983.

[BB18] J. Bernis and L. Bernis. *Analyse pour l'agrégation, 40 développements.* Ellipses, 2018.

[BCD21] Q. Berger, F. Caravenna, and P. Dai Pra. *Introduction aux probabilités : Modèles et applications.* Dunod, Collection Sciences sup, 2021.

[BG10] S. Benzoni-Gavage. *Calcul différentiel et Équations différentielles.* Dunod, 2010.

[Bil12] P. Billingsley. *Probability and measure, anniversary ed.* Wiley, New York, 2012.

[BMP05] V. Beck, J. Malick, and G. Peyré. *Objectif agrégation.* H&K, 2005.

[Bon06] J-M. Bony. *Cours d'analyse - Théorie des distributions et analyse de Fourier.* Les éditions de l'école polytechnique, 2006.

[Boya] F. Boyer. Agrégation externe de mathématiques : analyse numérique. https://www.math.univ-toulouse.fr/~fboyer/_media/enseignements/agreg/cours_an_agreg_fboyer_2014.pdf. Visité : 2022-02-03.

[Boyb] F. Boyer. Equations différentielles ordinaires. https://www.math.univ-toulouse.fr/~fboyer/_media/enseignements/agreg/cours_edo_agreg_fboyer.pdf. Visité : 2020-02-28.

[BP12] M. Briane and G. Pagès. *Analyse : Théorie de l'intégration.* Vuibert, 2012.

[Bré05] H. Brézis. *Analyse fonctionnelle - Théorie et applications.* Dunod, 2005.

[Car61] H. Cartan. *Théorie élémentaire des fonctions analytiques.* Hermann, 1961.

[Car67] H. Cartan. *Calcul différentiel.* Hermann, 1967.

[CL97] A. Chambert-Loir. *Exercices de mathématiques pour l'agrégation : analyse - Tome 1 - 2ème édition.* Masson, 1997.

[Dan] R. Danchin. Equations différentielles l3 mathématiques. https://perso.math.u-pem.fr/danchin.raphael/cours/equadiff.pdf. Visité : 2020-03-08.

[Dan07] J.F. Dantzer. *Mathématiques pour l'agrégation interne ; Analyse & Probabilités.* Vuibert, 2007.

[Dem16] J-P. Demailly. *Analyse numérique et équations différentielles - 4e édition.* EDP SCIENCES, 2016.

[Dev89] R. L. Devaney. *An introduction to chaotic dynamical systems.* Addison-Wesley, 1989.

[DM72] H. Dym and H.P. McKean. *Fourier Series and Integrals.* Academic press New York and London, 1972.

[Fej] J. Fejoz. Intégration et probabilités. https://www.ceremade.dauphine.fr/~fejoz/Integration/integration-probabilites.pdf. Visité : 2019-11-12.

[FFF12] D. Foata, A. Fuchs, and J. Franchi. *Calcul des probabilités*. Dunod, Collection Sciences sup, 2012.

[FGN07a] S. Francinou, H. Gianella, and S. Nicolas. *Exercices de mathématiques oraux X-ENS, algèbre 1*. Cassini, 2007.

[FGN07b] S. Francinou, H. Gianella, and S. Nicolas. *Exercices de mathématiques oraux X-ENS, analyse 1 (deuxième édition, revue et augmentée)*. Cassini, 2007.

[FGN09] S. Francinou, H. Gianella, and S. Nicolas. *Exercices de mathématiques oraux X-ENS, analyse 2*. Cassini, 2009.

[Fro] A. Frouvelle. Méthodes numériques, optimisation. https://www.ceremade.dauphine.fr/~amic/enseignement/MNO2017/MNO2017.pdf. Visité : 2022-02-03.

[GH13] T. Gallouët and R. Herbin. *Mesure, intégration, probabilités*. Ellipses, 2013.

[GK19] O. Garet and A. Kurtzmann. *De l'intégration aux probabilités, 2eme édition augmentée*. Ellipses, 2019.

[God98] R. Godement. *Analyse Mathématique I*. Springer, 1998.

[God01] R. Godement. *Analyse Mathématique III*. Springer, 2001.

[God03] R. Godement. *Analyse Mathématique II, 2ième édition corrigée*. Springer, 2003.

[Gon] D. Gontier. Méthodes numériques, optimisation. https://www.ceremade.dauphine.fr/~gontier/Publications/methodesNumeriques.pdf. Visité : 2022-02-03.

[Gou98] X. Gourdon. *Les Maths en tête, Mathématiques pour M' : Analyse*. Ellipses, 1998.

[Gou08a] X. Gourdon. *Les Maths en tête : Algèbre, 2e édition*. Ellipses, 2008.

[Gou08b] X. Gourdon. *Les Maths en tête : Analyse, 2e édition*. Ellipses, 2008.

[GT98] S. Gonnord and N. Tosel. *Thémes d'analyse pour l'agrégation - Calcul différentiel*. Ellipses, 1998.

[HL99] F. Hirsch and G. Lacombe. *Éléments d'analyse fonctionnelle*. Dunod, 1999.

[HW15] J. Hubbard and B. West. *Équations différentielles et systèmes dynamiques*. Cassini, 2015.

[Kat04] Y. Katznelson. *An introduction to harmonic analysis*. Cambridge University Press - Third corrected edition, 2004.

[Laa07] E. Laamri. *Mesures, intégration, convolution et transformée de Fourier des fonctions*. Dunod, 2007.

[Laz] L. Lazzarini. Analyse fonctionnelle. Polycopié du cours d'agrégation 2021.

[LFA93] J. Lelong-Ferrand and J.M. Arnaudiès. *Cours de mathématiques, tome 2, 4e édition : Analyse*. Dunod, 1993.

[LM99] G. Lacombe and P. Massat. *Analyse fonctionnelle - Exercices corrigés*. Dunod, 1999.

[Mar02] J.P. Marco. *Analyse pour la licence, 2e édition*. Dunod, 2002.

[Mar09] J.P. Marco. *Mathématiques Analyse L3 : Cours complet avec 600 tests et exercices corrigés*. Pearson Education, 2009.

[Mat] E. Matheron. Topologie, analyse fonctionnelle. http://matheron.perso.math.cnrs.fr/enseignement_fichiers/topanafonc.pdf. Visité : 2019-10-07.

[Min] V. Minerbe. Analyse complexe. https://webusers.imj-prg.fr/uploads/vincent.minerbe/Analysecomplexe/3M266.pdf. Visité : 2019-08-29.

[MT86] R. Mneimé and F. Testard. *Théorie des groupes de Lie classiques*. Hermann Paris, 1986.

[Nou06] I. Nourdin. *Agrégation de mathématiques, épreuve orale*. Dunod, 2006.

[Ouv07] J-Y. Ouvrard. *Probabilités 1 - Licence Capes*. Cassini, 2007.

[Ouv09] J-Y. Ouvrard. *Probabilités 2 - Master Agrégation*. Cassini, 2009.

[Per] D. Perrin. La suite logistique et le chaos. https://www.math.u-psud.fr/~perrin/Conferences/logistiqueDP.pdf. Visité : 2019-08-16.

[Pom94] A. Pommellet. *Cours d'analyse : Agrégation de mathématiques.* Ellipses, 1994.

[QQ17] M. Quéffelec and H. Quéffelec. *Analyse complexe et applications.* Calvage & Mounet, 2017.

[QZ13] H. Queffelec and C. Zuily. *Analyse pour l'agrégation - 4ème édition.* Dunod, 2013.

[RDO91] E. Ramis, C. Deschamps, and J. Odoux. *Cours de mathématiques spéciales, tome 3, Topologie et éléments d'analyse, 3e édition.* Masson, 1991.

[Rev97] D. Revuz. *Mesure et intégration.* Hermann, 1997.

[Rou03] F. Rouvière. *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation (deuxième édition, revue et augmentée).* Cassini, 2003.

[Rud95] W. Rudin. *Analyse fonctionelle.* Ediscience, 1995.

[Rud98] W. Rudin. *Analyse réelle et complexe, Cours et exercices.* Dunod, 1998.

[Sch97] L. Schwartz. *Théorie des distributions.* Hermann, 1997.

[Sch04] M. Schatzman. *Analyse numérique, une approche mathématique.* DUNOD, 2004.

[Ska04] G. Skandalis. *Topologie et analyse 3e année.* Dunod, 2004.

[SR08] J. Saint-Raymond. *Topologie, calcul differentiel et variable complexe, Nouvelle édition.* Calvage & Mounet, 2008.

[Tau06] P. Tauvel. *Analyse complexe pour la Licence 3.* Dunod, 2006.

[Ton] D. Tonon. Differential equations. https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxtYXRoZGFuaWVsYXRvbm9ufGd4OjFiNTExOTliYmM3MjVhZjU. Visité : 2020-02-28.

[Wag12a] C. Wagschal. *Dérivation, intégration.* Hermann, 2012.

[Wag12b] C. Wagschal. *Topologie et analyse fonctionnelle.* Hermann, 2012.

[Wil95] M. Willem. *Analyse harmonique réelle.* Hermann, 1995.

[Zav13] M. Zavidovique. *Un max de maths : Problèmes pour agrégatifs et mathématiciens en herbe ou confirmés.* Calvage & Mounet, 2013.

[Zui02] C. Zuily. *Éléments de distributions et d'équations aux dérivées partielles.* Dunod, 2002.

[Zyg03] A. Zygmund. *Trigonometric series.* Cambridge Mathematical Library - Third edition (volumes I & II combined), 2003.