

**ALGÈBRE ET
ANALYSE APPROFONDIES I**

MT241

B. Maurey

Année 2002-2003

Version remaniée pour MM3 2009-2010

Sommaire

Chapitre 1. Nombres réels. Suites numériques	1
1.1. Nombres réels	1
1.2. Limite des suites numériques	5
1.3. Suites de Cauchy	9
1.4. Topologie de \mathbb{R}	11
Chapitre 2. Séries numériques	17
2.1. Introduction	17
2.2. Séries convergentes à termes réels ou complexes	18
2.3. Séries à termes réels positifs	20
2.4. Comparaison séries–intégrales	23
2.5. Séries absolument convergentes	25
2.6. Séries semi-convergentes. Séries alternées	29
Chapitre 3. Intégrale de Riemann	33
3.1. Intégrale de Riemann sur un intervalle fermé borné $[a, b]$	33
3.2. Le cas vectoriel	42
3.3. Calculs approchés des intégrales simples	45
Chapitre 4. Rappels d’algèbre linéaire	49
4.1. Espaces vectoriels	49
4.2. Dimension	52
4.3. Sommes directes	54
4.4. Déterminants	59
4.5. Formes linéaires. Espace dual	66
Chapitre 5. Réduction des endomorphismes d’un espace vectoriel	71
5.1. Polynôme caractéristique ; vecteurs propres et valeurs propres	71
5.2. Sous-espaces propres d’un endomorphisme	78
5.3. Polynômes d’endomorphismes	84
5.4. Idéaux de polynômes. Polynôme minimal	88
Chapitre 6. Équations différentielles linéaires	95
6.1. Équations différentielles linéaires à coefficients constants	96
6.2. Systèmes différentiels linéaires	97
Index alphabétique	105
Index des notations	109

Chapitre 1. Nombres réels. Suites numériques

Ce chapitre est consacré à des rappels de première année et à quelques compléments sur les nombres réels, les suites réelles et complexes. Les ensembles de nombres seront désignés par les notations classiques : \mathbb{N} , \mathbb{Z} , \mathbb{Q} , \mathbb{R} , \mathbb{C} .

1.1. Nombres réels

Axiomes pour les nombres réels

Il y a de nombreuses façons de présenter les nombres réels en donnant un ensemble restreint de propriétés qui soit suffisant pour reconstruire toutes les autres propriétés des nombres réels. Certaines présentations insistent plutôt sur les opérations algébriques d'addition et de multiplication, et placent l'ordre au second plan. Nous choisirons ici de privilégier la notion d'ordre, et de nous appuyer sur la connaissance des propriétés de l'ensemble \mathbb{Q} des nombres rationnels.

On suppose donc connu l'ensemble \mathbb{Q} , avec son addition, sa multiplication, son ordre et les relations entre l'ordre et les opérations. L'ensemble \mathbb{R} des *nombres réels* peut alors être caractérisé à partir d'un petit nombre de propriétés « axiomatiques ». Ces propriétés seront classées ici en quatre points : (O) pour *ordre*, (E) pour *extension* des rationnels, (D) pour *densité* (des rationnels dans les réels) et (C) pour *coupures*.

(O) L'ensemble \mathbb{R} est totalement ordonné (c'est-à-dire qu'étant donnés $x, y \in \mathbb{R}$, on a $x \leq y$ ou bien $y \leq x$). L'ensemble \mathbb{R} n'a pas de plus grand élément et pas de plus petit élément.

(E) L'ensemble \mathbb{R} contient l'ensemble \mathbb{Q} des rationnels, et l'ordre sur \mathbb{R} prolonge l'ordre des rationnels : si q_1 et q_2 sont deux nombres rationnels, on a $q_1 \leq q_2$ pour l'ordre de \mathbb{R} si et seulement si on a $q_1 \leq q_2$ dans \mathbb{Q} .

(D) Étant donnés deux nombres réels x, y tels que $x < y$, il existe un nombre rationnel q tel que $x < q < y$.

(C) Pour toute partition de \mathbb{Q} en deux classes *non vides* I et J telles que

$$\forall q \in I, \forall r \in J, \quad q < r$$

(c'est-à-dire que tous les éléments de I sont plus petits que les éléments de J), il existe un nombre réel z tel que

$$\forall q \in I, \forall r \in J, \quad q \leq z \leq r.$$

La dernière propriété (C) s'appelle la *propriété des coupures*, parce qu'on appelle *coupure dans \mathbb{Q}* toute partition (I, J) de \mathbb{Q} vérifiant les propriétés énoncées dans (C). Le nombre z introduit dans la propriété (C) est unique (exercice) ; on dira que z est défini par la coupure (I, J). C'est l'unique nombre réel qui se trouve « coincé » entre les ensembles I et J. Si q, r sont deux nombres rationnels et si on a $q < z < r$, alors $q \in I$ et $r \in J$. Si z n'est pas rationnel, on peut dire que I est l'ensemble de tous les nombres rationnels q tels que $q < z$ et J l'ensemble de tous les nombres rationnels r tels que $r > z$; si z est lui-même rationnel, il y a deux cas possibles, selon que $z \in I$ ou $z \in J$.

Exemple. Désignons par J l'ensemble de tous les nombres rationnels $r > 0$ tels que $r^2 > 2$ et par I l'ensemble $\mathbb{Q} \setminus J$ complémentaire de J dans \mathbb{Q} . Le couple (I, J) est une coupure qui définit le nombre réel $\sqrt{2}$ (si A et B sont deux sous-ensembles d'un ensemble X, on note $A \setminus B$ l'ensemble des points de A qui ne sont pas dans B).

Proposition 1.1.1. Propriété des coupures. Pour toute partition de \mathbb{R} en deux classes non vides I et J telles que

$$\forall x \in I, \forall y \in J, \quad x < y,$$

il existe un nombre réel z unique tel que

$$\forall x \in I, \forall y \in J, \quad x \leq z \leq y.$$

Esquisse de démonstration. C'est presque la même chose que la propriété (C), sauf qu'on s'est donné maintenant une partition de \mathbb{R} au lieu d'une partition de \mathbb{Q} . On désigne alors par I' l'ensemble des rationnels qui sont dans I et par J' l'ensemble des rationnels qui sont dans J. On vérifie que (I', J') est une coupure de \mathbb{Q} (micro-exercice), qui définit un nombre réel z : c'est le nombre cherché.

Exercice. (A traiter à partir des propriétés (O, E, D, C) seulement, en oubliant momentanément ce qu'on a pu apprendre avant sur l'ensemble \mathbb{R})

– Montrer que pour tout nombre réel x il existe deux nombres rationnels p, q tels que $p < x < q$.

– Si x, y sont deux nombres réels, montrer que l'on a $x \leq y$ si et seulement si tout nombre rationnel q plus petit que x est plus petit que y .

– Montrer qu'un nombre réel x est égal à 0 si et seulement si pour tout rationnel $r > 0$, on a $-r \leq x \leq r$.

– Soit f une fonction croissante de \mathbb{R} dans \mathbb{R} , qui change de signe mais ne s'annule jamais ; montrer qu'il existe un nombre $z \in \mathbb{R}$ unique tel que $f(x) < 0$ pour tout $x < z$ et $f(y) > 0$ pour tout $y > z$.

Remarque. Approximation d'un nombre réel par des nombres rationnels. Si x est un nombre réel et si r est un nombre rationnel > 0 (qu'on peut imaginer *petit*), il existe deux nombres rationnels p et q tels que $p < x < q$ et $q - p < r$.

Justifions cette affirmation. Prenons un premier rationnel q_1 tel que $x < q_1$, et un autre rationnel u tel que $0 < u < r/2$. Il existe un entier $n > 0$ tel que $q_1 < nu$, donc $x < nu$, et de la même façon on peut trouver un entier relatif m tel que $mu < x$. L'ensemble des entiers relatifs $k \geq m$ tels que $ku \leq x$ est fini, puisqu'il est majoré par n et minoré par m . Soit k_0 son plus grand élément ; on aura alors $(k_0 + 1)u > x$, donc

$$p = k_0 u - u < k_0 u \leq x < (k_0 + 1)u = q$$

et on a bien obtenu $q - p = 2u < r$.

Majorant. Borne supérieure

Soit A un sous-ensemble de \mathbb{R} ; on dit qu'un nombre réel M est un *majorant* de l'ensemble A si tout élément x de A est inférieur ou égal à M,

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

Il est important (et facile) de savoir exprimer qu'un nombre y n'est pas un majorant de l'ensemble A : cela signifie qu'il existe au moins un nombre $x \in A$ tel que $x > y$. Si A est vide, on est conduit à dire que n'importe quel $M \in \mathbb{R}$ est un majorant de A. On dit que l'ensemble A est *majoré* s'il existe un majorant $M \in \mathbb{R}$ pour l'ensemble A. Un *minorant* de A est défini de façon analogue, et on dit que A est *minoré* s'il possède un minorant $m \in \mathbb{R}$. On dit que A est *borné* s'il est à la fois majoré et minoré.

Si A est un sous-ensemble non vide et majoré de \mathbb{R} , considérons l'ensemble J des majorants de A et désignons par I son complémentaire. Alors (I, J) est une coupure dans \mathbb{R} : tout d'abord, J est non vide par hypothèse. Puisque A est non vide, on peut trouver $a \in A$ et un réel x tel que $x < a$ parce que \mathbb{R} n'a pas de plus petit élément, donc x n'est pas un majorant de A , c'est-à-dire que $x \in I$, donc I est non vide. Soient ensuite $x \in I$ et $y \in J$; puisque x n'est pas un majorant de A , il existe un élément $a \in A$ tel que $x < a$, et $a \leq y$ puisque y est un majorant de A , donc $x < y$. On a bien montré que (I, J) est une coupure.

Puisque (I, J) est une coupure, il existe un unique nombre réel b défini par la coupure, c'est-à-dire tel que $x \leq b \leq y$ pour tous $x \in I, y \in J$. Nous allons voir que

Le nombre b est le plus petit majorant de A .

Tout d'abord, on sait que $b \leq y$ pour tout majorant y de A (c'est-à-dire tout $y \in J$). De plus, b est un majorant de A : sinon, il existerait $a \in A$ tel que $b < a$, puis un rationnel q tel que $b < q < a$ d'après la densité (D) ; alors q ne serait pas un majorant de A , donc $q \in I$ mais $b < q$, contredisant la définition de b par la coupure (I, J) . Par conséquent, b est un majorant de A , donc finalement b est le plus petit majorant de A .

Définition 1.1.1. On dit que b est la *borne supérieure* d'un sous-ensemble A de \mathbb{R} si b est le plus petit majorant de A .

On a donc établi une propriété fondamentale de l'ensemble \mathbb{R} :

Théorème 1.1.1. *Tout sous-ensemble A non vide et majoré de \mathbb{R} possède une borne supérieure.*

Si A possède un plus grand élément a , cet élément a est évidemment la borne supérieure de A . La *borne inférieure* d'un ensemble A , si elle existe, est le plus grand minorant de A . Tout sous-ensemble non vide et minoré de \mathbb{R} possède une borne inférieure. Quand elle existe, on désigne la borne supérieure de A par la notation $\sup A$, et $\inf A$ pour la borne inférieure.

Exemples. L'intervalle $A =]0, 1[$ admet 1 pour borne supérieure et 0 pour borne inférieure. Noter qu'aucun de ces deux nombres n'est élément de A dans cet exemple. L'ensemble $\{1, 1/2, 1/3, \dots, 1/n, \dots\}$ admet 1 pour plus grand élément, donc pour borne supérieure, et 0 pour borne inférieure.

Il est important de comprendre le fonctionnement de ces notions. Commençons par une propriété très simple.

Si $A \subset B \subset \mathbb{R}$, avec A non vide, B majoré, on a $\sup A \leq \sup B$. Dans la même situation $A \subset B \subset \mathbb{R}$, si A est non vide et B minoré, on a $\inf B \leq \inf A$.

Vérification. Le nombre $\sup B$ est par définition un majorant de B , donc c'est aussi un majorant de A puisque $A \subset B$. Mais puisque $\sup A$ est le *plus petit* majorant de A , on a bien $\sup A \leq \sup B$.

On va maintenant expliquer comment on peut définir les opérations algébriques sur les nombres réels en utilisant l'addition et la multiplication des rationnels et les propriétés d'ordre. Précisons la règle du jeu pour la lecture de ce qui suit : on va faire semblant pour un moment de ne plus savoir ce qu'est l'addition de deux nombres réels, et on va voir comment on pourrait la définir à partir de ce qui précède. La réalisation complète de ce programme est longue et pénible, et on ne fera ici que l'esquisser.

Addition des nombres réels

Étant donnés deux nombres réels x_1 et x_2 , on définit un nombre réel, noté provisoirement $[x_1 + x_2]$, par la formule

$$[x_1 + x_2] = \sup\{p_1 + p_2 : p_1, p_2 \in \mathbb{Q}, p_1 \leq x_1, p_2 \leq x_2\}.$$

Désignons par $S(x_1, x_2)$ l'ensemble des sommes $p_1 + p_2$, où p_1, p_2 sont rationnels, et $p_1 \leq x_1, p_2 \leq x_2$. Si x_1 et x_2 sont rationnels, on voit que $[x_1 + x_2] = x_1 + x_2$. En effet, dans ce cas, $x_1 + x_2$ est le plus grand élément de $S(x_1, x_2)$, donc aussi sa borne supérieure. Cette vérification permet de noter simplement $x_1 + x_2$ pour la somme de deux réels.

Exercice (pénible et un peu délicat).

- Vérifier que $x_1 + x_2 = x_2 + x_1$ pour tous réels x_1, x_2 .
- Si $x_1 \leq y_1$ et $x_2 \leq y_2$, montrer que $x_1 + x_2 \leq y_1 + y_2$.
- Montrer que $x_1 + (x_2 + x_3) = (x_1 + x_2) + x_3$ pour tous réels x_1, x_2, x_3 .
- Montrer que $x + 0 = x$ pour tout nombre réel x .

Pour terminer la mise en place des propriétés liées à l'addition, on définit un nombre réel noté $[-x]$ comme borne supérieure de l'ensemble $N(x)$ des rationnels q tels que $x \leq -q$. On montre alors que $[-x] + x = 0$. Cela permet de vérifier que si x est rationnel, $[-x]$ est bien égal au rationnel $-x$. Cette vérification étant faite, on notera simplement $-x$ pour l'opposé d'un nombre réel x . On voit que si $x \leq y$, alors $-y \leq -x$.

Multiplication

Le cas de la multiplication est plus embêtant, parce que la multiplication par un nombre réel < 0 va renverser l'ordre. On peut y arriver en définissant d'abord le produit de deux nombres réels ≥ 0 , comme borne supérieure d'un ensemble de produits de rationnels, puis en définissant les autres cas en raisonnant sur les signes. On commence donc par poser pour x_1, x_2 réels positifs ou nuls

$$[x_1 x_2] = \sup\{p_1 p_2 : p_1, p_2 \in \mathbb{Q}, 0 \leq p_1 \leq x_1, 0 \leq p_2 \leq x_2\}.$$

On vérifie que $[x_1 x_2] = x_1 x_2$ lorsque x_1 et x_2 sont rationnels ≥ 0 . On omettra ici la suite de la construction de la multiplication, qui commence ainsi : si $x_1 < 0$ et $x_2 \geq 0$, on posera $x_1 x_2 = -(-x_1)x_2$...

Il faut se rappeler que si x, y sont ≥ 0 , alors $xy \geq 0$. Il en résulte que si $a \leq b$ et $u \geq 0$, on a $au \leq bu$ (il suffit d'écrire $bu - au = (b - a)u \geq 0$).

Une construction des nombres réels

On pourra sans inconvénient sauter ce paragraphe en première lecture. L'objectif est d'expliquer comment on peut fabriquer un « modèle » pour l'ensemble \mathbb{R} en utilisant le langage de la théorie des ensembles, et en ne supposant connu que l'ensemble \mathbb{Q} des nombres rationnels. Ce modèle sera noté provisoirement \mathcal{R} au lieu de \mathbb{R} .

L'idée est assez simple. Si on « connaît » déjà l'ensemble \mathbb{R} , on voit bien qu'un nombre réel x est complètement déterminé par la connaissance de l'ensemble I_x de tous les nombres rationnels q tels que $q < x$. Si on oublie \mathbb{R} et qu'on cherche à l'introduire « proprement » à partir de \mathbb{Q} , on peut se dire qu'il suffit de remplacer « l'idée » de x par « l'idée » de I_x , c'est-à-dire l'idée d'un intervalle I de rationnels, majoré mais sans plus grand élément, et illimité à gauche.

On désignera donc par \mathcal{R} l'ensemble des sous-ensembles $I \subset \mathbb{Q}$ qui vérifient les trois propriétés suivantes :

1. $I \neq \emptyset$ et $I \neq \mathbb{Q}$
2. Si I contient un nombre rationnel r , alors I contient tous les nombres rationnels q tels que $q \leq r$.
3. Pour tout $r \in I$ il existe $r' \in I$ tel que $r < r'$ (l'ensemble I n'a pas de plus grand élément).

Du point de vue ensembliste, \mathcal{R} est un sous-ensemble de l'ensemble $\mathcal{P}(\mathbb{Q})$ des parties de \mathbb{Q} . Pour tout nombre rationnel q , posons

$$I(q) = \{r \in \mathbb{Q} : r < q\}.$$

On vérifie que $I(q)$ est un élément de \mathcal{R} . On a ainsi défini une application $j : q \rightarrow I(q)$ de \mathbb{Q} dans \mathcal{R} . Cette application j est injective, ce qui permettra de « plonger » \mathbb{Q} dans \mathcal{R} . Si on n'aime

pas « plonger », on pourra dire que \mathbb{R} sera la réunion de \mathbb{Q} (le vrai) avec la partie $\mathcal{R} \setminus j(\mathbb{Q})$ de \mathcal{R} formée de tous les $I \in \mathcal{R}$ qui ne sont pas de la forme $I(q)$ pour un rationnel q .

On définit maintenant l'ordre sur l'ensemble \mathcal{R} , en prenant l'ordre de l'inclusion des ensembles de rationnels : si $I_1, I_2 \in \mathcal{R}$, on dit que $I_1 \leq I_2$ si et seulement si $I_1 \subset I_2$. On vérifie que l'ensemble \mathcal{R} est totalement ordonné. De plus, l'ordre sur l'image $j(\mathbb{Q})$ de \mathbb{Q} est bien l'ordre de \mathbb{Q} , c'est-à-dire que l'on a $p_1 \leq p_2$ dans \mathbb{Q} si et seulement si $I(p_1) \subset I(p_2)$. Il est clair que \mathcal{R} n'a pas de plus grand élément ni de plus petit élément. On a donc vérifié les propriétés (O) et (E) pour l'ensemble \mathcal{R} .

Supposons que I_1, I_2 soient deux éléments de \mathcal{R} , tels que $I_1 < I_2$. Cela signifie qu'il existe un rationnel q tel que $q \in I_2$ mais $q \notin I_1$. Il existe aussi d'après la définition des éléments de \mathcal{R} un nombre rationnel $q' > q$ tel que $q' \in I_2$. Alors $I_1 \leq I(q) < I(q') < I_2$, ce qui montre la propriété (D).

Terminons avec la propriété (C) de coupure. Soit (I, J) une partition de \mathbb{Q} en deux classes non-vides telle que tout élément q de I et tout élément r de J vérifient $q < r$; désignons par I' l'ensemble de tous les éléments $q \in I$ pour lesquels il existe $q' \in I$ tel que $q < q'$. On vérifie alors que $I' \in \mathcal{R}$ et que $I(q) \subset I' \subset I(r)$ pour tous $q \in I, r \in J$. Autrement dit, $I' = z$ est l'élément cherché dans la propriété (C).

La droite achevée $\overline{\mathbb{R}}$

La *droite achevée*, notée $\overline{\mathbb{R}}$, est l'ensemble obtenu en ajoutant à l'ensemble \mathbb{R} deux nouveaux éléments, $+\infty$ et $-\infty$. Pour bien se mettre d'accord sur le langage employé, insistons un peu : ces deux nouveaux éléments $+\infty$ et $-\infty$ *ne sont pas* des nombres réels et *ne devront pas* être appelés nombres réels.

Il est utile de préciser un certain nombre de conventions concernant les symboles $+\infty$ et $-\infty$. Tout d'abord, si x est réel, on considère que $-\infty < x < +\infty$; on pose ensuite $x + (+\infty) = x - (-\infty) = +\infty$, $x - (+\infty) = x + (-\infty) = -\infty$; on pose aussi $(+\infty) + (+\infty) = +\infty$ et la convention analogue pour $-\infty$, à savoir $(-\infty) + (-\infty) = -\infty$.

Si $x > 0$, on pose $x \cdot (+\infty) = (-x) \cdot (-\infty) = +\infty$ et $x \cdot (-\infty) = (-x) \cdot (+\infty) = -\infty$; on pose aussi $(+\infty) \cdot (+\infty) = (-\infty) \cdot (-\infty) = +\infty$ et $1/(\infty) = 1/(-\infty) = 0$; les autres cas sont considérés indéterminés : $(+\infty) - (+\infty)$, $0 \cdot (+\infty)$ ou $0 \cdot (-\infty)$ et $1/0$.

On pourra noter que *tout sous-ensemble A de $\overline{\mathbb{R}}$ possède une borne supérieure dans $\overline{\mathbb{R}}$* (et une borne inférieure) : si A est vide, son plus petit majorant est l'élément $-\infty$; si A n'est pas majoré par un nombre réel, son plus petit majorant dans $\overline{\mathbb{R}}$ est $+\infty$. Si A est non vide et majoré par un nombre réel, on utilise la propriété de borne supérieure dans \mathbb{R} .

1.2. Limite des suites numériques

Une suite de nombres réels est une application $x : \mathbb{N} \rightarrow \mathbb{R}$; la notation habituelle pour une suite est (x_n) , ou plus précisément $(x_n)_{n \geq 0}$, plutôt que les notations d'applications telles que $n \rightarrow x(n)$.

Suite de nombres réels convergente vers un nombre réel

Définition 1.2.1. On dit que la suite de nombres réels (x_n) converge vers $\ell \in \mathbb{R}$ si pour tout $\varepsilon > 0$, on a pour tout entier n assez grand (dépendant de ε) l'inégalité $|x_n - \ell| < \varepsilon$. En symboles

$$\forall \varepsilon > 0, \exists N \in \mathbb{N}, \forall n \in \mathbb{N}, (n \geq N \Rightarrow |x_n - \ell| < \varepsilon).$$

On peut traduire la convergence de la suite (x_n) vers la limite ℓ en disant qu'il existe une application $\varepsilon \rightarrow N(\varepsilon) \in \mathbb{N}$, définie pour tout $\varepsilon > 0$, qui *contrôle la convergence* dans le sens suivant : pour tout $\varepsilon > 0$, on a

$$\forall n \in \mathbb{N}, (n \geq N(\varepsilon) \Rightarrow |x_n - \ell| < \varepsilon).$$

Par exemple, la convergence vers 0 de la suite $(1/n)$ est contrôlée par la fonction $N(\varepsilon) = [1/\varepsilon] + 1$ (la notation $[x]$ désigne la *partie entière* du nombre réel x , c'est-à-dire le plus grand entier relatif k tel que $k \leq x$). Si la fonction $\varepsilon \rightarrow N_1(\varepsilon)$ contrôle la convergence de la suite (x_n) vers ℓ et la fonction N_2 la convergence de la suite (y_n) vers m , on voit facilement que si on pose $N(\varepsilon) = \max\{N_1(\varepsilon/2), N_2(\varepsilon/2)\}$, la fonction $\varepsilon \rightarrow N(\varepsilon)$ contrôle la convergence de la suite $(x_n + y_n)$ vers $\ell + m$: c'est le théorème sur la somme de limites.

Le lecteur devra réviser les propriétés des limites : sommes, produits, quotients,...

Exercice. Étudier la suite définie par $x_0 = 1$, et $x_{n+1} = \frac{1}{2}(x_n + 5/x_n)$ pour tout $n \geq 0$ (on calculera $x_{n+1} \pm \sqrt{5}$, et on montrera que $1 - 2\sqrt{5}/(x_n + \sqrt{5}) = (x_n - \sqrt{5})/(x_n + \sqrt{5})$ tend vers 0).

Limite infinie

On dit que la suite de nombres réels (x_n) tend vers $+\infty$ si pour tout $A \in \mathbb{R}$ on a pour tout entier n assez grand (dépendant de A) l'inégalité $x_n > A$. En symboles

$$\forall A \in \mathbb{R}, \exists N \in \mathbb{N}, \forall n \in \mathbb{N}, (n \geq N \Rightarrow x_n > A).$$

Il existe évidemment une définition analogue pour $-\infty$ que le lecteur formulera lui-même. On réserve le terme de *suite convergente* aux suites qui convergent vers un nombre réel (c'est-à-dire que la limite est finie). Dans tous les cas la limite est unique quand elle existe, et on la note $\lim_n x_n$, finie ou $\pm\infty$.

Si (x_n) et (y_n) sont deux suites qui admettent les limites (peut-être infinies) ℓ et m , les suites $(x_n + y_n)$, $(x_n y_n)$ ou $(1/x_n)$ tendent respectivement vers $\ell + m$, ℓm ou $1/\ell$, chaque fois que les conventions naturelles pour l'arithmétique de $\overline{\mathbb{R}}$ (rappelées à la fin de la section précédente) donnent un sens à l'expression limite.

Dans certains cas une suite numérique ne commence pas avec $n = 0$, mais elle est seulement définie pour $n \geq n_0$; la notation $(x_n)_{n \geq n_0}$ rend compte de cette situation. Les modifications à faire dans ce cas à la définition des limites sont claires. La limite d'une suite ne change pas si on modifie un nombre fini de termes de la suite, ou si on décale tous les termes : si $\lim_n x_n = \ell \in \overline{\mathbb{R}}$, la suite (y_n) définie par $y_n = x_{n+1}$ vérifie aussi $\lim_n y_n = \ell$.

Principes simples

La propriété suivante est importante, c'est pratiquement la définition de la limite :

Si $a < \ell = \lim_n x_n \in \overline{\mathbb{R}}$, alors pour tout entier n assez grand on a l'inégalité $a < x_n$.

De même si $b > \ell = \lim_n x_n \in \overline{\mathbb{R}}$ on a $b > x_n$ pour tout entier n assez grand.

Vérification dans le cas $a < \ell$ et ℓ finie. Dans l'application de la définition de la limite, on choisit $\varepsilon = (\ell - a)/2 > 0$. Pour $n \geq N(\varepsilon)$, on aura $|x_n - \ell| < (\ell - a)/2$, ce qui donne $\ell - x_n < \ell - a$, donc $a < x_n$ pour tout $n \geq N(\varepsilon)$.

Exemple : suites équivalentes ; soit $(y_n)_{n \geq 0}$ une suite numérique telle que $y_n \neq 0$ pour tout n assez grand ; la suite (x_n) est dite équivalente à la suite (y_n) si x_n/y_n tend vers 1 quand $n \rightarrow +\infty$; alors pour tout $\varepsilon > 0$ on a pour n assez grand $(1 - \varepsilon) < x_n/y_n < (1 + \varepsilon)$. En particulier, on voit que $x_n/y_n > 0$ pour n assez grand (prendre $\varepsilon = 1/2$ par exemple), ce qui indique que x_n et y_n sont de même signe pour n assez grand.

Toute suite convergente (limite finie) est bornée.

Cela résulte du « principe simple » ci-dessus, appliqué avec par exemple $a = \ell - 1$ et $b = \ell + 1$.

Passage à la limite des inégalités larges

Il s'agit d'une remarque bien simple mais qui doit devenir un des automatismes fondamentaux de l'analyse : si $\ell = \lim x_n$ et si on a $x_n \leq M$ pour tout $n \geq n_0$, alors « l'inégalité passe à la limite » et on obtient $\ell \leq M$.

Important : seules les inégalités larges se conservent par passage à la limite.

Limites de fonctions et limites de suites

Soient $J =]c, d[$ un intervalle ouvert de \mathbb{R} et a un point de $[c, d]$; posons $I = J \setminus \{a\}$; si f est une fonction réelle définie dans I , qui vérifie

$$\lim_{x \rightarrow a, x \in I} f(x) = \ell$$

(avec $\ell \in \overline{\mathbb{R}}$) et si (x_n) est une suite de points de I telle que $x_n \rightarrow a$, alors la suite $(f(x_n))$ tend vers ℓ . Le résultat s'étend au cas où $d = a = +\infty$ ou bien $c = a = -\infty$.

Exemple : soit t un nombre réel fixé ; déterminer $\lim_n (1 + \frac{t}{n})^n$ en utilisant la limite du quotient $\ln(1+x)/x$ lorsque $x \rightarrow 0$ (le seul cas intéressant est le cas $t \neq 0$).

En particulier, ce qui précède s'applique si f est une fonction continue en un point. Soient J un intervalle ouvert de \mathbb{R} et a un point de J ; si f est une fonction réelle définie sur J , continue au point a , et si (x_n) est une suite de points de J telle que $x_n \rightarrow a$, alors la suite $(f(x_n))$ tend vers $f(a)$.

Suites croissantes et majorées

On rappelle que la borne supérieure d'un sous-ensemble A de \mathbb{R} est le plus petit majorant de A , et que tout sous-ensemble non vide et majoré de \mathbb{R} possède une borne supérieure, notée $\sup A$.

Théorème 1.2.1. *Toute suite croissante et majorée (x_n) de nombres réels converge vers un nombre $\ell \in \mathbb{R}$. La limite ℓ est la borne supérieure de l'ensemble*

$$\{x_0, x_1, \dots, x_n, \dots\}$$

des valeurs prises par la suite (x_n) . En particulier, on a $x_n \leq \ell$ pour tout n .

Autrement dit, une suite croissante de nombres réels est convergente si et seulement si elle est majorée.

Démonstration. L'ensemble A des valeurs prises par la suite (x_n) est non vide, et il est majoré par hypothèse. Nous pouvons donc considérer la borne supérieure $\ell = \sup(A)$. Soit $\varepsilon > 0$ quelconque ; puisque $\ell - \varepsilon < \ell$ et puisque ℓ est le plus petit majorant de A , nous savons que $\ell - \varepsilon$ ne peut pas être un majorant de A . Il existe donc au moins un élément x_N de A tel que $\ell - \varepsilon < x_N$ (l'entier N dépend de ε). Puisque la suite (x_n) est croissante, nous aurons $x_N \leq x_n$ pour tout $n \geq N$; par ailleurs $x_n \leq \ell$ puisque ℓ est un majorant de l'ensemble A . Au total nous avons pour tout entier $n \geq N$

$$\ell - \varepsilon < x_N \leq x_n \leq \ell,$$

ce qui entraîne que $|x_n - \ell| < \varepsilon$. Explicitons complètement la conclusion, pour une fois : nous venons bien de montrer qu'à tout $\varepsilon > 0$ donné, nous pouvons associer un entier $N = N(\varepsilon)$ tel que $|x_n - \ell| < \varepsilon$ pour tout $n \geq N$, donc nous avons montré que la suite (x_n) converge vers ℓ .

À titre d'exercice, on montrera de façon analogue le théorème qui suit :

Théorème 1.2.2. Soit f une fonction à valeurs réelles définie sur un intervalle $[a, b[$, croissante et majorée sur cet intervalle ; alors $f(x)$ tend vers une limite $\ell \in \mathbb{R}$ lorsque x tend vers b (par valeurs $x < b$). La limite ℓ est la borne supérieure de l'ensemble des valeurs prises par la fonction f sur l'intervalle $[a, b[$. En particulier, on a $f(x) \leq \ell$ pour tout $x \in [a, b[$. Le résultat reste valable pour un intervalle de la forme $[a, +\infty[$.

Convergence de suites complexes

On dit qu'une suite (z_n) de nombres complexes converge vers $\ell \in \mathbb{C}$ si la suite réelle des modules $|z_n - \ell|$ tend vers 0. On peut exprimer la convergence en utilisant les parties réelles et imaginaires : la suite (z_n) converge vers $\ell \in \mathbb{C}$ si et seulement si les deux suites réelles formées des parties réelles et imaginaires convergent respectivement vers celles de la limite ℓ ,

$$\lim_{n \rightarrow +\infty} z_n = \ell \Leftrightarrow \left(\lim_{n \rightarrow +\infty} \operatorname{Re} z_n = \operatorname{Re} \ell \text{ et } \lim_{n \rightarrow +\infty} \operatorname{Im} z_n = \operatorname{Im} \ell \right).$$

Il est parfois (et même souvent) nécessaire d'étudier une suite complexe directement comme suite complexe, alors que l'étude des parties réelles et imaginaires conduirait à des complications insurmontables. La même remarque s'applique à certains calculs matriciels ou vectoriels qui ne sont compréhensibles que si on raisonne globalement.

Reprendre l'exemple : $z_0 \in \mathbb{C}$, $z_{n+1} = \frac{1}{2}(z_n + b/z_n)$, avec b complexe, $b \neq 0$; discuter suivant la position de z_0 par rapport aux deux racines carrées de b .

Sous-suites

La notion de sous-suite est source de malentendus, et il faut se mettre d'accord sur le sens qu'on veut lui donner en mathématique. Étant donnée une suite (x_n) de nombres (ou une suite de n'importe quoi, ça serait pareil), la donnée d'une sous-suite consiste à sélectionner **dans l'ordre donné** et **sans répétition** certains éléments de la suite initiale, en conservant **une infinité** de termes de la suite donnée. Ce qui nous donne :

Définition 1.2.2. On dit que la suite $(y_k)_{k \geq 0}$ est une *sous-suite* de la suite $(x_n)_{n \geq 0}$ s'il existe une suite strictement croissante d'indices entiers $n_0 < n_1 < \dots < n_k < \dots$ tels que pour tout entier $k \geq 0$ on ait $y_k = x_{n_k}$.

Dans la pratique, on notera souvent directement $(x_{n_k})_k$ une sous-suite de la suite (x_n) . Si la suite (x_n) tend vers $\ell \in \overline{\mathbb{R}}$, toute sous-suite de la suite (x_n) tend vers la même limite ℓ ; bien sûr l'inverse n'est pas vrai : la suite $(x_n) = ((-1)^n)$ n'est pas convergente, mais possède des sous-suites convergentes, par exemple les sous-suites (x_{2n}) et (x_{2n+1}) . On verra plus loin un théorème général sur ce sujet de l'extraction de sous-suites convergentes (théorème de Bolzano-Weierstrass 1.4.2 : toute suite bornée de nombres réels admet des sous-suites convergentes).

Limite supérieure d'une suite bornée de nombres réels

Soit (x_n) une suite bornée de nombres réels ; désignons par I l'ensemble de tous les nombres réels x tels qu'il existe une infinité d'indices n vérifiant $x < x_n$. Puisque (x_n) est bornée, on peut choisir M tel que $-M \leq x_n \leq M$ pour tout $n \geq 0$. Il n'y a donc aucun indice n tel que $M < x_n$, par conséquent $M \notin I$, et on a $-M - 1 < x_n$ pour *tous* les indices entiers n , donc $-M - 1 \in I$. Par ailleurs il est clair que si $x \in I$ et $x' < x$, on a aussi $x' \in I$. Si on désigne par J le complémentaire de I dans \mathbb{R} , on voit que (I, J) est une coupure, qui définit donc un unique nombre réel z . Précisons les deux propriétés qui caractérisent complètement le nombre z :

- pour tout $x < z$, il existe une infinité d'indices n tels que $x < x_n$
- pour tout $y > z$, il n'existe qu'un nombre fini d'indices n tels que $y < x_n$.

Le nombre z ainsi défini s'appelle la *limite supérieure* de la suite (x_n) , et on le note

$$z = \limsup_{n \rightarrow +\infty} x_n.$$

Indiquons une autre approche. Pour chaque entier $m \geq 0$, considérons la borne supérieure

$$u_m = \sup\{x_m, x_{m+1}, x_{m+2}, \dots\}$$

de l'ensemble A_m formé par les termes de la suite au-delà du rang m . Il n'existe évidemment qu'un nombre fini d'indices k tels que $u_m < x_k$ (tous ces indices k sont $< m$, puisque u_m majore tous les termes x_n pour $n \geq m$), par conséquent $z \leq u_m$. D'autre part, les ensembles A_m étant décroissants par rapport à m , la suite (u_m) est décroissante ; puisqu'elle est minorée par z , elle converge vers une limite u , qui est en fait égale à z (si on avait $z < u$, on pourrait choisir v tel que $z < v < u$. Il n'y a alors qu'un nombre fini d'indices n tels que $v < x_n$, donc on peut trouver m tel que $x_n \leq v$ pour tout $n \geq m$, ce qui entraîne $u_m \leq v$, puis $u \leq u_m \leq v < u$, ce qui est impossible. On doit donc avoir $u = z$). On trouve ainsi une formule qui peut expliquer le nom de limite supérieure,

$$\limsup_{n \rightarrow +\infty} x_n = \lim_{m \rightarrow +\infty} \left(\sup_{n \geq m} x_n \right). \text{ On pose aussi } \liminf_{n \rightarrow +\infty} x_n = \lim_{m \rightarrow +\infty} \left(\inf_{n \geq m} x_n \right).$$

Propriétés de la limite supérieure.

- 1- Si $\lim_n x_n$ existe, on a $\limsup_n x_n = \lim_n x_n = \liminf_n x_n$. Réciproque ?
- 2- Si $\lambda > 0$, on a $\limsup_n \lambda x_n = \lambda \limsup_n x_n$.
- 3- Si $a_n \leq b_n$ pour $n \geq n_0$, on a $\limsup_n a_n \leq \limsup_n b_n$.
- 4- Si $z = \limsup_n x_n$, on peut trouver, aussi loin qu'on veut dans la suite, des termes x_n aussi près qu'on veut de z :

$$\forall \varepsilon > 0, \forall N, \exists n \geq N, |x_n - z| < \varepsilon.$$

Ceci signifie que $z = \limsup_n x_n$ est une *valeur d'adhérence* de la suite (x_n) .

Justifions ce qui précède. Étant donné $\varepsilon > 0$, on sait qu'il n'y a qu'un nombre fini d'indices n tels que $z + \varepsilon < x_n$, et il y a une infinité d'indices n tels que $z - \varepsilon < x_n$; on peut donc certainement trouver un indice $n \geq N$ vérifiant la seconde condition et le contraire de la première, $z - \varepsilon < x_n \leq z + \varepsilon$, ce qui donne $|x_n - z| < 2\varepsilon$.

Exemple, exercice. La suite $((-1)^n)$ possède exactement deux valeurs d'adhérence, 1 et -1 . Si y est une valeur d'adhérence de la suite (x_n) , montrer qu'on peut construire une sous-suite (x_{n_k}) qui converge vers y (on tient alors une démonstration de Bolzano-Weierstrass).

1.3. Suites de Cauchy

Si une suite (x_n) réelle ou complexe converge vers ℓ , on peut trouver pour tout $\varepsilon > 0$ un entier N tel que pour tout $m \geq N$, on ait $|x_m - \ell| < \frac{\varepsilon}{2}$. Si on prend un autre entier $n \geq N$, on aura aussi $|x_n - \ell| < \frac{\varepsilon}{2}$, donc par l'inégalité triangulaire on obtient $|x_m - x_n| < \varepsilon$ pour tous $m, n \geq N$, ce qui est une propriété vérifiée par toute suite convergente, mais où la valeur de la limite n'intervient pas. Nous allons donner un nom à la propriété que nous venons d'isoler.

Définition 1.3.1. Une suite réelle ou complexe (x_n) est une *suite de Cauchy* si pour tout $\varepsilon > 0$, il existe un entier N (dépendant de ε) tel que, pour tous les entiers $m, n \geq N$, on ait $|x_m - x_n| < \varepsilon$.

Toute suite convergente est donc de Cauchy. Une suite complexe est de Cauchy si et seulement si les parties réelles et imaginaires sont de Cauchy. Pour que (x_n) soit de Cauchy il suffit que

$$\forall \varepsilon > 0, \exists N, \forall n \geq N, |x_n - x_N| < \varepsilon.$$

Exercices.

1. Si (x_n) est une suite réelle ou complexe telle que $|x_{n+1} - x_n| \leq 2^{-n}$ pour tout $n \geq 0$, montrer que la suite (x_n) est de Cauchy.

2. Soit f une fonction réelle définie et dérivable sur \mathbb{R} , telle que $|f'(x)| \leq 1/2$ pour tout $x \in \mathbb{R}$; montrer que pour toute valeur de x_0 , la suite définie par $x_{n+1} = f(x_n)$ pour tout $n \geq 0$ est une suite de Cauchy.

Proposition 1.3.1. *Toute suite de Cauchy de nombres réels ou complexes est bornée.*

Démonstration. Soit (x_n) une suite de Cauchy réelle ou complexe; en prenant $\varepsilon = 1$ dans la définition, on trouve un entier N_1 tel que $|x_n - x_{N_1}| < 1$ pour tout $n \geq N_1$, donc $|x_n| \leq |x_{N_1}| + 1$ pour $n \geq N_1$ et finalement

$$M = \max\{|x_0|, \dots, |x_{N_1-1}|, |x_{N_1}| + 1\}$$

borne la suite (x_n) .

Convergence des suites de Cauchy. Espaces complets

Théorème 1.3.1. *Toute suite de Cauchy, réelle ou complexe, est convergente (bien sûr, la limite est réelle si la suite est réelle). Par conséquent, une suite de nombres réels ou complexes est convergente si et seulement si elle est de Cauchy.*

Démonstration. Il suffit de traiter le cas réel. Soit (x_n) une suite de Cauchy de nombres réels; on sait qu'elle est bornée, ce qui permet de définir sa limite supérieure $z = \limsup_n x_n$. On va montrer que la suite (x_n) converge vers z . Donnons nous un nombre $\varepsilon > 0$. Puisque la suite (x_n) est de Cauchy, il existe un entier N tel que pour tous $n, m \geq N$, on ait $|x_n - x_m| < \frac{\varepsilon}{2}$; d'après les propriétés de la limite supérieure, on sait que z est valeur d'adhérence de la suite (x_n) , donc il existe un entier $n_0 \geq N$ tel que $|x_{n_0} - z| < \varepsilon/2$. Pour tout $n \geq N$, on aura

$$|x_n - z| \leq |x_n - x_{n_0}| + |x_{n_0} - z| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2},$$

ce qui montre la convergence de (x_n) vers $z = \limsup_n x_n$.

On dit que \mathbb{R} et \mathbb{C} sont *complets*. On verra plus tard d'autres cadres dans lesquels seront définies la convergence des suites et les suites de Cauchy. On dira alors qu'un « espace » est complet si toute suite de Cauchy de cet espace converge vers une limite qui soit élément de cet espace (ainsi \mathbb{Q} n'est pas complet : toute suite de Cauchy de points de \mathbb{Q} a bien une limite dans \mathbb{R} , mais cette limite n'est pas forcément dans \mathbb{Q}).

Exemple. Soit f une fonction réelle définie et dérivable sur \mathbb{R} , telle que $|f'(x)| \leq 1/2$ pour tout $x \in \mathbb{R}$; pour toute valeur de x_0 , la suite définie par récurrence par la donnée de x_0 et par la relation $x_{n+1} = f(x_n)$ pour tout $n \geq 0$ converge vers l'unique solution de l'équation $f(x) = x$.

1.4. Topologie de \mathbb{R}

Théorème 1.4.1. *Toute suite de nombres réels admet des sous-suites monotones.*

Dans l'énoncé précédent, il faut entendre les mots *suites monotones* dans le sens de monotone au sens large. Bien entendu, on ne peut pas garantir qu'une suite de nombres réels possède toujours une sous-suite croissante, ou qu'elle possède toujours une sous-suite décroissante. On veut seulement dire qu'on a toujours l'une des deux possibilités.

Démonstration. Soit (x_n) une suite de nombres réels ; on va démontrer un résultat un tout petit peu plus précis que le résultat annoncé : il existe une sous-suite croissante au sens large, ou bien une sous-suite strictement décroissante. Disons que l'entier m est une *dernière chance* (sous-entendu : pour avoir une grande valeur de x_m) si

$$\forall n > m, \quad x_n < x_m.$$

Désignons par D le sous-ensemble de \mathbb{N} formé des dernières chances. Ou bien D est fini (peut-être vide), ou bien il est infini. Si D est fini, soit M un majorant de D ; alors aucun entier $m > M$ ne peut être une dernière chance, donc pour tout $m > M$ il existe $n > m$ tel que $x_n \geq x_m$. Il est clair dans ce cas qu'on peut définir par récurrence une sous-suite (x_{n_k}) , croissante au sens large : on pose $n_0 = M+1$; si on a déjà défini $n_0 < n_1 < \dots < n_k$ de façon que $x_{n_0} \leq x_{n_1} \leq \dots \leq x_{n_k}$, on aura $m = n_k > M$, donc il existe $n = n_{k+1} > n_k$ tel que $x_{n_{k+1}} \geq x_{n_k}$. La sous-suite (x_{n_k}) ainsi définie par récurrence est croissante (au sens large).

Dans le deuxième cas, l'ensemble D est infini ; écrivons le sous la forme

$$D = \{m_0 < m_1 < \dots < m_k < \dots\}.$$

On vérifie alors que par définition de D , la sous-suite (x_{m_k}) est décroissante (strictement).

Cas particulier. Soit (n_k) une suite d'entiers ≥ 0 ; on peut trouver, ou bien une sous-suite (n_{k_j}) strictement croissante, ou bien une sous-suite constante. En effet, d'après ce qui précède (avec une modification évidente) on peut trouver une sous-suite strictement croissante, ou bien une sous-suite décroissante. Mais une suite décroissante d'entiers ≥ 0 est constante à partir d'un certain rang.

Exercice (difficile). À chaque couple d'entiers (m, n) avec $m < n$ on associe une couleur, disons bleu ou rouge. Montrer qu'il existe une suite strictement croissante $n_0 < n_1 < \dots$ telle que tous les couples (n_i, n_j) (avec $i < j$) soient de la même couleur (théorème de Ramsey).

Théorème de Bolzano-Weierstrass 1.4.2. *Toute suite bornée de nombres réels admet des sous-suites convergentes.*

Démonstration. La sous-suite monotone fournie par le théorème précédent sera en plus ici une suite monotone *bornée*, donc convergente d'après le théorème 1.2.1.

Indiquons une autre démonstration, appelée démonstration *par dichotomie*. On suppose que la suite (x_n) est contenue dans l'intervalle $[0, 1]$. On construit par récurrence sur l'entier $k \geq 0$ des suites (a_k) , (b_k) de nombres réels et une suite d'entiers (n_k) qui vérifient l'hypothèse de récurrence $H(k)$ suivante (hypothèse en trois points) :

- la suite (finie) a_j , $j = 0, \dots, k$ est croissante, et la suite b_j , $j = 0, \dots, k$ est décroissante ; de plus $a_0 = 0$, $b_0 = 1$, et $b_j - a_j = 2^{-j}$ pour tout $j = 0, \dots, k$,
- les entiers (n_j) , $j = 0, \dots, k$ sont tels que $n_0 < n_1 < \dots < n_k$, et $x_{n_j} \in [a_j, b_j]$ pour tout $j = 0, \dots, k$,

– il existe une infinité d'indices n tels que $x_n \in [a_k, b_k]$.

Tous ces éléments étant donnés, expliquons comment passer de la propriété $H(k)$ à la propriété $H(k+1)$. Supposons $H(k)$ satisfaite. Posons $m = (a_k + b_k)/2$, le milieu du segment $[a_k, b_k]$. L'intervalle $[a_k, b_k]$ est recouvert par les deux intervalles $[a_k, m]$ et $[m, b_k]$, donc l'un de ces deux intervalles contient une infinité de termes de la suite (x_n) donnée, puisque d'après $H(k)$ il existe une infinité d'indices n tels que $x_n \in [a_k, b_k]$. Si c'est le premier intervalle, on pose $[a_{k+1}, b_{k+1}] = [a_k, m]$, c'est-à-dire que $a_{k+1} = a_k$ et $b_{k+1} = m$. Dans le cas contraire on pose $a_{k+1} = m$ et $b_{k+1} = b_k$. Dans tous les cas $a_k \leq a_{k+1} < b_{k+1} \leq b_k$ et $b_{k+1} - a_{k+1} = (b_k - a_k)/2 = 2^{-k-1}$. Puisque $[a_{k+1}, b_{k+1}]$ contient une infinité de termes de la suite, on peut certainement trouver un indice $N > n_k$ tel que $x_N \in [a_{k+1}, b_{k+1}]$. On pose alors pour finir $n_{k+1} = N$.

On a ainsi prouvé la possibilité de la construction par récurrence. On a alors deux suites adjacentes (a_k) et (b_k) , qui ont donc la même limite x , et de plus on a l'inégalité $a_k \leq x_{n_k} \leq b_k$ pour tout k , qui montre que la sous-suite (x_{n_k}) tend vers la même limite x .

Exercice. Autre application de la dichotomie : démontrer le théorème des valeurs intermédiaires en adaptant les idées précédentes.

Ensembles fermés de \mathbb{R}

Définition 1.4.1. On dit qu'un sous-ensemble $F \subset \mathbb{R}$ est *fermé* si pour toute suite (x_n) de points de F qui converge dans \mathbb{R} vers un point x , la limite x reste dans F .

Exemples.

1. Pour tous $a < b$ réels, l'ensemble $F = [a, b]$ est fermé. Supposons en effet que (x_n) soit une suite de points de F , et que (x_n) tende vers une limite $x \in \mathbb{R}$. Puisque $a \leq x_n \leq b$ pour tout n , on obtient $a \leq x \leq b$ par passage à la limite des inégalités larges, donc x est bien dans l'ensemble F .

2. Si $a < b$, l'intervalle $]a, b[$ n'est pas fermé.

3. L'ensemble vide est fermé par pinailage logique : en effet, il n'y a aucune suite de points de \emptyset , donc la condition de définition est satisfaite !

4. Ensembles définis par une inégalité large portant sur une fonction continue : soit $f : F \rightarrow \mathbb{R}$ une fonction continue définie sur une partie fermée F de \mathbb{R} ; pour tout $c \in \mathbb{R}$, les deux ensembles $\{x \in F : f(x) \leq c\}$ et $\{x \in F : f(x) \geq c\}$ sont fermés.

Proposition 1.4.1. Si F_1 et F_2 sont deux ensembles fermés, leur réunion $F_1 \cup F_2$ et leur intersection $F_1 \cap F_2$ sont des ensembles fermés. Plus généralement, l'intersection $\bigcap_{i \in I} F_i$ d'une famille quelconque d'ensembles fermés $(F_i)_{i \in I}$ est un ensemble fermé.

Rappel des propriétés des fonctions continues sur un fermé borné de \mathbb{R} (avec Bolzano)

Lorsque f est une fonction réelle *minorée* définie sur un ensemble non vide X , on peut considérer la borne inférieure $\inf f(X)$ de l'ensemble $f(X)$ de ses valeurs sur X , et on peut toujours trouver une suite (x_n) de points de X telle que

$$\inf f(X) = \lim_n f(x_n).$$

En effet, pour tout $n \geq 1$ on peut trouver par définition de la borne inférieure une valeur $f(x_n)$ de f en un certain point $x_n \in X$ telle que $\inf f(X) \leq f(x_n) < \inf f(X) + 1/n$, donc la suite des valeurs $f(x_n)$ converge vers $\inf f(X)$. Mais on ne peut pas en général trouver un point $x \in X$ tel que $f(x) = \inf f(X)$. Si un tel point x existe, il est clair que $f(x)$ est alors *la plus petite valeur de f sur l'ensemble X* , et on convient de noter dans ce cas

$$f(x) = \min f(X).$$

(et de façon analogue, on note $f(y) = \max f(X)$ si f atteint son maximum sur X au point $y \in X$). Nous allons voir que tel est le cas pour les fonctions continues définies sur un sous-ensemble fermé borné non vide de \mathbb{R} .

Théorème 1.4.3. *Si f est une fonction réelle définie et continue en tout point d'un ensemble fermé borné non vide $F \subset \mathbb{R}$, elle est bornée et elle atteint son maximum et son minimum sur F .*

Démonstration. Montrons d'abord que f est bornée sur F . Sinon, on pourrait trouver pour tout entier $n \geq 0$ un point $x_n \in F$ tel que $|f(x_n)| > n$. Par Bolzano, on trouve une sous-suite (x_{n_k}) convergente vers un certain point $x \in \mathbb{R}$, et on a $x \in F$ puisque F est fermé et que $x_{n_k} \in F$ pour tout k . On devrait alors avoir $f(x_{n_k}) \rightarrow f(x)$ par continuité de la fonction f au point x , mais $|f(x_{n_k})| > n_k$ tend vers l'infini, ce qui est contradictoire. On en déduit que f est bornée sur F .

L'ensemble $f(F)$ des valeurs de f sur F est non vide (puisque F est non vide) et borné, donc on peut définir sa borne inférieure $m = \inf f(F)$ et on peut trouver une suite de points $x_n \in F$ telle que $(f(x_n))$ converge vers m . D'après le théorème 1.4.2 (de Bolzano-Weierstrass), il existe une sous-suite (x_{n_k}) qui converge vers un point $x \in \mathbb{R}$. Puisque $x_{n_k} \in F$ pour tout k , on obtient $x \in F$ parce que F est fermé, donc x est bien dans l'ensemble de définition de f . Puisque f est continue au point x , on doit avoir $f(x) = \lim_k f(x_{n_k}) = m$, donc $f(x) \leq f(t)$ pour tout $t \in F$, c'est-à-dire que la fonction f atteint son minimum sur F au point x . La même démonstration s'applique pour trouver un point $y \in F$ où le maximum de f est atteint.

Corollaire 1.4.1. *Tout fermé F borné et non vide de \mathbb{R} possède un plus grand et un plus petit élément.*

Démonstration. La fonction continue f définie sur F par $f(x) = x$ atteint son maximum et son minimum sur F .

Corollaire 1.4.2. *Soit F un fermé non borné de \mathbb{R} et soit f une fonction réelle continue sur F telle que $f(x)$ tende vers $+\infty$ lorsque $|x| \rightarrow +\infty$ (par valeurs $x \in F$). La fonction f atteint son minimum sur l'ensemble F .*

Démonstration. Soit x_0 un point de F et posons $F_0 = \{x \in F : f(x) \leq f(x_0)\}$; cet ensemble F_0 est un fermé non vide. De plus il est borné (puisque f tend vers $+\infty$, il existe R tel que $f(y) > f(x_0)$ pour tout $y \in F$ tel que $|y| > R$, donc F_0 est borné par R), donc f atteint son minimum sur F_0 en un point x_1 et ce minimum est aussi le minimum de f sur l'ensemble F tout entier : si $z \in F$, ou bien $z \in F_0$ et on sait que $f(x_1) \leq f(z)$, ou bien $z \notin F_0$ et alors $f(x_1) \leq f(x_0) < f(z)$.

Corollaire 1.4.3. *Tout ensemble F fermé non vide et majoré de \mathbb{R} contient sa borne supérieure (il possède donc un plus grand élément).*

Démonstration. Si F est borné on connaît déjà le résultat. Sinon, on applique le corollaire précédent à la fonction $f(x) = -x$.

Théorème 1.4.4. *Pour toute suite décroissante $(F_n)_n$ de fermés bornés non vides de \mathbb{R} , l'intersection $\bigcap_n F_n$ est non vide.*

Démonstration. Pour tout $n \geq 0$ désignons par x_n le plus petit élément de F_n . On vérifie que la suite (x_n) est croissante (parce que $F_{n+1} \subset F_n$), et elle est majorée puisqu'elle est contenue dans l'ensemble borné F_0 . Elle est donc convergente vers un certain nombre x . Pour tout n_0 fixé, on peut dire que x est la limite de la suite $(x_n)_{n \geq n_0}$ de points du

fermé F_{n_0} , donc $x \in F_{n_0}$ pour tout n_0 , ce qui signifie que x est dans l'intersection de tous les F_n .

Définition 1.4.2. Soit A une partie non vide de \mathbb{R} et soit f une fonction réelle définie sur l'ensemble A ; on dit que f est *uniformément continue* sur A si pour tout $\varepsilon > 0$, il existe $\delta = \delta(\varepsilon) > 0$ tel que

$$\forall x, y \in A, \quad (|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon).$$

Exemples. La fonction $x \rightarrow x^2$ n'est pas uniformément continue sur \mathbb{R} (pour tout $\eta > 0$ donné, la fonction $(x + \eta/2)^2 - x^2 \geq \eta x$ tend vers $+\infty$ lorsque $x \rightarrow +\infty$). En revanche toute fonction de classe C^1 à dérivée bornée sur un intervalle I est uniformément continue sur I (appliquer la majoration des accroissements finis).

Si f est uniformément continue sur A , on définit son *module de continuité* ω_f , qui est une fonction définie sur $]0, +\infty[$ par la formule

$$\omega_f(\delta) = \sup\{|f(x) - f(y)| : x, y \in A, |x - y| < \delta\}.$$

Dire que f est uniformément continue signifie précisément que $\omega_f(\delta)$ tend vers 0 lorsque δ tend vers 0.

Théorème 1.4.5. *Toute fonction continue sur un fermé borné F (non vide) de \mathbb{R} est uniformément continue.*

Démonstration. Par l'absurde, avec Bolzano-Weierstrass : si $f : F \rightarrow \mathbb{R}$ n'est pas uniformément continue, il existe $\varepsilon > 0$ tel que pour tout $\delta > 0$, on puisse trouver deux points x, y de F tels que $|x - y| < \delta$ mais $|f(x) - f(y)| \geq \varepsilon$. On peut appliquer ceci en particulier pour chaque choix $\delta = 1/n$, pour tout $n \geq 1$. On obtient ainsi deux suites (x_n) et (y_n) de points de F telles que $|x_n - y_n| < 1/n$ et $|f(x_n) - f(y_n)| \geq \varepsilon$ pour tout $n \geq 1$. D'après Bolzano, on peut trouver une sous-suite (x_{n_k}) qui converge vers un point $x \in F$. Puisque $(x_n - y_n)$ tend vers 0, il en résulte que (y_{n_k}) converge aussi vers x . En passant à la limite dans les relations précédentes on obtient $|f(x) - f(x)| \geq \varepsilon$, contradiction.

Ouverts de \mathbb{R}

Définition 1.4.3. On dit qu'un sous-ensemble V de \mathbb{R} est un *voisinage* d'un point $x \in \mathbb{R}$ s'il existe $r > 0$ tel que $]x - r, x + r[\subset V$.

Par exemple, le segment $[0, 1]$ n'est pas un voisinage du point 0 (ni du point 1), mais c'est un voisinage de chacun des points x de l'intervalle ouvert $]0, 1[$.

Définition 1.4.4. On dit qu'un sous-ensemble Ω de \mathbb{R} est *ouvert* si Ω est un voisinage de chacun de ses points, c'est-à-dire si pour tout point $x \in \Omega$, il existe $r > 0$ tel que $]x - r, x + r[\subset \Omega$.

Exemples.

1. \emptyset et \mathbb{R} sont ouverts dans \mathbb{R} .
2. Un intervalle $]a, b[$ est ouvert dans \mathbb{R} . L'intervalle $]a, b]$ n'est pas ouvert (ni fermé!).
3. Si ω est un ouvert de \mathbb{R} , et f une fonction continue de \mathbb{R} dans \mathbb{R} , l'image inverse $\Omega = f^{-1}(\omega) = \{x \in \mathbb{R} : f(x) \in \omega\}$ est un ensemble ouvert de \mathbb{R} .

Proposition 1.4.2. *Propriétés des ouverts. L'intersection $\Omega_1 \cap \dots \cap \Omega_n$ d'une famille finie d'ouverts $\Omega_1, \dots, \Omega_n$ est un ouvert. La réunion d'une famille quelconque d'ouverts est un ouvert.*

Proposition 1.4.3. *Un ensemble $F \subset \mathbb{R}$ est fermé si et seulement si son complémentaire est ouvert.*

Démonstration. Soit Ω le complémentaire de F dans \mathbb{R} et supposons que Ω est un ensemble ouvert. Soit (x_n) une suite de points de F qui converge vers un point $x \in \mathbb{R}$, et montrons que $x \in F$; sinon, on aurait $x \in \Omega$, donc il existerait puisque Ω est ouvert un nombre $r > 0$ tel que $I =]x - r, x + r[\subset \Omega$, c'est-à-dire que I ne contiendrait aucun point de F . Ceci est impossible, puisque pour n assez grand on a $|x_n - x| < r$ mais $x_n \in F$.

Inversement supposons F fermé, et montrons que Ω est voisinage de chacun de ses points. Soit $x \in \Omega$; s'il n'existait aucun nombre $r > 0$ tel que l'intervalle $]x - r, x + r[$ soit contenu dans Ω , on pourrait pour tout $n \geq 1$, en posant $r = 1/n$, trouver au moins un point $x_n \in]x - 1/n, x + 1/n[$ qui ne soit pas dans Ω , c'est-à-dire que $x_n \in F$. Puisque $|x_n - x| < 1/n$, on voit que la suite (x_n) de points de F converge vers le point x qui n'est pas dans F , ce qui est impossible puisque F est fermé.

Exemple d'ensemble fermé « compliqué » : l'ensemble triadique de Cantor. C'est un exemple d'ensemble qui est très difficile à « sentir » autrement qu'en regardant son complémentaire. On définit donc une suite d'ouverts contenus dans $[0, 1]$ en posant $\Omega_1 =]1/3, 2/3[$, puis

$$\Omega_k = \{x \in [0, 1] : 1/3 < 3^k x - [3^k x] < 2/3\}$$

pour tout $k \geq 1$, où $[y]$ désigne la partie entière de y (le lecteur fera un petit dessin, pour $k = 1, 2, 3$). L'ensemble Δ est le fermé complémentaire dans $[0, 1]$ de l'ensemble ouvert $\bigcup_{n \geq 1} \Omega_n$ (si on veut, on peut dire que Δ est l'ensemble des $x \in [0, 1]$ tels que $\sin(3^k \pi x) \geq 0$ pour tout entier $k \geq 1$).

Composantes d'un ouvert de \mathbb{R}

Lemme. Soit Ω un ouvert de \mathbb{R} et soit $x \in \Omega$; ou bien $[x, +\infty[\subset \Omega$, ou bien il existe b tel que $x < b$, $[x, b[\subset \Omega$ et $b \notin \Omega$. De même à gauche, ou bien $] - \infty, x] \subset \Omega$, ou bien il existe a tel que $a < x$, $]a, x] \subset \Omega$ et $a \notin \Omega$. Tout ouvert de \mathbb{R} est donc réunion d'intervalles ouverts deux à deux disjoints.

Démonstration. On pose $A = \{y > x : y \notin \Omega\}$. Si A est vide, on a $[x, +\infty[\subset \Omega$; sinon, on peut considérer le fermé F formé des points $y \geq x$ qui ne sont pas dans Ω (intersection du complémentaire de Ω avec le fermé $[x, +\infty[$). Ce fermé est non vide, et minoré par x . Il contient donc un plus petit élément b . On a alors $b \notin \Omega$ et $[x, b[\subset \Omega$.

Exercice. Montrer qu'un ensemble à la fois ouvert et fermé dans \mathbb{R} est vide ou égal à \mathbb{R} (ou encore : si \mathbb{R} est la réunion $F_1 \cup F_2$ de deux fermés non vides, l'intersection $F_1 \cap F_2$ est un ensemble non vide).

Valeur d'adhérence d'une suite de nombres réels

Définition 1.4.5. Soit (x_n) une suite de nombres réels; on dit que $y \in \mathbb{R}$ est une *valeur d'adhérence* de la suite (x_n) si pour tout entier $N \geq 0$ et pour tout nombre réel $\varepsilon > 0$ il existe un entier $n \geq N$ tel que $|x_n - y| < \varepsilon$.

Exemples.

1. La suite $x_n = n$ n'admet aucune valeur d'adhérence dans \mathbb{R} .
2. Si (x_n) converge vers $x \in \mathbb{R}$, la valeur x est valeur d'adhérence de la suite, et c'est la seule valeur d'adhérence de la suite.
3. La suite (x_n) définie par $x_n = (-1)^n$ pour tout entier $n \geq 0$ admet exactement deux valeurs d'adhérence, 1 et -1 .
4. Si une sous-suite (x_{n_k}) de la suite (x_n) converge vers y , alors y est valeur d'adhérence de la suite (x_n) .
5. Pour toute suite bornée (x_n) , la valeur $\limsup_n x_n$ est valeur d'adhérence, et c'est la plus grande valeur d'adhérence de la suite; de la même façon, $\liminf_n x_n$ est la plus petite valeur d'adhérence de la suite (x_n) .

On va voir que l'exemple 4 est tout à fait général.

Proposition 1.4.4. *Le nombre y est valeur d'adhérence de la suite (x_n) si et seulement s'il existe une sous-suite de la suite (x_n) qui converge vers y .*

Démonstration. Dans la direction « si », c'est le contenu de l'exemple 4 qui précède. Inversement, supposons que y soit valeur d'adhérence. On va expliquer qu'on peut trouver une sous-suite (x_{n_k}) telle que $|x_{n_k} - y| < 2^{-k}$ pour tout entier $k \geq 0$. Si n_0, \dots, n_k sont déjà choisis on applique la définition de valeur d'adhérence avec $\varepsilon = 2^{-k-1}$ et $N = n_k + 1$ pour trouver $n_{k+1} = n \geq N > n_k$ tel que $|x_{n_{k+1}} - y| < 2^{-k-1}$.

Remarque. On retrouve Bolzano-Weierstrass en traduisant les informations précédentes. Si (x_n) est bornée, $x = \limsup_n x_n$ existe, x est une valeur d'adhérence de la suite (x_n) , donc il existe une sous-suite qui converge vers x .

Exercice. Montrer que si la suite (x_n) n'a aucune valeur d'adhérence, on a $\lim |x_n| = +\infty$.

Chapitre 2. Séries numériques

2.1. Introduction

Commençons par donner quelques exemples frappants de séries numériques, avant même de donner la définition mathématique de cette notion.

$$(A) \quad \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} + \cdots = 1.$$

$$(B) \quad 1 + \frac{1}{2} + \cdots + \frac{1}{n} + \cdots = +\infty.$$

$$(C) \quad 1 - \frac{1}{2} + \cdots + \frac{(-1)^{n+1}}{n} + \cdots = \ln 2.$$

$$(D) \quad 1 - 1 + \cdots + (-1)^n + \cdots = ???$$

Il faut certainement discuter ces exemples. La première saine réaction est la suivante : avant que l'on ait posé une définition mathématique précise, on a tout à fait le droit de dire que la notation

$$\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^n} + \cdots$$

ne veut rien dire (ce sont les trois petits points à la fin de la ligne précédente qui posent problème). Si on est dans un état d'esprit plus conciliant, on peut accepter d'essayer de deviner le sens qu'on donnera un peu plus loin. Essayons donc. Le premier exemple se comprend bien avec un schéma de $[0, 1]$: de l'intervalle $[0, 1]$, on efface la première moitié $[0, \frac{1}{2}[$. Il reste une longueur $1/2$ dans l'intervalle restant $[\frac{1}{2}, 1]$; on enlève maintenant la première moitié $[\frac{1}{2}, \frac{3}{4}[$ de l'intervalle restant, c'est-à-dire qu'on a enlevé en tout $\frac{1}{2} + \frac{1}{4}$, et il reste l'intervalle $[\frac{3}{4}, 1]$, de longueur $1/4$. On continue... On « voit bien » qu'on épuisera $[0, 1]$ en enlevant successivement $1/2, 1/4, 1/8, \dots$

Pour le second exemple on note que

$$1 \geq \frac{1}{2}; \quad \frac{1}{2} + \frac{1}{3} \geq 2 \times \frac{1}{4} = \frac{1}{2}; \quad \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} \geq 4 \times \frac{1}{8} = \frac{1}{2}, \dots$$

et de même pour tout entier $n \geq 1$

$$\frac{1}{2^{n-1}} + \cdots + \frac{1}{2^n - 1} \geq \frac{1}{2},$$

ce qui montre que

$$1 + \frac{1}{2} + \cdots + \frac{1}{2^n - 1} \geq \frac{n}{2}.$$

Le troisième exemple ne se devine pas, mais montre qu'une somme « infinie » peut devenir finie quand on change les signes. Dans le dernier exemple il y a un choix à faire ; on pourrait se débrouiller pour donner une valeur à cette expression (exercice hors sujet : laquelle ?), mais on choisira une définition pour laquelle la somme (D) ne sera pas définie ; on dira que cette série *diverge*.

2.2. Séries convergentes à termes réels ou complexes

Définition 2.2.1. On suppose donnés des nombres réels ou complexes u_0, u_1, \dots . On associe à ces nombres la suite (U_n) des *sommes partielles* définie par

$$\forall n \geq 0, \quad U_n = u_0 + \dots + u_n.$$

On dit que la série $\sum (u_n)_n$ est *convergente* si la suite $(U_n)_n$ des sommes partielles tend vers une limite finie (dans le cas réel ; vers une limite complexe dans le cas complexe). Dans ce cas on définit la *somme* U de la série par $U = \lim_n U_n$ et on introduit la notation

$$\sum_{n=0}^{+\infty} u_n = U = \lim_{n \rightarrow +\infty} U_n$$

ou bien

$$u_0 + u_1 + \dots + u_n + \dots = \lim_{n \rightarrow +\infty} U_n.$$

Remarque. Le lecteur pourra noter l'analogie avec la définition des intégrales convergentes $\int_1^{+\infty} f(t) dt$. Si f admet une primitive F , l'intégrale $\int_1^{+\infty} f(t) dt$ converge si et seulement si $F(t)$ tend vers une limite quand $t \rightarrow +\infty$. La suite des sommes partielles (U_n) joue donc un rôle analogue à celui de la primitive F .

La notation lourde mais correcte $\sum (u_n)_{n \geq 0}$ (que personne à part l'auteur du poly n'emploie) est censée aider à distinguer entre l'étude de la série $\sum (u_n)_{n \geq 0}$, et la valeur $\sum_{n=0}^{+\infty} u_n$ de la somme dans le cas où la série converge. Pratiquement, et sauf exception nous écrirons comme tout le monde « étudier la série $\sum u_n$ » (sans bornes de sommation). En revanche nous essaierons de réserver le symbole $\sum_{n=0}^{+\infty} u_n$ à la valeur de la somme.

Une série à termes complexes converge si et seulement si les deux séries à termes réels des parties réelles et parties imaginaires convergent, et dans le cas de convergence, la somme de la série complexe est égale à

$$\sum_{n=0}^{+\infty} \operatorname{Re} u_n + i \sum_{n=0}^{+\infty} \operatorname{Im} u_n.$$

Souvent, cette méthode n'est pas la bonne pour calculer la somme de la série (elle est cependant raisonnable pour déterminer si la série est convergente). Pour trouver la somme, il est souvent préférable d'adopter un point de vue global, et de considérer les sommes partielles complexes directement. On verra un premier exemple un peu plus loin avec le calcul de la somme de la série géométrique.

On dit que u_n est le *terme général* de la série $\sum (u_n)_{n \geq 0}$. Une série qui ne converge pas est appelée *série divergente*. La nature d'une série ne change pas si on modifie un nombre fini de ses termes (mais bien entendu, la *somme* de la série peut changer dans cette opération). On écrira « la série $\sum u_n$ converge », ou « la série $\sum u_n$ diverge », ou « étudier la série $\sum u_n$ », sans indiquer les bornes de sommation ; en revanche la notation $\sum_{n=n_0}^{+\infty} u_n$ représentera un *nombre* bien précis, égal à la somme de la série

$$u_{n_0} + u_{n_0+1} + \dots$$

(qui sera donc supposée convergente).

Si la série $\sum (u_n)_n$ converge, alors $\lim_n u_n = 0$; **l'inverse n'est pas vrai**, comme le montre l'exemple (B) de l'introduction. De façon tout à fait étonnante, un certain nombre d'étudiants persistent à oublier la phrase en gras qui précède.

Suites et séries

Donnons une interprétation « dynamique » pour essayer de balayer les confusions fâcheuses entre suites et séries. Quand on étudie une suite (x_n) , on peut penser que x_n représente la *position* à l'instant n , et que l'on s'intéresse à l'existence d'une position limite quand le temps n devient très grand. Quand on étudie une série $\sum u_n$, il faut penser que u_n représente le *déplacement* entre les instants $n - 1$ et n . Mais ce qui nous intéresse, c'est tout de même la position $U_n = u_0 + \dots + u_n$ à l'instant n . L'étude des séries donne des critères portant sur les *déplacements* qui permettent de garantir que la *position limite* existe.

On peut facilement retourner la correspondance séries-suites :

Théorème 2.2.1. *Soit (x_n) une suite de nombres réels ou complexes ; la série numérique $\sum (x_{n+1} - x_n)_{n \geq 0}$ converge si et seulement si la suite (x_n) tend vers une limite finie (ou une limite complexe). Posons plus précisément $u_0 = x_0$ et $u_{n+1} = x_{n+1} - x_n$ pour tout $n \geq 1$. Dans le cas où la série $\sum u_n$ converge, on a*

$$\lim_{n \rightarrow +\infty} x_n = \sum_{n=0}^{+\infty} u_n.$$

Démonstration. On voit que

$$U_n = u_0 + \dots + u_n = x_0 + (x_1 - x_0) + \dots + (x_n - x_{n-1}) = x_n,$$

par conséquent la suite des sommes partielles (U_n) de la série $\sum u_n$ est exactement la suite (x_n) . Par définition, la série $\sum u_n$ converge si et seulement si la suite (U_n) converge, c'est-à-dire si et seulement si la suite (x_n) converge. De plus, dans le cas de convergence, la somme $\sum_{n=0}^{+\infty} u_n$ de la série est par définition la limite de la suite (U_n) , c'est-à-dire ici $\lim_n x_n$.

Exemples.

1. Montrer que

$$\frac{1}{1.2} + \frac{1}{2.3} + \dots + \frac{1}{(n+1)(n+2)} + \dots = 1.$$

2. Montrer que la série

$$\sum \ln\left(1 + \frac{1}{n+1}\right)$$

diverge. Dans ces deux exemples, on calcule directement (explicitement) les sommes partielles U_n pour tout $n \geq 0$.

Proposition 2.2.1. *Soient $\sum u_n$ et $\sum v_n$ deux séries convergentes, et soit $\lambda \in \mathbb{C}$; la série $\sum (\lambda u_n + v_n)$ est convergente, et*

$$\sum_{n=0}^{+\infty} (\lambda u_n + v_n) = \lambda \sum_{n=0}^{+\infty} u_n + \sum_{n=0}^{+\infty} v_n.$$

Série géométrique

Soit $z \in \mathbb{C}$; on étudie la série $\sum z^n$. Pour qu'elle converge, il est nécessaire que $z^n \rightarrow 0$, donc nécessaire que $|z| < 1$. Inversement, lorsque $|z| < 1$, on sait que $z^{n+1} \rightarrow 0$ et la formule

$$1 + z + \dots + z^n = \frac{1}{1-z} - \frac{z^{n+1}}{1-z}$$

montre que la série converge. En résumé :

La série géométrique $\sum z^n$ converge si et seulement si $|z| < 1$. Dans le cas $|z| < 1$, on a

$$\sum_{n=0}^{+\infty} z^n = \frac{1}{1-z}.$$

Remarque. Dans le cas complexe, le calcul de la somme serait bien plus difficile si on séparait partie réelle et partie imaginaire.

Exercices et exemples.

1. (Question préparatoire) Montrer que pour tout $a \in \mathbb{C}$, on a

$$\lim_n \frac{a^n}{n!} = 0.$$

2. Séries obtenues par la formule de Taylor-Lagrange, ou Taylor avec reste intégral : montrer que pour tout $x \in \mathbb{R}$,

$$e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \dots$$

Étudier la série de Taylor d'une fonction réelle f définie sur \mathbb{R} , indéfiniment dérivable et telle que $|f^{(n)}(x)| \leq 1$ pour tout $n \geq 1$ et tout $x \in \mathbb{R}$. Traiter les exemples de $\sin x$ et $\cos x$.

3. Montrer la convergence de la série de Taylor de $-\ln(1-x)$ pour $x = -1$ (ou par la même méthode, pour $-1 \leq x \leq 1/2$). Montrer que

$$\ln 2 = 1 - \frac{1}{2} + \dots + \frac{(-1)^{n+1}}{n} + \dots$$

2.3. Séries à termes réels positifs

On dit que $\sum u_n$ est une série à termes réels positifs si $u_n \geq 0$ pour tout $n \geq 0$. Dans ce cas il est clair que la suite (U_n) des sommes partielles est croissante, donc d'après le théorème 1.2.1 on peut dire :

Théorème 2.3.1. Une série $\sum u_n$ à termes positifs est convergente si et seulement si la suite (U_n) des sommes partielles est majorée. Lorsque $\sum u_n$ converge, on a pour tout entier $n \geq 0$ l'inégalité

$$U_n \leq \sum_{m=0}^{+\infty} u_m.$$

Remarque. Si on a seulement $u_n \geq 0$ pour $n \geq n_0$, la première partie du théorème subsiste, mais pas l'inégalité de la deuxième partie de l'énoncé.

Remarque. Une série à termes positifs diverge si et seulement si $U_n \rightarrow +\infty$. On écrit parfois $\sum u_n = +\infty$ pour dire qu'une telle série diverge, et inversement on peut écrire $\sum u_n < +\infty$ pour exprimer qu'une série **à termes positifs** est convergente.

Théorème de comparaison 2.3.2. Supposons que $0 \leq u_n \leq v_n$ pour tout $n \geq 0$.

a. Si $\sum v_n$ converge, alors $\sum u_n$ converge et

$$\sum_{n=0}^{\infty} u_n \leq \sum_{n=0}^{\infty} v_n.$$

b. Si $\sum u_n$ diverge, alors $\sum v_n$ diverge.

Dans le cas où les hypothèses $0 \leq u_n \leq v_n$ ne sont vérifiées que pour $n \geq n_0$, on obtient encore les résultats de comparaison pour la nature des deux séries (qui n'est pas modifiée quand on change un nombre fini de termes) mais bien entendu l'inégalité entre les sommes des deux séries n'est plus nécessairement vraie.

Exemples.

1. La série $\sum (2^{-n}/n)_{n \geq 1}$ converge.

2. Pour tout $x \in [0, 1[$, la série $\sum n^2 x^n$ converge (comparer à la série $\sum y^n$ pour un y tel que $x < y < 1$).

3. Pour toute suite (k_n) d'entiers égaux à $0, 1, \dots, 9$, la série $\sum k_n/10^n$ converge. Montrer que tout nombre $x \in [0, 1]$ est la somme d'au moins une série de la forme

$$\frac{k_1}{10} + \frac{k_2}{100} + \frac{k_3}{1000} + \dots$$

avec $k_i \in \{0, 1, \dots, 9\}$ pour tout $i \geq 1$. Quelle est la somme de la série

$$\sum_{n=0}^{+\infty} \frac{9}{10^{n+1}} = \frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \dots ?$$

On dit qu'un développement décimal est *impropre* si toutes les décimales à partir d'un certain rang sont égales à 9, comme dans l'exemple précédent. Tout nombre décimal possède un développement propre et un développement impropre.

4. Généraliser au développement en base b , où b est un entier ≥ 2 .

Indiquons un cas particulier utile du principe de comparaison : supposons que l'on ait $u_n, v_n > 0$ pour tout entier $n \geq 0$ et que $u_{n+1}/u_n \leq v_{n+1}/v_n$; on vérifie qu'alors $u_n \leq (u_0/v_0) v_n$, ce qui permet d'appliquer le théorème de comparaison. Dans le cas où les hypothèses $u_n, v_n > 0$ et $u_{n+1}/u_n \leq v_{n+1}/v_n$ ne sont satisfaites qu'à partir des rangs $n \geq n_0$, on obtient de même $u_n \leq (u_{n_0}/v_{n_0}) v_n$ pour tout $n \geq n_0$, ce qui permet encore d'utiliser les principes de comparaison.

On dit que deux suites réelles ou complexes (x_n) et (y_n) sont *équivalentes*, et on note $x_n \sim y_n$ si pour n assez grand x_n et y_n sont non nuls et si $\lim_n x_n/y_n = 1$. On peut étendre la définition aux suites qui peuvent s'annuler pour certaines valeurs des indices, en disant que (x_n) et (y_n) sont équivalentes s'il existe une suite (a_n) telle que $\lim_n a_n = 1$ et telle que $y_n = a_n x_n$ pour tout $n \geq n_0$ (bien entendu, $a_n = y_n/x_n$ chaque fois que $x_n \neq 0$).

Corollaire 2.3.1. Si $u_n \sim v_n$ et $\sum u_n$ à termes positifs, les deux séries $\sum u_n$ et $\sum v_n$ sont de même nature.

Exemples.

1. La série $\sum (1/n^2)_{n \geq 1}$ converge ; on peut la comparer à la série convergente déjà étudiée plus haut, $\sum (1/n(n+1))_{n \geq 1}$.

2. La série $\sum(1 - \cos(1/n))_{n \geq 1}$ est de même nature que la série $\sum 1/n^2$, elle est donc convergente.

3. On sait que la série $\sum(1/n)_{n \geq 1}$ diverge, mais on peut retrouver ce résultat d'une autre façon que dans l'introduction en comparant cette série à la série $\sum(\ln(1+1/n))_{n \geq 1}$ précédemment étudiée en exemple.

ATTENTION! Le corollaire précédent est **FAUX** pour les séries dont les termes changent de signe.

Les bons vieux critères (Cours d'Analyse de Cauchy ~ 1820)

Mentionnons deux critères de comparaison avec la série géométrique, que nous appellerons **A** et **B** (et qui ont quelques variantes). **On suppose ici** $u_n \geq 0$ pour tout entier n assez grand.

Critère A. *S'il existe $x < 1$ tel que $u_n^{1/n} \leq x$ pour tout entier n assez grand, la série $\sum u_n$ converge. Si $u_n^{1/n} \geq 1$ pour une infinité d'entiers n , la série $\sum u_n$ diverge.*

La deuxième partie du critère est évidente, puisque si $u_n^{1/n} \geq 1$ pour une infinité d'entiers, on a $u_n \geq 1$ pour une infinité d'entiers, et u_n ne tend pas vers 0. La vérification de la première partie du critère **A** est très simple. Si $u_n^{1/n} \leq x < 1$ pour tout $n \geq n_0$, on a directement $u_n \leq x^n$ pour $n \geq n_0$, et x^n est le terme général d'une série convergente puisque $x < 1$. La première partie du critère peut s'exprimer de façon plus rapide avec la notion de limite supérieure :

Critère A1. *Si $\limsup_n u_n^{1/n} < 1$, la série $\sum u_n$ converge. Si $\limsup_n u_n^{1/n} > 1$, la série $\sum u_n$ diverge.*

Si $\limsup_n u_n^{1/n} > 1$, on est sûr que (u_n) ne tend pas vers 0, ce qui justifie la deuxième partie de **A1**. **ATTENTION** : si $\limsup_n u_n^{1/n} = 1$, il est possible que $\lim_n u_n = 0$, et on ne peut pas conclure (voir plus loin). Le critère ci-dessus implique le cas particulier suivant :

Critère A'. *Supposons que $\lim_n u_n^{1/n}$ existe. Si $\lim_n u_n^{1/n} < 1$, la série $\sum u_n$ converge. Si $\lim_n u_n^{1/n} > 1$, la série $\sum u_n$ diverge.*

On remarquera bien qu'il n'y a pas de conclusion possible dans le cas limite où $\lim_n u_n^{1/n} = 1$. En effet, on a vu que $\sum(1/n)_{n \geq 1}$ diverge alors que $\sum(1/n^2)_{n \geq 1}$ converge. Pourtant, on a $\lim_n u_n^{1/n} = 1$ dans les deux cas (petit exercice de calcul de limite).

Le second critère utilise les quotients u_{n+1}/u_n : on suppose donc ici, non seulement que $u_n \geq 0$ mais aussi que $u_n > 0$ pour tout entier n assez grand.

Critère B. *S'il existe $x < 1$ tel que $u_{n+1}/u_n \leq x$ pour tout entier n assez grand, la série $\sum u_n$ converge. Si $u_{n+1}/u_n \geq 1$ pour tout entier n assez grand, la série $\sum u_n$ diverge.*

Ici encore la deuxième partie du critère est évidente, puisque si $u_{n+1}/u_n \geq 1$ pour tout $n \geq n_0$, on aura $u_n \geq u_{n_0} > 0$ qui ne tend pas vers 0. La vérification de la première partie du critère **B** est très simple. Si $u_{n+1}/u_n \leq x < 1$ pour $n \geq n_0$, on a en posant $v_n = x^n$ que $u_{n+1}/u_n \leq x = v_{n+1}/v_n$ pour $n \geq n_0$, donc on peut majorer $\sum u_n$ par une série géométrique convergente. Si $\lim_n u_{n+1}/u_n$ existe, remarquons que si $\lim_n u_{n+1}/u_n < 1$, on est dans le premier cas du critère **B**, et si $\lim_n u_{n+1}/u_n > 1$,

on est dans le deuxième cas. Par conséquent, le second critère ci-dessus implique le cas particulier suivant :

Critère B'. Supposons que $\lim_n u_{n+1}/u_n$ existe. Si $\lim_n u_{n+1}/u_n < 1$, la série $\sum u_n$ converge. Si $\lim_n u_{n+1}/u_n > 1$, la série $\sum u_n$ diverge.

On remarquera ici encore qu'il n'y a pas de conclusion possible dans le cas limite où $\lim_n u_{n+1}/u_n = 1$ (pour la même raison que pour le critère A').

Exercices.

1. Montrer que si $\lim_n u_{n+1}/u_n$ existe, alors

$$\lim_n u_n^{1/n} = \lim_n \frac{u_{n+1}}{u_n}.$$

2. Applications des critères. Retrouver la convergence pour tout $a > 0$ de la série exponentielle $\sum a^n/(n!)$, étudier la convergence des séries $\sum n^b a^n$ pour $a \geq 0$, $a \neq 1$. Vérifier que les deux critères échouent sur la série de Riemann $\sum n^b$.

3. Le critère $\sum 2^n u_{2^n}$ pour les séries à termes positifs décroissants :

Si $0 \leq u_{n+1} \leq u_n$ pour tout $n \geq 0$ (c'est-à-dire si le terme général est décroissant), la série $\sum u_n$ converge si et seulement si la série $\sum 2^n u_{2^n}$ converge.

(Remarquer que $2^k u_{2^{k+1}} \leq u_{2^k} + \dots + u_{2^{k+1}-1} \leq 2^k u_{2^k}$ pour tout $k \geq 0$). Montrer que l'application de cette règle à la série de Riemann $\sum 1/n^\alpha$, dans le cas $\alpha \geq 0$, ramène à une série géométrique et permet de conclure. Étudier de même les séries de Bertrand $\sum \left(\frac{1}{n(\ln n)^\alpha}\right)_{n \geq 1}$.

2.4. Comparaison séries-intégrales

Proposition 2.4.1. Soit $f : [1, +\infty[\rightarrow [0, +\infty[$ une fonction décroissante et positive ; on a pour tout entier $n \geq 2$

$$\sum_{i=2}^n f(i) \leq \int_1^n f(t) dt \leq \sum_{i=1}^{n-1} f(i).$$

En conséquence, la série $\sum f(n)$ converge si et seulement si l'intégrale $\int_1^{+\infty} f(t) dt$ converge.

On peut donner des informations plus précises. Si $\sum u_n$ est une série convergente, on appelle *reste d'ordre n* de la série la quantité

$$R_n = \sum_{m=0}^{+\infty} u_m - U_n = u_{n+1} + u_{n+2} + \dots$$

Proposition 2.4.2. Soit $f : [1, +\infty[\rightarrow [0, +\infty[$ une fonction décroissante et positive ; considérons la série de terme général $u_n = f(n)$. On a lorsque l'intégrale $\int_1^{+\infty} f(t) dt$ converge l'encadrement

$$\int_{n+1}^{+\infty} f(t) dt \leq R_n = f(n+1) + f(n+2) + \dots \leq \int_n^{+\infty} f(t) dt$$

pour tout entier $n \geq 1$. Dans le cas où l'intégrale est divergente,

$$f(1) + f(2) + \dots + f(n) = U_n \sim \int_1^n f(t) dt \quad (n \rightarrow +\infty).$$

Séries de Riemann : étude par comparaison à une intégrale

Théorème 2.4.1. La série de Riemann $\sum (1/n^\alpha)_{n \geq 1}$ est convergente si et seulement si $\alpha > 1$. De plus pour $\alpha > 1$

$$\frac{1}{(n+1)^\alpha} + \frac{1}{(n+2)^\alpha} + \dots \sim \frac{n^{1-\alpha}}{\alpha-1}.$$

Pour $\alpha = 1$,

$$1 + \frac{1}{2} + \dots + \frac{1}{n} \sim \ln n$$

et pour $\beta < 1$

$$1 + \frac{1}{2^\beta} + \dots + \frac{1}{n^\beta} \sim \frac{n^{1-\beta}}{1-\beta}.$$

Exercice. Constante d'Euler : montrer l'existence de la limite

$$C = \lim_n \left(1 + \frac{1}{2} + \dots + \frac{1}{n} - \ln n \right)$$

en étudiant la série $\sum u_k$ de terme général $u_k = 1/k - \ln(1 + 1/k)$ (défini pour $k \geq 1$).

Retrouver la somme de la série $\sum_{n=1}^{+\infty} (-1)^{n+1}/n$ en utilisant l'existence de la limite précédente.

Règle $n^\alpha u_n$

On suppose toujours ici que $\sum u_n$ est une série à termes positifs.

Critère C. S'il existe un nombre réel $\alpha > 1$ tel que la suite $(n^\alpha u_n)$ reste majorée, alors la série $\sum u_n$ est convergente. S'il existe $\delta > 0$ tels que $nu_n \geq \delta$ pour tout entier n assez grand, la série $\sum u_n$ est divergente.

On obtient encore comme conséquence :

Critère C'. S'il existe $\alpha > 1$ tel que la suite $(n^\alpha u_n)$ tende vers une limite finie, alors la série $\sum u_n$ est convergente. Si la suite (nu_n) tend vers $+\infty$ ou vers une limite > 0 , la série $\sum u_n$ est divergente.

Exemples. Étudier les séries

$$\sum \left(\frac{\ln n}{n^2} \right)_{n \geq 1} ; \quad \sum \left(\frac{1}{\sqrt{n}(\ln n)^2} \right)_{n \geq 1}.$$

Exemple où le critère **C** ne donne rien : la série de Bertrand $\sum \frac{1}{n(\ln n)^\alpha}$, à traiter par comparaison avec une intégrale.

Exercice. Forme faible de la formule de Stirling. Montrer que

$$x_n = \frac{n!}{n^{n+1/2} e^{-n}}$$

tend vers une limite non nulle ℓ (en fait $\ell = \sqrt{2\pi}$, mais on ne le montrera pas ici).

En passant au log on se ramène à montrer que

$$\left(n + \frac{1}{2} \right) \ln n - n - \ln(n!)$$

tend vers une limite réelle. On pose $g(x) = (x - 1/2) \ln x - x$ et on montre que $g(x+1) - g(x) - \ln x \sim 1/(12x^2)$ quand $x \rightarrow +\infty$. On en déduit la convergence de la série de terme général $u_k = g(k+1) - g(k) - \ln k$, $k \geq 1$. L'étude des sommes partielles de cette série donne le résultat.

2.5. Séries absolument convergentes

Définition 2.5.1. On dit que la série à termes réels ou complexes $\sum u_n$ est *absolument convergente* si la série des modules $\sum |u_n|$ est convergente.

Le *critère de Cauchy* pour la convergence des séries à termes réels ou complexes consiste à exprimer que la suite (U_n) des sommes partielles est de Cauchy. La série $\sum u_n$ vérifie donc le critère de Cauchy si pour tout $\varepsilon > 0$, il existe un entier $N = N(\varepsilon)$ tel que pour tous entiers m, n tels que $N \leq m < n$ on ait

$$|U_n - U_m| = |u_{m+1} + \cdots + u_n| < \varepsilon.$$

Puisque la suite (U_n) converge si et seulement si elle est une *suite* de Cauchy (théorème 1.3.1), on voit qu'une série à termes réels ou complexes est convergente si et seulement si elle vérifie le critère de Cauchy.

Théorème 2.5.1. *Toute série absolument convergente $\sum u_n$ est convergente, et*

$$\left| \sum_{n=0}^{+\infty} u_n \right| \leq \sum_{n=0}^{+\infty} |u_n|.$$

Cette dernière inégalité peut être vue comme une forme d'« inégalité triangulaire infinie ».

Démonstration. Posons

$$U_n = u_0 + \cdots + u_n; \quad V_n = |u_0| + \cdots + |u_n|.$$

Supposons $m < n$. On a

$$|U_n - U_m| = |u_{m+1} + \cdots + u_n| \leq |u_{m+1}| + \cdots + |u_n| = V_n - V_m.$$

Puisque la série $\sum |u_n|$ converge, la suite (V_n) est convergente, donc de Cauchy, ce qui entraîne que (U_n) est de Cauchy, donc convergente. De plus $|U_n| \leq V_n$ pour tout $n \geq 0$ par l'inégalité triangulaire, d'où l'inégalité entre le module de la somme de la série et la somme de la série des modules, par passage à la limite des inégalités larges.

Indiquons une deuxième démonstration que nous qualifierons de « piétonne ». Pour chaque entier $n \geq 0$, posons $u_n = v_n - w_n$, où on a posé $v_n = u_n, w_n = 0$ dans le cas $u_n > 0$ et $v_n = 0, w_n = -u_n$ dans le cas $u_n \leq 0$. On remarque que $v_n, w_n \geq 0$ et $v_n, w_n \leq v_n + w_n = |u_n|$. La convergence de la série $\sum |u_n|$ implique donc par le théorème de comparaison 2.3.2 la convergence des deux séries $\sum v_n$ et $\sum w_n$, qui donne par différence la convergence de $\sum u_n$.

Exemples. La série géométrique complexe $\sum z^n$ est absolument convergente pour $|z| < 1$. La série exponentielle complexe $\sum z^n/n!$ est absolument convergente pour tout $z \in \mathbb{C}$. La série $\sum (-1)^{n+1}/n$ est convergente, mais pas absolument convergente (on l'appelle *série harmonique alternée*).

Sommation par paquets des séries absolument convergentes

Si $\sum u_n$ est une série convergente, on peut être amené à s'intéresser à la série des termes pairs,

$$u_0 + u_2 + \cdots + u_{2n} + \cdots$$

Il faut voir que cette nouvelle série peut diverger (c'est le cas par exemple pour la série harmonique alternée mentionnée ci-dessus). Si elle converge on dira qu'on a fait la somme du « paquet » des termes pairs. Si $\sum u_n$ est absolument convergente, la série des termes pairs est encore absolument convergente, donc convergente.

Soit $\sum u_n$ une série absolument convergente à termes réels ou complexes, et soit A un sous-ensemble de \mathbb{N} ; on définit une suite $(\chi_A(n))$ de la façon suivante : $\chi_A(n) = 1$ si $n \in A$ et 0 sinon. On a $|\chi_A(n)u_n| \leq |u_n|$, donc la série $\sum \chi_A(n)u_n$ est absolument convergente. On pose

$$\sum_{n \in A} u_n = \sum_{n=0}^{+\infty} \chi_A(n) u_n.$$

On dira qu'on a effectué la somme du « paquet » A de la série. Dans ce paragraphe, on posera pour simplifier $\sigma(A) = \sum_{n \in A} u_n$. Si A et B sont deux sous-ensembles de \mathbb{N} disjoints, on a $\chi_{A \cup B}(n) = \chi_A(n) + \chi_B(n)$ pour tout $n \geq 0$, donc $\sigma(A \cup B) = \sigma(A) + \sigma(B)$. Cette propriété d'additivité s'étend à toute famille finie A_0, A_1, \dots, A_k de paquets disjoints. On va maintenant étudier ce qui se passe pour une famille *infinie* de paquets disjoints.

Supposons fixée une série $\sum u_n$ absolument convergente à termes réels ou complexes. On posera donc $\sigma(A) = \sum_{n \in A} u_n$ et de la même façon $\tau(A) = \sum_{n \in A} |u_n|$, pour tout $A \subset \mathbb{N}$. Pour tout A on a d'après le théorème 2.5.1 l'inégalité

$$|\sigma(A)| \leq \tau(A).$$

Si $A \subset B$, il est clair que $\chi_A(n) \leq \chi_B(n)$ pour tout $n \geq 0$, donc

$$0 \leq \tau(A) \leq \tau(B) \leq \tau(\mathbb{N}) = \sum_{n=0}^{+\infty} |u_n| < +\infty.$$

Théorème 2.5.2. *Soit $\sum v_n$ une série à termes positifs; supposons choisie une suite $(A_k)_{k \geq 0}$ de « paquets » deux à deux disjoints qui recouvre \mathbb{N} , et telle que $\tau_k = \sum_{n \in A_k} v_n$ converge pour tout entier $k \geq 0$. Alors la série $\sum \tau_k$ est de même nature que la série $\sum v_n$, et dans le cas où ces deux séries convergent,*

$$\sum_{n=0}^{+\infty} v_n = \sum_{k=0}^{+\infty} \left(\sum_{n \in A_k} v_n \right).$$

Démonstration. Posons $C = \{0, \dots, N\}$; chacun des entiers $j \leq N$ appartient à un certain ensemble A_{k_j} . Si on pose $K = \max\{k_0, k_1, \dots, k_N\}$ on voit que l'ensemble $A = A_0 \cup A_1 \cup \dots \cup A_K$ contient tous les éléments de C , donc

$$\tau(C) = v_0 + \dots + v_N \leq \tau(A) = \sum_{k=0}^K \tau_k \leq \sum_{k=0}^{+\infty} \tau_k.$$

Si $\Sigma = \sum_{k=0}^{+\infty} \tau_k < +\infty$, il en résulte que la suite des sommes partielles (V_N) est majorée par Σ , donc la série $\sum v_n$ converge et $\sum_{n=0}^{+\infty} v_n \leq \Sigma$. Inversement on a déjà observé que

$$\tau_0 + \dots + \tau_K = \tau(A_0 \cup \dots \cup A_K) \leq \tau(\mathbb{N}) = \sum_{n=0}^{+\infty} v_n$$

pour tout K , ce qui donne l'autre direction.

Remarque. Supposons donnée une famille $(u_{n,k})$ à deux indices de nombres réels positifs. On peut se demander ce que pourrait signifier la somme $\sum_{n,k=0}^{+\infty} u_{n,k}$. Il y a au moins deux solutions naturelles : on pourrait regrouper tous ces nombres par paquets A_n comprenant tous les couples (n,k) pour un n fixé, k variant dans \mathbb{N} , ou bien par paquets B_k regroupant tous les couples (n,k) pour un k fixé. Le résultat précédent nous permet de comprendre que ces deux solutions donnent le même résultat (ainsi que n'importe quel découpage de l'ensemble des couples d'ailleurs). Attention : ça ne marche plus du tout si le signe du terme général $u_{n,k}$ change.

Théorème 2.5.3. Soit $\sum u_n$ une série absolument convergente à termes réels ou complexes ; supposons choisie une suite $(A_k)_{k \geq 0}$ de « paquets » deux à deux disjoints qui recouvre \mathbb{N} . Alors la série $\sum \sigma_k$ de terme général $\sigma_k = \sum_{n \in A_k} u_n$ est absolument convergente et

$$\sum_{n=0}^{+\infty} u_n = \sum_{k=0}^{+\infty} \left(\sum_{n \in A_k} u_n \right).$$

Démonstration. Il suffit de traiter le cas réel. Comme dans la démonstration dite piétonne, écrivons $u_n = v_n - w_n$, avec $v_n, w_n \geq 0$ et $v_n + w_n = |u_n|$. Puisque $v_n, w_n \leq |u_n|$, les deux séries à termes positifs $\sum v_n$ et $\sum w_n$ sont convergentes, ce qui permet d'appliquer le résultat précédent. Le résultat en découle.

Changement de l'ordre des termes

Proposition 2.5.1. Soit π une bijection de \mathbb{N} sur \mathbb{N} et soit $\sum u_n$ une série absolument convergente ; la série permutée $\sum u_{\pi(n)}$ est absolument convergente et

$$\sum_{n=0}^{+\infty} u_{\pi(n)} = \sum_{n=0}^{+\infty} u_n.$$

Démonstration. C'est un cas particulier du principe des paquets. Il suffit de poser $A_k = \{\pi(k)\}$ pour tout entier $k \geq 0$.

ATTENTION. Ce résultat est FAUX pour les séries qui ne sont pas absolument convergentes. Par exemple, on peut trouver une permutation de la série $\sum_{n \geq 1} (-1)^{n+1}/n$ qui est divergente. On peut aussi trouver des permutations telles que la nouvelle série soit convergente, mais avec une somme différente.

Exercice (difficile). Soit $\sum u_n$ une série convergente telle que $\sum |u_n| = +\infty$; pour tout nombre réel x , il existe une permutation π des entiers telle que $\sum_{n=0}^{+\infty} u_{\pi(n)} = x$.

Produit de séries numériques absolument convergentes

Étant données deux séries $\sum u_n$ et $\sum v_n$, on introduit une nouvelle série $\sum w_n$ appelée série produit et dont le terme général w_n est défini par

$$w_n = \sum_{k=0}^n u_k v_{n-k}.$$

Théorème 2.5.4. Si les deux séries $\sum u_n$ et $\sum v_n$ sont absolument convergentes, la série produit est absolument convergente, et

$$\sum_{n=0}^{+\infty} w_n = \left(\sum_{n=0}^{+\infty} u_n \right) \left(\sum_{n=0}^{+\infty} v_n \right).$$

Démonstration. On va appliquer deux fois la sommation par paquets à la série obtenue en alignant les composants de w_0, w_1, w_2, \dots

$$t_0 = u_0 v_0; t_1 = u_0 v_1, t_2 = u_1 v_0; t_3 = u_0 v_2, t_4 = u_1 v_1, t_5 = u_2 v_0; \dots$$

(si on veut une définition précise du terme général t_n de cette série, on peut dire que $t_n = u_q v_{p-q}$ si $n = p(p+1)/2 + q$, $p \geq 0$ et $0 \leq q \leq p$). On définit les premiers paquets de sommation (A_k) qui correspondent aux composants de w_k , c'est-à-dire

$$A_k = \{k(k+1)/2, \dots, (k+1)(k+2)/2 - 1\}.$$

La somme sur A_k donne donc w_k . La deuxième famille de paquets (B_k) est définie en prenant tous les termes t_n qui contiennent le facteur u_k , et la somme du paquet B_k est égale à $u_k(\sum_{n=0}^{+\infty} v_n)$. En appliquant deux fois la sommation par paquets, on a

$$\sum_{n=0}^{+\infty} w_n = \sum_{n=0}^{+\infty} t_n = \left(\sum_{k=0}^{+\infty} u_k \right) \left(\sum_{k=0}^{+\infty} v_k \right).$$

Mais il faut aussi justifier que la série $\sum t_n$ est absolument convergente. Si on effectue la somme des $|t_n|$ du paquet B_k , on obtient

$$\tau_k = |u_k| \sum_{n=0}^{+\infty} |v_n|$$

qui est finie puisque $\sum |v_n| < +\infty$, et ensuite

$$\sum |t_n| = \sum_k \tau_k = \left(\sum_{k=0}^{+\infty} |u_k| \right) \left(\sum_{n=0}^{+\infty} |v_n| \right) < +\infty$$

d'après le théorème 2.5.2.

Exercice.

1. Donner une formule explicite pour les ensembles B_k (si on n'a pas peur des séries doubles, on évite ces problèmes assez artificiels de numérotage des termes (t_n) : on posera simplement $t_{n,k} = u_n v_k$, et le découpage qui donne le terme général de la série produit correspond au découpage de l'ensemble de tous les couples (n, k) selon la valeur de la somme $n+k$).

2. Montrer que le théorème du produit de séries reste vrai si $\sum u_n$ est absolument convergente et $\sum v_n$ simplement convergente (il faut donner une démonstration complètement différente de la précédente).

La fonction exponentielle complexe

On pose pour tout $z \in \mathbb{C}$

$$\varphi(z) = 1 + z + \frac{z^2}{2!} + \dots + \frac{z^n}{n!} + \dots$$

On a déjà vu que $\varphi(x) = e^x$ pour tout $x \in \mathbb{R}$. Posons pour $a, b \in \mathbb{C}$

$$u_n = \frac{a^n}{n!}, \quad v_n = \frac{b^n}{n!}.$$

On vérifie avec la formule du binôme que

$$w_n = \sum_{k=0}^n u_k v_{n-k} = \frac{(a+b)^n}{n!}$$

ce qui montre par le théorème des produits de séries que $\varphi(a+b) = \varphi(a)\varphi(b)$ pour tous nombres complexes a, b . Par ailleurs, on a vu avec la formule de Taylor, pour tout $y \in \mathbb{R}$

$$\sin y = y - \frac{y^3}{3!} + \frac{y^5}{5!} + \cdots + (-1)^n \frac{y^{2n+1}}{(2n+1)!} + \cdots = \operatorname{Im} \varphi(iy)$$

$$\cos y = 1 - \frac{y^2}{2!} + \frac{y^4}{4!} + \cdots + (-1)^n \frac{y^{2n}}{(2n)!} + \cdots = \operatorname{Re} \varphi(iy)$$

ce qui montre que pour tout $y \in \mathbb{R}$

$$\varphi(iy) = \cos y + i \sin y = e^{iy}.$$

Finalement, si $z = x + iy$, on a

$$\varphi(z) = \varphi(x)\varphi(iy) = e^x e^{iy} = e^x (\cos y + i \sin y).$$

Il est donc tout à fait raisonnable de poser pour tout nombre complexe z

$$e^z = \varphi(z) = 1 + z + \frac{z^2}{2!} + \cdots + \frac{z^n}{n!} + \cdots$$

On peut également prolonger aux nombres complexes les fonctions ch et sh, en posant

$$\operatorname{ch}(z) = \frac{e^z + e^{-z}}{2} \quad \text{et} \quad \operatorname{sh}(z) = \frac{e^z - e^{-z}}{2}$$

pour tout $z \in \mathbb{C}$. On peut de même prolonger les fonctions sin et cos aux valeurs complexes, en utilisant les développements en série ci-dessus. On remarque que $\operatorname{ch}(ix) = \cos(x)$, $\operatorname{sh}(ix) = i \sin(x)$ pour tout nombre réel (ou complexe) x .

2.6. Séries semi-convergentes. Séries alternées

Considérons ici des groupements de termes par paquets qui sont des *intervalles* : on se donne une suite $n_0 < n_1 < \dots$ et on considère les ensembles $A_0 = \{0, 1, \dots, n_0\}$, $A_1 = \{n_0 + 1, \dots, n_1\}$ et plus généralement pour tout $k \geq 1$ on pose $A_k = \{n_{k-1} < n \leq n_k\}$. On pose aussi

$$\sigma_k = \sum_{n \in A_k} u_n = u_{n_{k-1}+1} + \cdots + u_{n_k},$$

$$\Sigma_k = \sigma_0 + \cdots + \sigma_k = u_0 + \cdots + u_{n_k} = U_{n_k}.$$

Si la série $\sum u_n$ converge, la série $\sum \sigma_k$ converge aussi puisque la suite (Σ_k) de ses sommes partielles est une sous-suite de la suite (U_n) des sommes partielles de $\sum u_n$. L'inverse n'est pas toujours vrai, mais c'est l'inverse qui serait intéressant. On va ajouter des hypothèses pour pouvoir l'obtenir.

Théorème 2.6.1. Groupements de longueurs bornées. *On suppose données une suite $n_0 < n_1 < \dots$ d'entiers et une série $\sum u_n$ telles que*

— les longueurs des intervalles (A_k) sont bornées, c'est-à-dire qu'il existe M tel que pour tout k , on ait $n_k - n_{k-1} \leq M$.

— le terme général u_n tend vers 0 ;

alors la série $\sum u_n$ converge si et seulement si la série $\sum \sigma_k$ converge.

Démonstration. Il reste seulement à montrer que si $\sum \sigma_k$ est convergente, la série $\sum u_n$ est convergente aussi. Posons $\Sigma = \sum_{k=0}^{+\infty} \sigma_k$ et montrons que la suite (U_n) converge vers Σ . Soit $\varepsilon > 0$ donné; on peut trouver un entier k_0 tel que $|\Sigma - \Sigma_k| < \varepsilon/2$ pour tout $k \geq k_0$. Puisque $u_n \rightarrow 0$, on peut ensuite trouver un entier k_1 tel que $k_1 \geq k_0$, et tel que $|u_n| < \varepsilon/(2M)$ pour tout $n > n_{k_1}$. Pour tout entier $n \geq N = n_{k_1} + 1$, on vérifie que $|U_n - \Sigma| < \varepsilon$. En effet, il existe un entier $k \geq k_1$ tel que $n_k < n \leq n_{k+1}$, et alors

$$|U_n - \Sigma_k| = |U_n - U_{n_k}| = |u_{n_k+1} + \dots + u_n| \leq M \frac{\varepsilon}{2M} = \frac{\varepsilon}{2}.$$

Applications.

1. La série $\sum (-1)^{n+1}/\sqrt{n}$ converge : en groupant par deux, on obtient une série à termes positifs de terme général équivalent à $cn^{-3/2}$, donc convergente.

2. Une permutation des termes d'une série simplement convergente peut changer la somme de la série (ou bien la rendre divergente).

La série

$$1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \dots$$

est une permutation de la série harmonique alternée. En groupant en paquets de longueur 2, 1, 2, 1, ..., on trouve exactement la moitié de la série harmonique, et la nouvelle somme est donc la moitié de l'ancienne.

Exercice. Exprimer la permutation de \mathbb{N} qui transforme la série harmonique alternée en la série précédente. Retrouver la somme de la série précédente en utilisant l'exercice sur la constante d'Euler. Montrer que pour tout $x \in \mathbb{R}$ on peut trouver une variante des permutations précédentes qui donnera la somme x pour la permutée de la série harmonique alternée correspondante (en jouant simplement sur la proportion des termes positifs parmi les n premiers de la série permutée).

Exemple de séries à terme général équivalent non positif de nature différente. Étudier les séries

$$\sum \frac{(-1)^n}{\sqrt{n+1}}, \quad \sum \ln\left(1 + \frac{(-1)^n}{\sqrt{n+1}}\right).$$

On a vu que la première série est convergente. Pour la seconde : avec un DL de $\ln(1+u)$ à l'ordre 3, on fait apparaître la série convergente $\sum (-1)^n/\sqrt{n}$, une série absolument convergente de terme général équivalent à $cn^{-3/2}$ et la série divergente $\sum 1/n$. Ceci montre que la deuxième série diverge.

ATTENTION. On voit donc que deux séries de termes généraux équivalents peuvent être de nature différente.

Théorème 2.6.2 (des séries alternées). Soit $\sum u_n$ une série à termes réels telle que $u_{n+1}u_n < 0$ pour tout $n \geq 0$, et telle que $|u_n|$ soit décroissant vers 0. La série $\sum u_n$ converge.

Démonstration. Supposons par exemple $u_0 > 0$; on constate que $U_{2n-1} \leq U_{2n+1} \leq U_{2n+2} \leq U_{2n}$ pour tout entier $n \geq 1$; la sous-suite impaire de la suite (U_n) est croissante et majorée, et la sous-suite paire est décroissante et minorée, donc ces deux suites sont convergentes. Puisque $U_{2n} - U_{2n+1} = -u_{2n+1}$ tend vers 0, les deux sous-suites ont la même limite et la suite (U_n) converge.

On peut aussi utiliser un groupement par paquets de longueur deux. On se ramène donc à une série $\sum \sigma_k$ de terme général $\sigma_k = u_{2k} + u_{2k+1} \geq 0$. Il suffit de montrer que ses sommes partielles sont bornées. Or on a, en posant $v_n = |u_n|$

$$\Sigma_k = (v_0 - v_1) + (v_2 - v_3) + \dots + (v_{2k} - v_{2k+1}) =$$

$$= v_0 - (v_1 - v_2) - (v_3 - v_4) - \cdots - (v_{2k-1} - v_{2k}) - v_{2k+1} \leq v_0 = u_0.$$

Exemples. La série

$$\sum \frac{(-1)^n}{n^\beta}$$

converge pour tout $\beta > 0$. Montrer que l'intégrale

$$\int_0^{+\infty} \frac{\sin x}{x} dx$$

est convergente en étudiant la série de terme général $u_n = \int_{n\pi}^{(n+1)\pi} (\sin x)/x dx$.

La transformation d'Abel

Cette transformation, qui ressemble d'une certaine façon à l'intégration par parties, consiste à remarquer que

$$U_0 v_0 + \cdots + U_n v_n = U_n V_n - (u_1 V_0 + u_2 V_1 + \cdots + u_n V_{n-1}),$$

où on a posé comme d'habitude $U_n = u_0 + \cdots + u_n$ et $V_n = v_0 + \cdots + v_n$. On obtient immédiatement le principe général suivant :

Proposition 2.6.1. *Si la suite $(U_n V_n)$ tend vers une limite finie, les deux séries $\sum U_n v_n$ et $\sum u_{n+1} V_n$ sont de même nature.*

Théorème 2.6.3. *Soit $\sum u_n$ une série à termes réels ou complexes de la forme $\sum \lambda_n v_n$, où $\sum |\lambda_{n+1} - \lambda_n| < +\infty$; on suppose de plus*

ou bien que la série $\sum v_n$ converge,

ou bien que $\lim \lambda_n = 0$ et que la suite (V_n) des sommes partielles de la série $\sum v_n$ est bornée ($V_n = v_0 + \cdots + v_n$).

Dans chacun de ces deux cas, la série $\sum \lambda_n v_n$ est convergente.

On remarquera que l'hypothèse $\sum |\lambda_{n+1} - \lambda_n| < +\infty$ est satisfaite dans le cas particulier où (λ_n) est monotone et bornée.

Démonstration. On écrit

$$\lambda_0 v_0 + \cdots + \lambda_n v_n = \lambda_0 V_0 + \lambda_1 (V_1 - V_0) + \cdots + \lambda_{n-1} (V_{n-1} - V_{n-2}) + \lambda_n (V_n - V_{n-1}) =$$

$$(\lambda_0 - \lambda_1) V_0 + \cdots + (\lambda_{n-1} - \lambda_n) V_{n-1} + \lambda_n V_n.$$

La série $\sum (\lambda_n - \lambda_{n+1}) V_n$ est absolument convergente (parce que (V_n) est bornée dans les deux cas) et la suite $(\lambda_n V_n)$ est convergente dans chacun des deux cas proposés (la suite (λ_n) est convergente puisque la série $\sum (\lambda_{n+1} - \lambda_n)$ converge).

Exemple. Étudier la série

$$\sum \frac{e^{in\alpha}}{n}$$

où α est réel.

Exercices sur le chapitre 2

Étudier les séries

$$\begin{aligned} & \sum \int_0^1 x^n (\ln(x))^2 dx; \quad \sum \frac{(-1)^n}{n - \ln n}; \quad \sum e^{-\sqrt{n}}; \\ & \sum e^{-n^2} \int_0^n e^{u^2} du; \quad \sum \frac{1}{(1 + 1/\sqrt{n})^n}; \quad \sum \frac{\cos(\ln n)}{n}; \\ & \sum \sin(n! \pi e); \quad \sum \frac{n!}{n^n}; \quad \sum \ln\left(1 + \frac{(-1)^n}{n^\alpha}\right). \end{aligned}$$

3. Intégrale de Riemann

Supposons que l'on veuille calculer l'intégrale

$$\int_0^1 e^{-t^2} dt.$$

Très facile, direz-vous : il suffit de prendre une primitive F de la fonction $f : x \rightarrow e^{-x^2}$ et de calculer $F(1) - F(0)$. Seulement voilà, il y a un problème : on ne sait pas exprimer une primitive de cette fonction f à partir des fonctions déjà connues.

Mais pourtant, direz-vous encore, on a appris que toute fonction continue admet des primitives ; il y a quelque chose qui cloche là-dedans.

Justement, c'est bien vrai qu'il existe une primitive, mais en gros la seule façon de la calculer est de poser

$$F(x) = \int_0^x e^{-t^2} dt$$

et on tourne en rond !

La moralité est la suivante : ce qu'on apprend en terminale sur le calcul d'intégrales à l'aide de primitives est très utile dans beaucoup de situations, chaque fois que l'on sait exprimer une primitive de la fonction à intégrer à partir de fonctions déjà connues ; mais pour l'essentiel, il s'agit d'une escroquerie : la notion fondamentale à définir est la notion d'intégrale (de Riemann), l'existence des primitives en sera une conséquence.

3.1. Intégrale de Riemann sur un intervalle fermé borné $[a, b]$.

Le premier objectif est d'introduire une classe de fonctions très simples, les *fonctions en escalier*, et de définir leur intégrale. La définition est plus commode si on met en place un minimum de terminologie au préalable : soient $a \leq b$ deux nombres réels ; on appelle *subdivision* π de l'intervalle $[a, b]$ la donnée d'un entier $n \geq 1$ et d'une suite de points $x_0 = a < x_1 < \dots < x_n = b$. On dira que les intervalles $]x_0, x_1[$, $]x_1, x_2[$, \dots , $]x_{n-1}, x_n[$ sont les intervalles (ouverts) de la subdivision, et on dira que x_0, x_1, \dots, x_n sont les *points de subdivision*.

On dit que la subdivision $\sigma = (z_0 = a < z_1 < \dots < z_p = b)$ est *plus fine* que la subdivision $\pi = (x_0 = a < x_1 < \dots < x_n = b)$ si chacun des points x_i , $i = 0, \dots, n$ fait partie de l'ensemble $\{z_0, \dots, z_p\}$ des points de subdivision de σ , c'est-à-dire que chaque intervalle $]x_{i-1}, x_i[$ de la subdivision π est « découpé » en intervalles plus petits $]z_{j-1}, z_j[$ dans la subdivision σ .

Étant données deux subdivisions π et ρ quelconques, on peut toujours trouver une subdivision σ plus fine que π et que ρ en même temps : il suffit de prendre un ensemble $z_0 < \dots < z_p$ de points de subdivision qui contienne tous les points de subdivision de π et de ρ .

Définition 3.1.1. On dit qu'une fonction réelle φ définie sur $[a, b]$ est une *fonction en escalier* s'il existe une subdivision $\pi = (x_0 = a < x_1 < \dots < x_n = b)$ de l'intervalle telle que φ soit constante sur chacun des intervalles ouverts $]x_{i-1}, x_i[$ de la subdivision, pour tout $i = 1, \dots, n$.

Si cette condition est réalisée, on dira que π est une subdivision *adaptée* à la fonction en escalier φ . Bien entendu, toute subdivision σ plus fine que π sera encore adaptée à φ , puisque chaque intervalle de la subdivision plus fine est contenu dans un intervalle de π , sur lequel la fonction φ est constante. Si ψ est une deuxième fonction en escalier sur $[a, b]$ et ρ une subdivision adaptée à ψ , on peut trouver une subdivision σ plus fine que π et que ρ ; cette subdivision σ sera adaptée à la fois à φ et à ψ . Les deux fonctions φ et ψ seront donc constantes sur tous les intervalles ouverts de la subdivision σ , ce qui implique que toute fonction combinaison linéaire $\lambda\varphi + \mu\psi$ de ces deux fonctions en escalier sera elle aussi constante sur les intervalles de σ :

si φ et ψ sont en escalier sur $[a, b]$, toutes les combinaisons linéaires $\lambda\varphi + \mu\psi$ sont en escalier ; l'ensemble des fonctions réelles en escalier sur $[a, b]$ est un espace vectoriel réel.

Exercice. Pour tous c, d tels que $a \leq c \leq d \leq b$ désignons par $x \rightarrow \chi_{[c,d]}(x)$ la fonction égale à 1 si $c \leq x \leq d$ et à 0 sinon. Montrer que l'espace des fonctions en escalier sur $[a, b]$ est l'espace vectoriel engendré par les fonctions $\chi_{(c,d)}$, lorsque c et d varient.

Plus précisément, si $\pi = (x_0 = a < \dots < x_n = b)$ est adaptée à la fonction en escalier φ , alors φ est combinaison linéaire des fonctions $\chi_{[x_i, x_j]}$, $0 \leq i \leq j \leq n$.

Pour la même raison le maximum $\max(\varphi, \psi)$ — ou le minimum $\min(\varphi, \psi)$, ou plus généralement n'importe quelle fonction $x \rightarrow F(\varphi(x), \psi(x))$ — de deux fonctions en escalier sur $[a, b]$ est une fonction en escalier sur $[a, b]$.

Soient φ une fonction en escalier sur $[a, b]$ et $\pi = (x_0 = a < x_1 < \dots < x_n = b)$ une subdivision adaptée à φ ; choisissons un point quelconque $\xi_i \in]x_{i-1}, x_i[$, pour tout $i = 1, \dots, n$. La fonction φ sera alors constamment égale à $c_i = \varphi(\xi_i)$ sur l'intervalle $]x_{i-1}, x_i[$. Il est naturel d'essayer de définir l'intégrale de φ par l'expression

$$E(\varphi, \pi) = \sum_{i=1}^n (x_i - x_{i-1}) \varphi(\xi_i) = \sum_{i=1}^n (x_i - x_{i-1}) c_i.$$

Pour être tout à fait rigoureux, il faut montrer que l'expression précédente ne dépend que de φ , et pas de la subdivision π adaptée à φ . C'est assez facile à comprendre mais terriblement pénible à écrire (et à lire!). Pour le faire on remarque que $E(\varphi, \pi)$ ne change pas si on remplace π par une subdivision plus fine $\sigma = (z_0 = a < z_1 < \dots < z_p = b)$. On peut passer de π à σ par étapes, en ajoutant à chaque étape un seul nouveau point de subdivision. Pour montrer que $E(f, \pi)$ ne varie pas dans cette chaîne d'opérations, il suffit de le vérifier pour une étape. Supposons donc que l'on passe de π à σ en une seule étape, en ajoutant un unique nouveau point z de subdivision, et que $x_{i_0-1} < z < x_{i_0}$. La fonction φ est constante égale à c_{i_0} sur l'intervalle $]x_{i_0-1}, x_{i_0}[$, donc aussi sur chacun des deux nouveaux intervalles $]x_{i_0-1}, z[$ et $]z, x_{i_0}[$. On aura donc

$$\begin{aligned} E(\varphi, \sigma) &= \sum_{j=1}^{i_0-1} (x_j - x_{j-1}) c_j + (z - x_{i_0-1}) c_{i_0} + (x_{i_0} - z) c_{i_0} + \sum_{j=i_0+1}^n (x_j - x_{j-1}) c_j = \\ &= \sum_{j=1}^{i_0-1} (x_j - x_{j-1}) c_j + (x_{i_0} - x_{i_0-1}) c_{i_0} + \sum_{j=i_0+1}^n (x_j - x_{j-1}) c_j = E(\varphi, \pi). \end{aligned}$$

Si maintenant ρ est une autre subdivision adaptée à φ , on peut considérer une subdivision σ plus fine que π et ρ , et d'après ce qui précède,

$$E(\varphi, \pi) = E(\varphi, \sigma) = E(\varphi, \rho).$$

Définition 3.1.2. Soit φ une fonction réelle en escalier définie sur l'intervalle $[a, b]$ et soit $\pi = (x_0 = a < x_1 < \dots < x_n = b)$ une subdivision adaptée à φ , c'est-à-dire que φ est constante égale à c_i sur chacun des intervalles ouverts $]x_{i-1}, x_i[$ de la subdivision, pour $i = 1, \dots, n$. On définit l'intégrale de φ sur $[a, b]$ par l'expression

$$\int_a^b \varphi(t) dt = \sum_{i=1}^n (x_i - x_{i-1}) c_i.$$

Parfois, on notera simplement $\int_a^b \varphi$ pour abrégier.

Proposition 3.1.1. Linéarité et majoration de l'intégrale des fonctions en escalier. Si φ et ψ sont deux fonctions réelles en escalier sur $[a, b]$, on a

$$\forall \lambda, \mu \in \mathbb{R}, \quad \int_a^b (\lambda\varphi + \mu\psi)(t) dt = \lambda \int_a^b \varphi(t) dt + \mu \int_a^b \psi(t) dt;$$

$$\left(\varphi \leq \psi \text{ sur } [a, b]\right) \Rightarrow \left(\int_a^b \varphi(t) dt \leq \int_a^b \psi(t) dt\right); \quad \left|\int_a^b \varphi(t) dt\right| \leq \int_a^b |\varphi(t)| dt.$$

Démonstration. Il suffit de choisir une subdivision σ qui soit adaptée à la fois à la fonction φ et à ψ . Les deux premières propriétés sont alors très faciles à vérifier sur la formule $E(\varphi, \sigma)$ qui définit l'intégrale. La troisième s'obtient en encadrant f entre $-|f|$ et $|f|$, et en appliquant les deux premières propriétés, ou bien directement par l'inégalité triangulaire appliquée à l'expression $E(\varphi, \pi)$.

Remarque. En dehors de la deuxième propriété qui concerne l'ordre, tout marcherait aussi bien pour des fonctions en escalier à valeurs vectorielles (voir la section suivante), que ce soit pour la définition de l'intégrale ou pour sa linéarité. On pourrait aussi adapter la troisième propriété de la proposition précédente pour montrer que la norme de l'intégrale vectorielle est majorée par l'intégrale de la norme.

Intégrabilité par encadrement

Soit f une fonction réelle bornée sur $[a, b]$, disons bornée par le nombre C , c'est-à-dire que $|f| \leq C$ sur $[a, b]$; on peut alors trouver des fonctions en escalier $\varphi_1 \leq f$, par exemple la fonction constante $\varphi_1 = -C$, et également des fonctions en escalier $\varphi_2 \geq f$, par exemple $\varphi_2 = C$. Fixons une fonction en escalier $\varphi_2 \geq f$. On a alors pour toute fonction en escalier $\varphi_1 \leq f$ l'inégalité $\varphi_1 \leq \varphi_2$, donc $\int_a^b \varphi_1(t) dt \leq \int_a^b \varphi_2(t) dt$. L'ensemble des intégrales $\int_a^b \varphi_1(t) dt$ est donc (non vide et) majoré par $\int_a^b \varphi_2(t) dt$; sa borne supérieure M_1 vérifie donc l'inégalité $M_1 \leq \int_a^b \varphi_2(t) dt$. Si maintenant on fait varier φ_2 , on voit que la borne inférieure m_2 de l'ensemble des valeurs des intégrales $\int_a^b \varphi_2(t) dt$ (où on suppose toujours $\varphi_2 \geq f$) vérifie $m_2 \geq M_1$. La fonction f est sympathique lorsque ces deux nombres sont égaux : c'est dans ce cas qu'on dit que f est intégrable au sens de Riemann.

Définition 3.1.3. Soit f une fonction bornée à valeurs réelles définie sur $[a, b]$; on dit que la fonction f est Riemann-intégrable sur $[a, b]$ si pour tout nombre réel $\varepsilon > 0$, il existe deux fonctions en escalier φ_1 et φ_2 sur $[a, b]$ telles que $\varphi_1 \leq f \leq \varphi_2$ sur $[a, b]$ et

$$\int_a^b \varphi_2(t) dt - \int_a^b \varphi_1(t) dt = \int_a^b (\varphi_2 - \varphi_1)(t) dt < \varepsilon.$$

Cette définition revient exactement à dire que les deux nombres M_1 et m_2 introduits ci-dessus sont égaux. On définit alors l'intégrale de f comme étant la valeur commune,

$$\int_a^b f(t) dt = \sup \left\{ \int_a^b \varphi_1(t) dt : \varphi_1 \leq f \right\} = \inf \left\{ \int_a^b \varphi_2(t) dt : \varphi_2 \geq f \right\}$$

où les fonctions φ_1 et φ_2 sont bien entendu choisies en escalier. Parfois, on notera simplement $\int_a^b f$ pour abrégier.

Remarque. Si f est une fonction Riemann-intégrable sur l'intervalle $[a, b]$, elle est bornée sur $[a, b]$ **par définition**.

Remarque idéologique. L'intégrale est une limite de sommes de « petits bouts ». Dans un problème physique, l'intégrale a la même « dimension » (au sens des physiciens) que les petits bouts $f(t) dt$. Un certain nombre d'étudiants se sont laissé enfoncer dans le crâne en terminale l'idée qu'une intégrale représente une surface. Si cette interprétation est utile dans certains cas, elle est bien nuisible à la compréhension de la plupart des intégrales physiques (telles que travail, flux, etc. . .).

Exercice. Soit $A = [0, 1] \cap \mathbb{Q}$ l'ensemble des nombres rationnels de $[0, 1]$; montrer que la fonction indicatrice de A n'est pas R-intégrable (exercice très classique).

Le fait que cette fonction χ_A ne soit pas R-intégrable est un défaut de l'intégrale de Riemann. On a de bonnes raisons de penser que cette fonction devrait avoir une intégrale nulle, dans une bonne théorie de l'intégration. C'est ce qui se passe avec la théorie plus moderne de l'intégrale de Lebesgue.

Intégrabilité par approximation

Pour démontrer certaines propriétés, et pour traiter ensuite le cas vectoriel (section suivante), il est commode de transformer un peu la caractérisation de l'intégrabilité par encadrement pour obtenir une caractérisation *par approximation en escalier*. On obtient une définition équivalente de l'intégrabilité d'une fonction réelle f définie sur $[a, b]$ en disant : pour tout $\varepsilon > 0$, il existe deux fonctions en escalier φ et ψ sur $[a, b]$ telles que

$$|f - \varphi| \leq \psi \quad \text{et} \quad \int_a^b \psi < \varepsilon.$$

Pour passer de la formulation par encadrement, avec deux fonctions φ_1, φ_2 en escalier telles que $\varphi_1 \leq f \leq \varphi_2$ à celle de la ligne précédente, il suffit de poser par exemple $\varphi = \varphi_1$ et $\psi = \varphi_2 - \varphi_1$; dans ce cas $|f - \varphi| = f - \varphi_1 \leq \varphi_2 - \varphi_1 = \psi$. Inversement, étant données φ et ψ qui donnent l'approximation, il suffit de poser $\varphi_1 = \varphi - \psi$, $\varphi_2 = \varphi + \psi$.

Si f est R-intégrable et si on dispose d'une approximation de la forme $|f - \varphi| \leq \psi$, avec φ, ψ en escalier, l'encadrement $\varphi - \psi \leq f \leq \varphi + \psi$, la définition de l'intégrale de f et les propriétés de l'intégrale des fonctions en escalier montrent que $\int_a^b \varphi - \int_a^b \psi \leq \int_a^b f \leq \int_a^b \varphi + \int_a^b \psi$ donc

$$(*) \quad \left| \int_a^b f - \int_a^b \varphi \right| \leq \int_a^b \psi.$$

Exercice. Supposons que la fonction f définie sur $[a, b]$ vérifie la propriété suivante : pour tout $\varepsilon > 0$, il existe deux fonctions R-intégrables g et h sur $[a, b]$ telles que $|f - g| \leq h$ et $\int_a^b h < \varepsilon$. Montrer que f est R-intégrable sur $[a, b]$.

Proposition 3.1.2. Linéarité et majoration. Si f_1 et f_2 sont R-intégrables sur $[a, b]$, leurs combinaisons linéaires sont R-intégrables sur $[a, b]$ et

$$\forall \lambda_1, \lambda_2 \in \mathbb{R}, \quad \int_a^b (\lambda_1 f_1 + \lambda_2 f_2)(t) dt = \lambda_1 \int_a^b f_1(t) dt + \lambda_2 \int_a^b f_2(t) dt.$$

Si f et g sont R-intégrables sur $[a, b]$ et si $f \leq g$ sur $[a, b]$, il en résulte que

$$\int_a^b f(t) dt \leq \int_a^b g(t) dt.$$

Si f est R-intégrable sur $[a, b]$, la fonction $t \rightarrow |f(t)|$ est R-intégrable sur $[a, b]$ et on a l'inégalité

$$\left| \int_a^b f(t) dt \right| \leq \int_a^b |f(t)| dt.$$

Démonstration. Montrons d'abord que $\lambda_1 f_1 + \lambda_2 f_2$ est R-intégrable. Supposons donné $\varepsilon > 0$. Choisissons quatre fonctions en escalier $\varphi_1, \psi_1, \varphi_2, \psi_2$ telles que $|f_j - \varphi_j| < \psi_j$ et $\int_a^b \psi_j < \varepsilon$ pour $j = 1, 2$. On sait d'après la relation (*) que cela implique l'inégalité $|\int_a^b f_j - \int_a^b \varphi_j| \leq \int_a^b \psi_j < \varepsilon$. Posons $f = \lambda_1 f_1 + \lambda_2 f_2$, $\varphi = \lambda_1 \varphi_1 + \lambda_2 \varphi_2$ et d'autre part $\psi = |\lambda_1| \psi_1 + |\lambda_2| \psi_2$. Supposons pour fixer les idées que $|\lambda_1| + |\lambda_2| \leq 1$. On aura

$$\left| \int_a^b f - \int_a^b \varphi \right| = |(\lambda_1 f_1 + \lambda_2 f_2) - (\lambda_1 \varphi_1 + \lambda_2 \varphi_2)| \leq |\lambda_1| \psi_1 + |\lambda_2| \psi_2 = \psi$$

ainsi que

$$\int_a^b \psi = \int_a^b (|\lambda_1| \psi_1 + |\lambda_2| \psi_2) < (|\lambda_1| + |\lambda_2|) \varepsilon \leq \varepsilon,$$

et cette approximation est possible pour tout $\varepsilon > 0$, ce qui montre que $f = \lambda_1 f_1 + \lambda_2 f_2$ est R-intégrable. On sait alors que son intégrale vérifie

$$\left| \int_a^b f - \int_a^b \varphi \right| = \left| \int_a^b f - (\lambda_1 \int_a^b \varphi_1 + \lambda_2 \int_a^b \varphi_2) \right| \leq \int_a^b \psi < \varepsilon$$

et par ailleurs

$$\left| (\lambda_1 \int_a^b f_1 + \lambda_2 \int_a^b f_2) - (\lambda_1 \int_a^b \varphi_1 + \lambda_2 \int_a^b \varphi_2) \right| < (|\lambda_1| + |\lambda_2|) \varepsilon \leq \varepsilon,$$

ce qui montre que

$$\left| \int_a^b (\lambda_1 f_1 + \lambda_2 f_2)(t) dt - (\lambda_1 \int_a^b f_1(t) dt + \lambda_2 \int_a^b f_2(t) dt) \right| < 2\varepsilon$$

pour tout $\varepsilon > 0$, ce qui donne l'égalité voulue.

Terminons avec les deux dernières propriétés. Supposons que $f \leq g$. L'ensemble des fonctions en escalier $\leq f$ est contenu dans celui des fonctions en escalier $\leq g$, donc le sup

$\int_a^b f$ des intégrales de la première famille est inférieur ou égal au $\sup \int_a^b g$ de la deuxième famille. La vérification du dernier point est très simple par approximation : si $|f - \varphi| \leq \psi$ et $\int_a^b \psi < \varepsilon$, on aura aussi

$$||f| - |\varphi|| \leq \psi,$$

donc $|\varphi|$ est une fonction en escalier qui donne l'approximation souhaitée pour la fonction valeur absolue $|f|$. L'inégalité sur les intégrales résulte de la deuxième propriété, appliquée à l'encadrement $-|f| \leq f \leq |f|$.

Exercice. Si $\varphi(x_1, x_2)$ est une fonction lipschitzienne sur \mathbb{R}^2 , et si f, g sont deux fonctions réelles R-intégrables sur $[a, b]$, montrer que $t \rightarrow \varphi(f(t), g(t))$ est R-intégrable sur $[a, b]$. Cas particulier : $\max(f, g), \min(f, g)$.

Proposition 3.1.3. Soient f_1 et f_2 deux fonctions Riemann-intégrables sur $[a, b]$; la fonction produit $f_1 f_2$ est Riemann-intégrable sur $[a, b]$.

Démonstration. Soit M une borne pour les fonctions $|f_1|$ et $|f_2|$, et soit φ_1 une approximation en escalier de f_1 , telle que $|f_1 - \varphi_1| \leq \psi_1$, avec ψ_1 en escalier et $\int_a^b \psi_1 < \varepsilon/(2M)$. Puisque $|f_1| \leq M$, on peut supposer que $|\varphi_1| \leq M$ (micro-exercice). Soit de même φ_2 une approximation en escalier de f_2 , telle que $|f_2 - \varphi_2| \leq \psi_2$, avec ψ_2 en escalier et $\int_a^b \psi_2 < \varepsilon/(2M)$; on a alors

$$f_1 f_2 - \varphi_1 \varphi_2 = (f_1 - \varphi_1) f_2 + \varphi_1 (f_2 - \varphi_2),$$

donc

$$|f_1 f_2 - \varphi_1 \varphi_2| \leq M\psi_1 + M\psi_2 = \psi,$$

et on obtient $\int_a^b \psi < \varepsilon$.

ATTENTION. Le résultat précédent ne marche que parce que les fonctions R-intégrables sont bornées par définition. Dans une théorie plus générale de l'intégration (intégrales généralisées absolument convergentes, ou bien théorie de Lebesgue), il faudra supposer que **l'une des deux** fonctions intégrables f_1 ou f_2 est de plus **bornée** pour pouvoir conclure que $f_1 f_2$ est intégrable.

Si on considère par exemple la théorie de l'intégrale généralisée absolument convergente sur $[0, 1]$, on voit que $f(x) = x^{-1/2}$ est « intégrable » sur $[0, 1]$, mais son carré $f^2(x) = f(x) \cdot f(x) = 1/x$ ne l'est plus.

Sommes de Riemann, théorème de Riemann

Étant donnée une subdivision π de $[a, b]$, on dira que

$$\delta(\pi) = \max\{x_i - x_{i-1} : 1 \leq i \leq n\}$$

est le *pas*, ou le *module* de la subdivision. On définira une *subdivision pointée* (π, ξ) de l'intervalle $[a, b]$ en choisissant pour chaque indice $i = 1, \dots, n$ un point ξ_i quelconque dans l'intervalle fermé $[x_{i-1}, x_i]$ de la subdivision π . Étant donnée une subdivision pointée (π, ξ) , on associe à chaque fonction réelle f définie sur $[a, b]$ la *somme de Riemann*

$$\Sigma_{\pi, \xi}(f) = \sum_{i=1}^n (x_i - x_{i-1}) f(\xi_i).$$

Il serait peut être préférable que la notation de la somme de Riemann rappelle l'intervalle $[a, b]$, mais on peut prétendre avec un peu de mauvaise foi que $[a, b]$ est implicitement donné par la subdivision π .

Les sommes de Riemann sont les intégrales de certaines fonctions en escalier : si ξ est un choix de points pour la subdivision π , la somme de Riemann $\Sigma_{\pi, \xi}(f)$ est l'intégrale de la fonction en escalier

$$\varphi = \sum_{i=1}^n f(\xi_i) \chi_{]x_{i-1}, x_i[}$$

qui est égale à $f(\xi_i)$ sur chaque intervalle ouvert $]x_{i-1}, x_i[$. On voit facilement que :

si f et g sont deux fonctions réelles quelconques définies sur $[a, b]$ et si (π, ξ) est une subdivision pointée de $[a, b]$, on a

$$\Sigma_{\pi, \xi}(\lambda f + \mu g) = \lambda \Sigma_{\pi, \xi}(f) + \mu \Sigma_{\pi, \xi}(g); \quad (f \leq g \text{ sur } [a, b]) \Rightarrow (\Sigma_{\pi, \xi}(f) \leq \Sigma_{\pi, \xi}(g)).$$

Remarque. Si A est un sous-ensemble de \mathbb{R} , on notera χ_A la fonction définie sur \mathbb{R} par $\chi_A(x) = 1$ si $x \in A$ et $\chi_A(x) = 0$ si $x \notin A$. On dit que χ_A est la *fonction caractéristique*, ou bien *fonction indicatrice* de A .

Soit c un nombre tel que $a \leq c \leq b$ et désignons par J l'un quelconque des deux intervalles $]c, b]$ ou $[c, b]$; soit (π, ξ) une subdivision pointée de l'intervalle $[a, b]$, avec $\pi = (x_0 = a < \dots < x_n = b)$; on a alors en posant $J =]c, b]$

$$(**) \quad \left| \Sigma_{\pi, \xi}(\chi_J) - \int_a^b \chi_J(t) dt \right| = \left| \Sigma_{\pi, \xi}(\chi_J) - (b - c) \right| \leq \delta(\pi).$$

On voit donc que les sommes de Riemann de la fonction χ_J convergent vers l'intégrale $\int_a^b \chi_J = b - c$ de la fonction χ_J lorsque le pas $\delta(\pi)$ de la subdivision π tend vers 0. On verra un peu plus loin que cette convergence des sommes de Riemann vers l'intégrale est vraie pour toutes les fonctions Riemann-intégrables (c'est ce que nous appelons *théorème de Riemann*).

L'inégalité $(**)$ ci-dessus est presque évidente, ce sont uniquement les problèmes de « bord » qui peuvent nous embêter un peu. Si c n'est pas point de subdivision de π , il existe exactement un intervalle $[x_{k-1}, x_k]$ de la subdivision π qui contient le point c , et $x_{k-1} < c < x_k$. On est donc sûr que $\chi_J(\xi_i) = 0$ pour $i < k$ et $\chi_J(\xi_i) = 1$ si $i > k$, et seule la valeur de $\chi_J(\xi_k)$ est incertaine, 0 ou 1 selon que ξ_k est ou n'est pas dans J . On a donc

$$b - x_k = \sum_{i=k+1}^n (x_i - x_{i-1}) \cdot 1 \leq \Sigma_{\pi, \xi}(\chi_J) \leq \sum_{i=k}^n (x_i - x_{i-1}) \cdot 1 = b - x_{k-1},$$

et ces deux valeurs extrêmes qui encadrent $b - c$ diffèrent entre elles de $x_k - x_{k-1} \leq \delta(\pi)$, donc $|\Sigma_{\pi, \xi}(\chi_J) - (b - c)| \leq \delta(\pi)$. Le cas où c est point de subdivision est un peu plus embêtant, et laissé au lecteur.

Proposition 3.1.4. Pour toute fonction φ en escalier sur $[a, b]$, les sommes de Riemann $\Sigma_{\pi, \xi}(\varphi)$ convergent vers $\int_a^b \varphi$ lorsque $\delta(\pi)$ tend vers 0.

Démonstration. La fonction φ est combinaison linéaire de fonctions indicatrices d'intervalles, que l'on peut choisir de la forme précédente χ_J . Il suffit donc d'appliquer la linéarité des sommes de Riemann et la remarque précédente.

Théorème de Riemann

Théorème 3.1.1. Soit f une fonction à valeurs réelles définie sur $[a, b]$; si la fonction f est Riemann-intégrable sur $[a, b]$, alors pour tout nombre réel $\varepsilon > 0$, il existe $\eta > 0$ tel que pour toute subdivision pointée (π, ξ) de $[a, b]$ telle que $\delta(\pi) < \eta$ on ait

$$\left| \Sigma_{\pi, \xi}(f) - \int_a^b f(t) dt \right| < \varepsilon.$$

Autrement dit, les sommes de Riemann convergent vers l'intégrale de f lorsque le pas de la subdivision tend vers 0.

Inversement, si les sommes de Riemann de f convergent vers un nombre réel I lorsque le pas de la subdivision π tend vers 0, la fonction f est R-intégrable sur $[a, b]$ et son intégrale est égale à I .

Démonstration. Supposons que f soit R-intégrable, et déduisons que les sommes de Riemann convergent vers $\int_a^b f$. Soit $\varepsilon > 0$ donné et soit $\varphi_1 \leq f \leq \varphi_2$ un encadrement de f par deux fonctions en escalier, tel que $(\int_a^b f) - \varepsilon/2 < \int_a^b \varphi_1 \leq \int_a^b \varphi_2 < (\int_a^b f) + \varepsilon/2$. Puisque φ_1 et φ_2 sont en escalier, on sait que leurs sommes de Riemann respectives tendent vers les intégrales, donc il existe $\eta > 0$ tel que $|\Sigma_{\pi, \xi}(\varphi_j) - \int_a^b \varphi_j| < \varepsilon/2$ pour toute subdivision pointée (π, ξ) de module $< \eta$, et pour $j = 1, 2$. Il en résulte que

$$\left(\int_a^b f \right) - \varepsilon < \left(\int_a^b \varphi_1 \right) - \frac{\varepsilon}{2} < \Sigma_{\pi, \xi}(\varphi_1) \leq \Sigma_{\pi, \xi}(f) \leq$$

$$\Sigma_{\pi, \xi}(\varphi_2) \leq \left(\int_a^b \varphi_2 \right) + \frac{\varepsilon}{2} < \left(\int_a^b f \right) + \varepsilon,$$

ce qui montre que $|\Sigma_{\pi, \xi}(f) - \int_a^b f| < \varepsilon$ pour toute subdivision pointée de module $< \eta$, et termine la démonstration du premier point.

Supposons inversement que les sommes de Riemann de f convergent vers un nombre réel I lorsque le pas de la subdivision π tend vers 0. Choisissons $\varepsilon > 0$ et soit $\eta > 0$ tel que $|\Sigma_{\pi, \xi}(f) - I| < \varepsilon/4$ pour toute subdivision pointée (π, ξ) de pas $< \eta$. Soit π une subdivision de diamètre $< \eta$; pour chaque intervalle $I_j =]x_{j-1}, x_j[$ de π , choisissons un point $\xi_j \in I_j$ tel que $f(\xi_j) \sim \inf f(I_j)$, disons $f(\xi_j) < \inf f(I_j) + \varepsilon/(4(b-a))$. La fonction en escalier φ_1 égale à $\inf f(I_j)$ sur chaque intervalle $]x_{j-1}, x_j[$ et à $f(x_j)$ au point x_j vérifie $\varphi_1 \leq f$ et

$$0 \leq \Sigma_{\pi, \xi}(f) - \int_a^b \varphi_1 < \varepsilon/4$$

donc $|I - \int_a^b \varphi_1| < \varepsilon/2$. De même on pourra trouver φ_2 en escalier telle que $f \leq \varphi_2$ et $|I - \int_a^b \varphi_2| < \varepsilon/2$, ce qui donne bien un encadrement de f par deux fonctions en escalier avec $\int_a^b (\varphi_2 - \varphi_1) < \varepsilon$. L'intégrale de f étant comprise entre celles de φ_1 et φ_2 , il en résulte aussi que $|I - \int_a^b f| < \varepsilon$, et ceci pour tout $\varepsilon > 0$, donc $\int_a^b f = I$.

Exemple. Si f est R-intégrable sur $[0, 1]$, on a

$$\int_0^1 f(t) dt = \lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{i=1}^n f\left(\frac{i}{n}\right).$$

Cette formule s'applique plus ou moins directement dans un grand nombre d'exercices, par exemple dans l'exercice classique suivant : montrer que

$$\lim_{n \rightarrow +\infty} \sum_{k=1}^n \frac{1}{n+k} = \ln 2.$$

Fonctions intégrables usuelles

Théorème 3.1.2. Intégrabilité des fonctions continues. *Si f est continue sur $[a, b]$, elle est R-intégrable sur $[a, b]$.*

Démonstration. On sait que toute fonction réelle continue f sur un intervalle fermé borné $[a, b]$ est uniformément continue, ce qui nous dit qu'étant donné $\varepsilon > 0$, il existe $\eta > 0$ tel que $|f(x) - f(y)| < \varepsilon/(b-a)$ chaque fois que $|x - y| < \eta$. Il en résulte que pour toute subdivision π de module $< \eta$, la fonction en escalier φ_π définie par $\varphi_\pi(x) = f(x_i)$ si $x \in]x_{i-1}, x_i]$, pour $i = 1, \dots, n$ et $\varphi_\pi(a) = f(a)$, c'est-à-dire

$$\varphi_\pi = f(a)\chi_{[a,a]} + \sum_{i=1}^n f(x_i)\chi_{]x_{i-1}, x_i]}$$

vérifie $|f - \varphi_\pi| < \varepsilon/(b-a)$ sur l'intervalle $[a, b]$. On obtient une approximation en escalier de f en posant $\varphi = \varphi_\pi$ et en prenant pour ψ la fonction constante $\varepsilon/(b-a)$. On a $|f - \varphi| \leq \psi$ et $\int_a^b \psi = \varepsilon$.

Théorème 3.1.3. *Si f est de classe C^1 sur $[a, b]$, on a*

$$f(b) - f(a) = \int_a^b f'(t) dt.$$

Démonstration. Considérons une subdivision $\pi = (x_0, \dots, x_n)$. Pour tout $i = 1, \dots, n$ on peut trouver d'après le théorème des accroissements finis un point $\xi_i \in [x_{i-1}, x_i]$ tel que

$$f(x_i) - f(x_{i-1}) = (x_i - x_{i-1})f'(\xi_i)$$

donc $\Sigma_{\pi, \xi}(f') = f(b) - f(a)$. Quand on fait tendre le diamètre de π vers 0, cette somme de Riemann particulière doit tendre vers l'intégrale de f' , puisque f' est R-intégrable.

Théorème 3.1.4. *Si f est une fonction monotone sur l'intervalle borné $[a, b]$, alors f est Riemann-intégrable sur $[a, b]$.*

Démonstration. Supposons par exemple f croissante. Découpons l'intervalle $[a, b]$ en n parties égales, chacune de longueur $h = (b-a)/n$, et limitées par les points $x_0 = a < x_1 = a + h < x_2 = a + 2h < \dots < x_n = b$. Définissons deux fonctions étagées φ_1 et φ_2 sur $[a, b]$ en posant $\varphi_1(x) = f(x_{i-1})$ pour tout $x \in [x_{i-1}, x_i[$, et $\varphi_2(x) = f(x_i)$ pour tout $x \in]x_{i-1}, x_i]$, pour tout $i = 1, \dots, n$. Pour être complet, on pose aussi $\varphi_1(b) = f(b)$ et $\varphi_2(a) = f(a)$. On voit que $\varphi_1 \leq f \leq \varphi_2$,

$$\int_a^b \varphi_1 = h(f(a) + f(a+h) + \dots + f(a+(n-1)h))$$

et

$$\int_a^b \varphi_2 = h(f(a+h) + f(a+2h) + \dots + f(b)).$$

On a donc $\int_a^b (\varphi_2 - \varphi_1) = h(f(b) - f(a)) = (b-a)(f(b) - f(a))/n$, quantité que l'on peut rendre aussi petite que l'on veut, en choisissant l'entier n très grand.

Relation de Chasles

Soient $a < b < c$ et f une fonction à valeurs réelles définie sur $[a, c]$; alors f est Riemann-intégrable sur $[a, c]$ si et seulement si elle est intégrable sur $[a, b]$ et sur $[b, c]$, et dans ce cas

$$\int_a^c f = \int_a^b f + \int_b^c f.$$

On note que $\int_a^a f = 0$. Dans le cas où $b < a$, on convient que $\int_a^b f(t) dt = -\int_b^a f(t) dt$. Si on se donne a, b, c dans un ordre quelconque, on a alors la relation

$$\int_a^c f(t) dt = \int_a^b f(t) dt + \int_b^c f(t) dt.$$

Rappelons les propriétés de majoration : si $a \leq b$, on a $\left| \int_a^b f(t) dt \right| \leq \int_a^b |f(t)| dt$, mais dans le cas où on ne sait pas que $a \leq b$, il faut écrire la majoration sous la forme

$$\left| \int_a^b f(t) dt \right| \leq \left| \int_a^b |f(t)| dt \right|.$$

En particulier, si on a $|f(t)| \leq M$ pour tout point t entre a et b , on en déduit la majoration $\left| \int_a^b f(t) dt \right| \leq M |b - a|$, quelle que soit la position respective de a et b .

Intégrale et primitives

Soit f une fonction réelle continue définie sur un intervalle ouvert I ; soit a un point fixé de I et posons

$$F(x) = \int_a^x f(t) dt$$

pour tout $x \in I$ (avec la convention ci-dessus lorsque $x < a$).

Théorème 3.1.5. *La fonction F est une primitive de f sur I . On a donc montré que toute fonction continue sur un intervalle ouvert I admet des primitives.*

Démonstration. Montrons que pour tout $x \in I$, le nombre $f(x)$ est la dérivée de F au point x . On a

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| = \left| \frac{1}{h} \int_x^{x+h} (f(t) - f(x)) dt \right| \leq \left| \frac{1}{h} \int_x^{x+h} |f(t) - f(x)| dt \right| \leq \varepsilon$$

si $|h|$ est choisi assez petit pour que $x+h \in I$ et $|f(t) - f(x)| < \varepsilon$ pour tout t entre les points x et $x+h$.

3.2. Le cas vectoriel.

Dans ce paragraphe on considère des fonctions définies sur un intervalle $[a, b]$ et à valeurs dans un espace vectoriel réel F de dimension finie. On supposera que F est muni d'une norme. On appellera *fonction en escalier* de $[a, b]$ dans F une fonction φ de $[a, b]$ dans F pour laquelle il existe un entier $n \geq 1$ et une subdivision $x_0 = a < x_1 < \dots < x_n = b$ de $[a, b]$ telle que φ soit constante sur chaque intervalle $]x_{i-1}, x_i[$, pour $i = 1, \dots, n$.

Si c_i désigne la valeur constante (vectorielle) de φ sur l'intervalle $]x_{i-1}, x_i[$, on définit l'intégrale de φ sur l'intervalle $[a, b]$ par la formule

$$\int_a^b \varphi(t) dt = \sum_{i=1}^n (x_i - x_{i-1}) c_i \in F.$$

(on doit vérifier que le résultat ne dépend que de φ : la vérification est la même que dans le cas réel). On vérifie facilement les propriétés de linéarité de l'intégrale des fonctions en escalier vectorielles, ainsi que la majoration $\|\int_a^b \varphi\| \leq \int_a^b \|\varphi\|$ (appliquer l'inégalité triangulaire à la somme de vecteurs qui définit l'intégrale).

Définition 3.2.1. Soit f une fonction définie sur $[a, b]$ et à valeurs dans F ; on dit que f est *intégrable au sens de Riemann* (nous écrirons en abrégé : R-intégrable) si pour tout $\varepsilon > 0$ donné, on peut trouver une fonction *vectorielle* φ en escalier et ψ en escalier réelle telles que

$$\forall x \in [a, b], \|f(x) - \varphi(x)\| \leq \psi(x) \text{ et } \int_a^b \psi(t) dt < \varepsilon.$$

La définition du vecteur *intégrale* demande un petit raisonnement. Considérons une suite quelconque d'approximations $\|f - \varphi_m\| \leq \psi_m$, avec (φ_m) suite de fonctions en escalier vectorielles et (ψ_m) suite de fonctions en escalier réelles telle que $\lim_m \int_a^b \psi_m = 0$. Étant donné m et p , on pourra écrire $\|\varphi_m - \varphi_p\| \leq \|f - \varphi_m\| + \|f - \varphi_p\| \leq \psi_m + \psi_p$, ce qui entraîne que $\int_a^b \|\varphi_m - \varphi_p\|$ tend vers 0 lorsque $m, p \rightarrow +\infty$. Il en résulte que

$$\left\| \int_a^b \varphi_m - \int_a^b \varphi_p \right\|$$

tend aussi vers 0, donc la suite des intégrales $(\int_a^b \varphi_m)$ est une suite de Cauchy dans l'espace F , donc une suite convergente puisque F est complet (comme tout espace vectoriel réel de dimension finie). Il est naturel de définir l'intégrale de f comme étant le vecteur limite I de la suite des intégrales des φ_m , mais il reste un petit point à compléter : la limite ne dépend pas de la suite (φ_m, ψ_m) choisie. Si (φ, ψ) est une approximation quelconque de f telle que $\|f - \varphi\| \leq \psi$ et $\int_a^b \psi < \varepsilon$, on aura

$$\|\varphi_m - \varphi\| \leq \|f - \varphi_m\| + \|f - \varphi\| \leq \psi_m + \psi.$$

Si on fait tendre m vers l'infini, l'intégrale de φ_m tend vers I et celle de ψ_m vers 0, ce qui entraîne que

$$\left\| I - \int_a^b \varphi \right\| \leq \int_a^b \psi < \varepsilon.$$

On voit alors que le vecteur $I \in F$ est uniquement déterminé par cette propriété : c'est l'unique vecteur I de F tel que pour tout $\varepsilon > 0$ donné et pour toute approximation $\|f - \varphi\| \leq \psi$, $\int_a^b \psi < \varepsilon$, on ait $\left\| I - \int_a^b \varphi \right\| < \varepsilon$. On notera encore l'intégrale par le symbole

$$\int_a^b f(t) dt = I \in F$$

ou bien $\int_a^b f$ pour abrégé.

Remarque hors programme. Cette définition reste la définition correcte si on cherche à étendre la définition de l'intégrabilité à des fonctions à valeurs dans un espace vectoriel normé de dimension infinie (qu'on aura besoin de supposer complet, pour faire converger les suites de Cauchy ci-dessus).

Supposons donnée une base de F et choisissons pour norme la norme euclidienne des coordonnées dans cette base. Si $t \rightarrow f(t)$ est une fonction de $[a, b]$ dans F , désignons par f_i les fonctions coordonnées de f (ce sont des fonctions à valeurs réelles). On voit facilement que f est Riemann-intégrable au sens de la définition précédente si et seulement si toutes les fonctions coordonnées f_i sont Riemann-intégrables : on peut en effet approcher f par une fonction en escalier vectorielle φ si et seulement si on peut approcher toutes les fonctions coordonnées f_i par des fonctions en escalier réelles φ_i . Dans ce cas, l'intégrale de Riemann de f est le vecteur de F dont les composantes dans la base choisie sont les intégrales des fonctions f_i .

Il en résulte plusieurs conséquences faciles, obtenues simplement en passant aux fonctions coordonnées : si f est \mathbb{R} -intégrable sur $[a, b]$, elle est bornée. Si f est continue de $[a, b]$ dans l'espace vectoriel F , elle est \mathbb{R} -intégrable. Si f est de classe C^1 de $[a, b]$ dans F , on a $f(b) - f(a) = \int_a^b f'(t) dt$.

Étant donnée une subdivision pointée (π, ξ) de l'intervalle $[a, b]$, on associe à chaque fonction f définie sur $[a, b]$ et à valeurs dans F la *somme de Riemann* (vectorielle)

$$\Sigma_{\pi, \xi}(f) = \sum_{i=1}^n (x_i - x_{i-1})f(\xi_i) \in F.$$

En appliquant le théorème de Riemann aux fonctions coordonnées, on obtient le même énoncé dans le cas vectoriel :

Soit f une fonction vectorielle définie sur $[a, b]$; si la fonction f est Riemann-intégrable sur $[a, b]$, alors pour tout nombre réel $\varepsilon > 0$, il existe $\eta > 0$ tel que pour toute subdivision pointée (π, ξ) de $[a, b]$ telle que $\delta(\pi) < \eta$ on ait

$$\left\| \Sigma_{\pi, \xi}(f) - \int_a^b f(t) dt \right\| < \varepsilon.$$

Autrement dit, les sommes de Riemann convergent vers l'intégrale de f lorsque le pas de la subdivision tend vers 0.

Proposition 3.2.1. Linéarité et majoration. Si f et g sont deux fonctions \mathbb{R} -intégrables sur $[a, b]$, à valeurs dans F , la fonction $\lambda f + \mu g$ est intégrable sur $[a, b]$ et on a

$$\int_a^b (\lambda f + \mu g) = \lambda \int_a^b f + \mu \int_a^b g.$$

De plus, la fonction réelle $t \rightarrow \|f(t)\|$ est \mathbb{R} -intégrable, et on a

$$\left\| \int_a^b f(t) dt \right\| \leq \int_a^b \|f(t)\| dt.$$

Démonstration. On montre que la fonction $t \rightarrow \|f(t)\|$ est \mathbb{R} -intégrable quand f l'est comme on a montré que $|f|$ est intégrable dans le cas réel. Pour vérifier les autres propriétés, il suffit de passer à la limite dans les propriétés correspondantes des sommes de Riemann, qui sont évidentes.

Le résultat précédent s'applique en particulier à la majoration de l'intégrale des fonctions à valeurs complexes, en considérant \mathbb{C} comme un espace normé sur \mathbb{R} et le module d'un nombre complexe comme la norme de cet espace.

Corollaire 3.2.1. Si f est une fonction Riemann-intégrable sur $[a, b]$, à valeurs complexes, la fonction réelle $t \rightarrow |f(t)|$ est \mathbb{R} -intégrable, et on a

$$\left| \int_a^b f(t) dt \right| \leq \int_a^b |f(t)| dt.$$

Proposition 3.2.2. Soient f, f_1, f_2 trois fonctions à valeurs dans F et \mathbb{R} -intégrables sur $[a, b]$, et h réelle et \mathbb{R} -intégrable sur $[a, b]$; les fonctions suivantes sont Riemann-intégrables sur $[a, b]$:

$$hf, f_1 \cdot f_2 \text{ (produit scalaire), } t \rightarrow \|f(t)\|.$$

Démonstration par approximation.

Un exemple d'application de la majoration de la norme

On va retrouver une version simple du théorème des Accroissements Finis vectoriel. Supposons f de classe C^1 d'un ouvert U de E (espace vectoriel réel de dimension finie) à valeurs dans F . On suppose que U contient un segment vectoriel $[A, B]$. On a alors en posant $g(t) = f(A + t(B - A))$ pour $t \in [0, 1]$

$$f(B) - f(A) = g(1) - g(0) = \int_0^1 g'(t) dt \in F$$

ce qui donne la majoration

$$\begin{aligned} \|f(B) - f(A)\| &\leq \int_0^1 \|g'(t)\| dt \leq \\ &\leq \sup_{t \in [0,1]} \|g'(t)\| = \|B - A\| \sup\{\|df_x\| : x \in [A, B]\}. \end{aligned}$$

Le résultat reste vrai en supposant simplement que f soit différentiable dans U (c'est-à-dire en enlevant l'hypothèse de continuité de la différentielle), mais il demande alors une démonstration différente (bien que l'idée de fond soit la même ; voir le chapitre de Calcul Différentiel).

3.3. Calculs approchés des intégrales simples

Méthode des trapèzes

Soient f une fonction de classe C^2 sur l'intervalle $[a, b]$, N un entier ≥ 1 et $h = (b - a)/N$. On approche l'intégrale de f entre a et b par la formule

$$T_N = h \left(\frac{1}{2}f(a) + f(a + h) + f(a + 2h) + \dots + f(a + (N - 1)h) + \frac{1}{2}f(b) \right)$$

et on obtient la majoration suivante pour l'erreur

$$\left| \int_a^b f(t) dt - T_N \right| \leq \frac{(b - a)M_2}{12} h^2$$

où

$$M_2 = \max\{|f''(x)| : a \leq x \leq b\}.$$

La méthode des trapèzes revient à approcher la fonction f sur chaque intervalle de longueur h de la subdivision par une fonction affine qui coïncide avec f en chaque extrémité de l'intervalle.

L'évaluation de l'erreur est fondée sur le calcul suivant : on considère un petit intervalle $[u, v]$, on désigne par m son milieu et on pose $h = v - u$. On désigne par $a(t)$ la fonction affine telle que $a(u) = f(u)$ et $a(v) = f(v)$. On obtient par deux intégrations par parties convenablement choisies un calcul de l'erreur commise dans l'intégrale de u à v en remplaçant f par la fonction affine a :

$$\begin{aligned} \int_u^v (f(t) - a(t)) dt &= [(f(t) - a(t))(t - m)]_u^v - \int_u^v (f'(t) - a'(t))(t - m) dt \\ &= [(f'(t) - a'(t))(h^2/8 - (t - m)^2/2)]_u^v - \int_u^v f''(t)(h^2/8 - (t - m)^2/2) dt = \\ &= - \int_u^v f''(t)(h^2/8 - (t - m)^2/2) dt. \end{aligned}$$

On observe que $h^2/8 - (t - m)^2/2$ reste positif sur l'intervalle $[u, v]$, ce qui permet de majorer la valeur absolue de l'intégrale précédente par

$$M_2 \int_u^v (h^2/8 - (t - m)^2/2) dt = M_2 \frac{h^3}{12}$$

ce qui donne le résultat annoncé en ajoutant tous les petits bouts. . .

Méthode de Simpson

Soient f une fonction de classe C^4 sur l'intervalle $[a, b]$, N un entier ≥ 1 et posons $h = (b - a)/(2N)$; on découpe l'intervalle $[a, b]$ en un nombre pair $2N$ d'intervalles de longueur h . La méthode de Simpson, qui est beaucoup plus performante que la méthode des trapèzes dans la plupart des cas, consiste à approcher f par un trinôme sur chaque groupe de deux intervalles successifs. Puisqu'on est passé d'approximation affine à approximation trinôme, on pourrait s'attendre à ce que l'ordre de l'erreur passe de h^2 à h^3 , mais il se produit un petit miracle, qui fait que l'ordre de l'erreur sera en fait h^4 .

On approche l'intégrale par

$$S_N = \frac{h}{3} \left(f(a) + 4f(a+h) + 2f(a+2h) + 4f(a+3h) + \dots \right. \\ \left. + \dots + 2f(a+(2N-2)h) + 4f(a+(2N-1)h) + f(b) \right)$$

et on obtient la majoration suivante pour l'erreur :

$$\left| \int_a^b f(t) dt - S_N \right| \leq \frac{(b-a)M_4}{180} h^4$$

où

$$M_4 = \max\{|f^{(4)}(x)| : a \leq x \leq b\}.$$

Lemme. Pour tout polynôme P de degré ≤ 3 , on a l'égalité

$$\int_{u-h}^{u+h} P(t) dt = \frac{h}{3}(P(u-h) + 4P(u) + P(u+h)).$$

Démonstration. Soit Q une primitive de P ; alors Q est un polynôme de degré ≤ 4 . D'après la formule de Taylor pour les polynômes,

$$Q(u+h) = Q(u) + hQ'(u) + \frac{h^2}{2}Q''(u) + \frac{h^3}{6}Q^{(3)}(u) + \frac{h^4}{24}Q^{(4)}(u),$$

et de même pour $Q(u-h)$, donc

$$\int_{u-h}^{u+h} P(t) dt = 2hQ'(u) + \frac{h^3}{3}Q^{(3)}(u) = 2hP(u) + \frac{h^3}{3}P''(u).$$

En appliquant maintenant la formule de Taylor au calcul de $P(u-h) + 4P(u) + P(u+h)$, on vérifie le lemme. Une autre façon de procéder consiste à remarquer d'abord qu'on peut supposer $u = 0$ par translation (la fonction $x \rightarrow P(x-u)$ est une autre fonction polynomiale de même degré), puis qu'il suffit de montrer le lemme pour les quatre fonctions $1, x, x^2$ et x^3 , puisqu'on obtiendra tous les polynômes de degré ≤ 3 par combinaisons linéaires de ces quatre. Il reste à observer que $\int_{-h}^h dt = 2h = \frac{h}{3}(1+4+1)$, $\int_{-h}^h t dt = 0 = \frac{h}{3}(-h+0+h)$, $\int_{-h}^h t^2 dt = 2h^3/3 = \frac{h}{3}((-h)^2+0+h^2)$, $\int_{-h}^h t^3 dt = 0 = \frac{h}{3}((-h)^3+0+h^3)$.

La formule précise du calcul de l'erreur donnée ci-dessus pour la méthode de Simpson est assez délicate à démontrer; la preuve suit un chemin analogue à celui que nous avons indiqué pour la méthode des trapèzes. Indiquons une approche plus simple, mais qui fournit de moins bonnes constantes. Soit $[u-h, u+h]$ « un bout » dans l'application de la méthode de Simpson; considérons la formule de Taylor à l'ordre 4, appliquée à la fonction f au point u . Soit P le polynôme de degré 3 figurant avant le reste (intégral ou de Lagrange); nous savons que

$$|f(x) - P(x)| \leq \frac{(x-u)^4}{24} M_4.$$

Il en résulte

$$\left| \int_{u-h}^{u+h} f(t) dt - \int_{u-h}^{u+h} P(t) dt \right| \leq M_4 \frac{h^5}{60}.$$

Par ailleurs, en posant $S_{u,h}(g) = (h/3)(g(u-h) + 4g(u) + g(u+h))$, on obtient

$$|S_{u,h}(f) - S_{u,h}(P)| \leq \frac{h}{3} M_4 \frac{h^4}{12} = M_4 \frac{h^5}{36}.$$

L'erreur sur un « bout » est donc $\leq 2M_4h^5/45$. En multipliant par le nombre N de bouts, on obtient une majoration de l'erreur totale par $(b-a)M_4h^4/45$.

Chapitre 4. Rappels d'algèbre linéaire

4.1. Espaces vectoriels

Pour « fabriquer » un *espace vectoriel*, il faut d'abord se donner un ensemble E (dont les éléments seront appelés *vecteurs*), puis donner deux opérations, l'addition des vecteurs d'une part, et d'autre part la multiplication des vecteurs par des nombres appelés *scalaires*; on suppose que ces données vérifient les axiomes des espaces vectoriels, que nous ne rappellerons pas ici; l'ensemble des scalaires doit être un *corps*, appelé le *corps de base* de l'espace vectoriel E ; ce corps sera noté \mathbb{K} . On se limitera le plus souvent dans ce cours au cas où $\mathbb{K} = \mathbb{R}$ ou $\mathbb{K} = \mathbb{C}$; le corps \mathbb{K} pourra être aussi un *sous-corps de \mathbb{C}* , par exemple $\mathbb{K} = \mathbb{Q}$. Strictement parlant, on devrait dire « l'espace vectoriel $(E, +, \cdot)$ » mais l'usage courant est de dire simplement « l'espace vectoriel E ».

Pour tout entier $n \geq 1$, l'espace vectoriel \mathbb{K}^n est l'ensemble des n -uplets (a_1, \dots, a_n) d'éléments de \mathbb{K} , muni des opérations usuelles d'addition et de multiplication par les éléments de \mathbb{K} (les « scalaires »). Nous associerons souvent à un élément x de \mathbb{K}^n une matrice colonne X pour pouvoir appliquer le calcul matriciel,

$$x = (a_1, \dots, a_n) \in \mathbb{K}^n, \quad X = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

(attention aux conventions : (a_1, \dots, a_n) est un n -uplet, pas une matrice ligne; une matrice ligne ne contient pas de virgules).

La *base canonique* de l'espace vectoriel \mathbb{K}^n est formée des vecteurs $\mathbf{e}_1 = (1, 0, \dots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \dots, 0)$, jusqu'à $\mathbf{e}_n = (0, \dots, 0, 1)$. Dans cette base le vecteur x ci-dessus peut s'écrire $x = a_1 \mathbf{e}_1 + \dots + a_n \mathbf{e}_n$.

Sous-espaces vectoriels et applications linéaires

Définitions.

4.1.1. Sous-espace vectoriel. Soit E un espace vectoriel sur \mathbb{K} ; on dit qu'un sous-ensemble F de E est un *sous-espace vectoriel de E* si les trois conditions suivantes sont vérifiées :

- Le vecteur nul 0_E de E appartient à F , c'est-à-dire $0_E \in F$
- Pour tous vecteurs x, y de F , on a $x + y \in F$
- Pour tout vecteur $x \in F$ et tout $\lambda \in \mathbb{K}$, on a $\lambda x \in F$.

On peut résumer ces trois conditions en disant que F est **non vide** et que $\lambda x + y \in F$ pour tous $x, y \in F$ et tout $\lambda \in \mathbb{K}$.

4.1.2. Application linéaire. Soient E_1 et E_2 deux espaces vectoriels sur \mathbb{K} ; on dit qu'une application $u : E_1 \rightarrow E_2$ est une *application linéaire* de E_1 dans E_2 si

- Pour tous vecteurs $x, y \in E_1$, on a $u(x + y) = u(x) + u(y)$
- Pour tout vecteur $x \in E_1$ et tout $\lambda \in \mathbb{K}$, on a $u(\lambda x) = \lambda u(x)$.

On peut résumer en disant que $u(\lambda x + y) = \lambda u(x) + u(y)$ pour tous $x, y \in E_1$ et tout $\lambda \in \mathbb{K}$. On rappelle que $u(0_{E_1}) = 0_{E_2}$, et $u(-x) = -u(x)$ pour tout vecteur $x \in E_1$.

Exemples de sous-espaces vectoriels.

1. Les ensembles $\{0_E\}$ et E sont des sous-espaces vectoriels de E .
2. Pour tout vecteur $x \in E$ non nul, on peut considérer la *droite vectorielle*

$$\mathbb{K}x = \{\lambda x : \lambda \in \mathbb{K}\}.$$

C'est un sous-espace vectoriel de E . Dans le cas où $\mathbb{K} = \mathbb{R}$, le sous-espace vectoriel $\mathbb{R}x$ correspond vraiment à une droite au sens usuel, de vecteur directeur x et passant par le point origine 0_E . Dans le cas complexe (si $\mathbb{K} = \mathbb{C}$), le lecteur devra bien comprendre qu'une « droite complexe » ne correspond pas à l'idée géométrique usuelle de droite. Une *droite affine* n'est pas en général un sous-espace vectoriel (elle ne passe pas toujours par 0); c'est un ensemble de la forme $a + \mathbb{K}x$; une telle droite affine passe par le point $a \in E$, et elle admet le vecteur $x \neq 0_E$ comme vecteur directeur. Un *plan vectoriel* est un sous-espace vectoriel de la forme $\{\lambda x + \mu y : \lambda, \mu \in \mathbb{K}\}$, où x et y sont deux vecteurs non \mathbb{K} -colinéaires.

3. L'intersection $E_1 \cap E_2$ de deux sous-espaces vectoriels E_1 et E_2 de E est un sous-espace vectoriel de E .

4. Étant donnés deux sous-espaces vectoriels F_1 et F_2 de E , on définit la *somme des deux sous-espaces* par

$$F_1 + F_2 = \{y_1 + y_2 : y_1 \in F_1, y_2 \in F_2\}.$$

Par exemple, la somme de deux droites vectorielles distinctes est un plan vectoriel. Cette opération est associative : $(F_1 + F_2) + F_3 = F_1 + (F_2 + F_3)$. On définit de même la somme $F_1 + \dots + F_n$ d'une famille finie F_1, \dots, F_n de sous-espaces vectoriels de E .

5. Produit fini d'espaces vectoriels

Si E_1, \dots, E_n sont des espaces vectoriels sur \mathbb{K} , l'espace produit $E_1 \times \dots \times E_n$, qui est l'ensemble des n -uplets (x_1, \dots, x_n) où $x_i \in E_i$ pour chaque $i = 1, \dots, n$, est muni d'une structure naturelle d'espace vectoriel sur \mathbb{K} qui découle des structures de E_1, \dots, E_n : les opérations de l'espace vectoriel $E_1 \times \dots \times E_n$ sont définies par

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n),$$

$$a(x_1, \dots, x_n) = (ax_1, \dots, ax_n)$$

pour tout $a \in \mathbb{K}$. L'espace vectoriel \mathbb{K}^n est un cas particulier de cette opération produit, le cas où tous les facteurs E_i sont égaux à \mathbb{K} . L'application de $E_1 \times \dots \times E_n$ dans E_i définie par $\pi_i(x_1, \dots, x_n) = x_i$ s'appelle la *i -ème projection* de l'espace produit sur le i ème facteur E_i .

Exemples d'applications linéaires.

1. L'application nulle, les projections d'un espace vectoriel produit sur les facteurs sont des applications linéaires. L'*application identique* de E ou *identité* de E , qui sera notée Id_E dans ces notes, et qui est définie par $\text{Id}_E(x) = x$ pour tout $x \in E$, est évidemment une application linéaire de E dans E . On appelle *endomorphisme* de E toute application linéaire de E dans lui-même.

2. La somme de deux applications linéaires est une application linéaire. L'ensemble de toutes les applications linéaires de E dans F est l'espace vectoriel $\mathcal{L}(E, F)$.

3. On appelle *forme linéaire* une application linéaire de E dans le corps \mathbb{K} . Le *dual* de E est l'espace $\mathcal{L}(E, \mathbb{K})$ de toutes les formes linéaires sur E . On le note en général E^* .

4. Quand elle est définie, la composition $v \circ u$ de deux applications linéaires u et v est une application linéaire.

5. Étant donné un système $\mathbf{x} = (x_1, \dots, x_p)$ de vecteurs d'un espace vectoriel E , l'application de \mathbb{K}^p dans E définie par

$$\varphi_{\mathbf{x}}(\lambda_1, \dots, \lambda_p) = \sum_{j=1}^p \lambda_j x_j$$

est une application linéaire.

Une application linéaire u de E dans F est un *isomorphisme* (d'espaces vectoriels) s'il existe une application linéaire $v : F \rightarrow E$ telle que $v \circ u = \text{Id}_E$ et $u \circ v = \text{Id}_F$. Dans ce cas l'application v est l'application réciproque de u (au sens ensembliste) ; pour que l'application linéaire u soit un isomorphisme, il suffit que u soit bijective de E sur F ; l'application réciproque $v = u^{-1}$ sera automatiquement linéaire.

Soit $u \in \mathcal{L}(E, F)$ une application linéaire ; le *noyau* de u est l'ensemble des vecteurs $x \in E$ tels que $u(x) = 0_F$; c'est un sous-espace vectoriel de E et on le note $\ker u$. L'*image* de u est l'ensemble $u(E)$ des éléments de F qui sont l'image d'au moins un élément de l'espace E ; c'est un sous-espace vectoriel de F .

Rappel. Une application linéaire $u \in \mathcal{L}(E, F)$ est injective si et seulement si $\ker u = \{0_E\}$.

Exemple. La somme $F_1 + F_2$ de deux sous-espaces de E est l'image de $F_1 \times F_2$ par l'application linéaire u de $F_1 \times F_2$ dans E définie par $u(x_1, x_2) = x_1 + x_2$. Le noyau de u est le sous-espace de $F_1 \times F_2$ formé de tous les couples $(x, -x)$ où x varie dans l'intersection $F_1 \cap F_2$. L'application u est donc injective si et seulement si $F_1 \cap F_2 = \{0\}$; dans ce cas u sera un isomorphisme de $F_1 \times F_2$ sur $F_1 + F_2$.

Exemples mettant en jeu des espaces vectoriels de dimension infinie

L'ensemble S des séries convergentes, muni de la somme $(\sum u_n) + (\sum v_n) = \sum (u_n + v_n)$ et de l'opération $\lambda \sum u_n = \sum \lambda u_n$, est un espace vectoriel (de dimension infinie) ; l'espace des séries absolument convergentes est un sous-espace vectoriel de S , l'application qui associe à chaque série convergente la somme de la série est une application linéaire de S dans $\mathbb{K} = \mathbb{R}$ ou \mathbb{C} (forme linéaire). La transformation d'Abel fournit des exemples d'endomorphismes de l'espace vectoriel S : si (λ_n) est une suite réelle décroissante tendant vers 0, l'application $\sum v_n \rightarrow \sum \lambda_n v_n$ est linéaire de S dans S .

L'ensemble $C(\mathbb{R})$ des fonctions réelles continues sur \mathbb{R} est muni d'une structure naturelle d'espace vectoriel sur \mathbb{R} où les opérations sont la somme de deux fonctions et la multiplication d'une fonction par une constante réelle,

$$(f + g)(x) = f(x) + g(x), \quad (af)(x) = af(x).$$

Les fonctions de classe C^1 sur \mathbb{R} forment un sous-espace vectoriel $C^1(\mathbb{R})$ de l'espace vectoriel $C(\mathbb{R})$. Les fonctions de classe C^2 forment un autre sous-espace $C^2(\mathbb{R})$, contenu dans $C^1(\mathbb{R})$. Étant données deux fonctions a et b sur \mathbb{R} , les fonctions f de $C^2(\mathbb{R})$ telles que

$$\forall x \in \mathbb{R}, \quad f''(x) = a(x)f'(x) + b(x)f(x)$$

forment un sous-espace vectoriel de $C^2(\mathbb{R})$.

L'application qui associe à toute fonction $f \in C(\mathbb{R})$ sa primitive F nulle en 0 est une application linéaire de $C(\mathbb{R})$ dans $C^1(\mathbb{R})$. L'application qui associe à toute fonction $f \in C(\mathbb{R})$ la quantité $\int_0^1 f(t) dt$ est une application linéaire de $C(\mathbb{R})$ dans \mathbb{R} (forme linéaire).

Combinaisons linéaires

Soit x_1, \dots, x_n une famille de vecteurs d'un espace vectoriel E sur le corps \mathbb{K} ; on dit qu'un vecteur $z \in E$ est une *combinaison linéaire* des vecteurs x_1, \dots, x_n s'il existe des coefficients $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ tels que

$$z = \lambda_1 x_1 + \dots + \lambda_n x_n.$$

On notera (dans ces notes de cours) $[x_1, \dots, x_n]$ l'ensemble des combinaisons linéaires des vecteurs x_1, \dots, x_n . Dans le cas d'un seul vecteur, on a $[x] = \mathbb{K}x$.

On voit que $[x_1, \dots, x_n]$ est l'image de \mathbb{K}^n par l'application linéaire $\varphi_{\mathbf{x}}$ définie dans l'exemple 5 ci-dessus. Par conséquent $[x_1, \dots, x_n]$ est un sous-espace vectoriel de E . Inversement, tout sous-espace F de E qui contient les vecteurs x_1, \dots, x_n contiendra aussi toutes les combinaisons linéaires de ces vecteurs, donc $[x_1, \dots, x_n] \subset F$. Autrement dit, le sous-espace vectoriel $[x_1, \dots, x_n]$ est le plus petit sous-espace vectoriel de E qui contienne les vecteurs x_1, \dots, x_n . On l'appelle le *sous-espace vectoriel engendré* par les vecteurs x_1, \dots, x_n .

Remarque. Si $u : E \rightarrow F$ est linéaire, on a $u([y_1, \dots, y_n]) = [u(y_1), \dots, u(y_n)]$.

4.2. Dimension

À tout système $\mathbf{x} = (x_1, \dots, x_n)$ de n vecteurs de E , on a associé une application linéaire $\varphi_{\mathbf{x}}$ de \mathbb{K}^n dans E , définie par

$$\forall (\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n, \quad \varphi_{\mathbf{x}}(\lambda_1, \dots, \lambda_n) = \lambda_1 x_1 + \dots + \lambda_n x_n \in E.$$

On a dit que l'image $\varphi_{\mathbf{x}}(\mathbb{K}^n)$ est égale à l'ensemble $[x_1, \dots, x_n]$ des combinaisons linéaires des vecteurs x_1, \dots, x_n .

Système libre, générateur

Définitions.

4.2.1. On dit que le système de vecteurs $\mathbf{x} = (x_1, \dots, x_n)$ est *libre* ou *linéairement indépendant* si l'application $\varphi_{\mathbf{x}}$ est injective, c'est-à-dire si

$$\forall (\lambda_1, \dots, \lambda_n) \in \mathbb{K}^n, \quad (\lambda_1 x_1 + \dots + \lambda_n x_n = 0_E \Rightarrow \lambda_1 = \lambda_2 = \dots = \lambda_n = 0).$$

4.2.2. On dit que le système de vecteurs $\mathbf{x} = (x_1, \dots, x_n)$ est un *système de générateurs* pour E si l'application $\varphi_{\mathbf{x}}$ est surjective de \mathbb{K}^n sur E , c'est-à-dire si tout vecteur de E peut s'exprimer comme combinaison linéaire des vecteurs x_1, \dots, x_n .

Remarques.

1. Image linéaire d'un système générateur : si $u \in \mathcal{L}(E, F)$, et si le système de vecteurs (x_1, \dots, x_n) est générateur pour E , alors $(u(x_1), \dots, u(x_n))$ est générateur pour l'image $u(E)$. En particulier, si u est surjective de E sur F , l'image d'un système de générateurs pour E est un système de générateurs pour F .

2. Image inverse d'un système libre : si $u \in \mathcal{L}(E, F)$, si $(u(x_1), \dots, u(x_n))$ est libre dans F , alors (x_1, \dots, x_n) est libre dans E .

3. Si un système de vecteurs est libre, tous ses sous-systèmes sont libres.

4. Un système de vecteurs (x_1, \dots, x_n) est libre si et seulement si aucun des vecteurs du système ne peut s'exprimer comme combinaison linéaire des **autres** vecteurs du système.

Lemme. Expression de l'indépendance linéaire par conditions successives. *Le système de vecteurs $\mathbf{x} = (x_1, \dots, x_n)$ est libre si et seulement si*

$$x_1 \neq 0_E \text{ et } \forall i = 1, \dots, n-1, \quad x_{i+1} \notin [x_1, \dots, x_i].$$

Démonstration abrégée. Supposons que la condition ci-dessus soit satisfaite, et montrons que le système \mathbf{x} est libre. Supposons donc que $\sum_{i=1}^n \lambda_i x_i = 0_E$ et montrons que tous les coefficients λ_i sont nuls ; s'il existait des λ_i non nuls, on pourrait considérer le dernier (le plus grand) indice p tel que $\lambda_p \neq 0$, et on aurait alors puisque $\lambda_{p+1} = \dots = \lambda_n = 0$,

$$\lambda_1 x_1 + \dots + \lambda_p x_p = 0_E,$$

donc

$$x_p = -\frac{1}{\lambda_p}(\lambda_1 x_1 + \dots + \lambda_{p-1} x_{p-1}) \in [x_1, \dots, x_{p-1}],$$

ce qui contredirait notre condition. Par conséquent tous les λ_i sont nuls, et on a donc montré que le système (x_1, \dots, x_n) est libre.

Autre expression de l'indépendance linéaire : un système de vecteurs (x_1, \dots, x_n) est libre si et seulement si

$$\{0_E\} \neq [x_1] \neq [x_1, x_2] \neq [x_1, x_2, x_3] \neq \dots \neq [x_1, \dots, x_n].$$

Corollaire 4.2.1. *Si (x_1, \dots, x_k) est libre et $y \notin [x_1, \dots, x_k]$, alors (x_1, \dots, x_k, y) est libre.*

Base d'un espace vectoriel

Un système $\mathbf{x} = (x_1, \dots, x_n)$ de vecteurs de E est une *base* de E s'il est à la fois libre et générateur pour E . Cela revient à dire que l'application $\varphi_{\mathbf{x}}$ est injective et surjective, donc que $\varphi_{\mathbf{x}}$ est un isomorphisme de \mathbb{K}^n sur E .

Exemple. Si $E = [x]$, avec $x \neq 0_E$, tout vecteur $y \neq 0_E$ est une base de E .

Si $\mathbf{e} = (e_1, \dots, e_n)$ est une base de E , on définit les *fonctions coordonnées* e_j^* , pour $j = 1, \dots, n$ par

$$e_j^*(a_1 e_1 + \dots + a_n e_n) = a_j.$$

Ce sont des formes linéaires. Elles forment une base du dual E^* , appelée la *base duale* \mathbf{e}^* de la base \mathbf{e} .

Les théorèmes sur la dimension pour les espaces engendrés par une partie finie

Dans ce paragraphe on suppose que l'espace vectoriel E étudié est différent de $\{0_E\}$.

Théorème 4.2.1. (Théorème de la base incomplète). *Si $\mathbf{x} = (x_1, \dots, x_p)$ est libre (ou bien si le système \mathbf{x} est le système vide) et si $\mathbf{y} = (y_1, \dots, y_q)$ est générateur pour E , il existe une base de E contenant tous les vecteurs du système \mathbf{x} et complétée par certains des vecteurs du système \mathbf{y} .*

Corollaire 4.2.2. *Si E est engendré par une partie finie, il existe une base pour E . Plus précisément, si E est engendré par q vecteurs on peut trouver une base de n vecteurs pour E , avec $n \leq q$.*

Démonstration. On applique le théorème de la base incomplète au système vide et à un système de générateurs pour E .

Théorème 4.2.2. Si E possède une base formée de n vecteurs, toutes les bases de E ont n vecteurs. Le nombre n s'appelle la dimension de E , et on le note $\dim E$.

Dans le cas de l'espace vectoriel $E = \{0\}$ réduit au vecteur nul, on convient de poser $\dim\{0\} = 0$.

Remarques.

- a. Si $\dim E = n$, l'espace E ne peut pas contenir $n + 1$ vecteurs indépendants.
- b. Si F est un sous-espace de E et si $\dim E = n$, on a $\dim F \leq n$; si $\dim F = n = \dim E$, alors $F = E$.
- c. Si $n = \dim E$ et si (x_1, \dots, x_n) est un système libre formé de n vecteurs de E , c'est une base de E .

Montrons le point a : si E contenait $n + 1$ vecteurs indépendants, il contiendrait une base ayant au moins $n + 1$ vecteurs d'après le théorème de la base incomplète, ce qui contredit $\dim E = n$. Le premier point de b résulte de a . Pour le second, si on avait $F \subset E$, différent de E et de même dimension n que E , on aurait, en prenant une base (y_1, \dots, y_n) de F et un vecteur $x \notin F$, un système libre (y_1, \dots, y_n, x) de $n + 1$ vecteurs dans E , ce qui est impossible d'après le point a . Pour le point c , on remarque que si $F = [x_1, \dots, x_n]$, on aura $\dim F = n = \dim E$, donc $F = E$.

Proposition 4.2.1. Soit u une application linéaire de E dans F . Si E est de dimension finie, on a

$$\dim E = \dim \ker u + \dim u(E).$$

En particulier, si $\dim F = \dim E$, une application linéaire de E dans F est injective si et seulement si elle est surjective.

4.3. Sommes directes

On a défini précédemment la somme de deux sous-espaces vectoriels F_1 et F_2 d'un espace vectoriel E ,

$$F_1 + F_2 = \{y_1 + y_2 : y_1 \in F_1, y_2 \in F_2\}.$$

On dit que la somme $F_1 + F_2$ est une *somme directe* si $F_1 \cap F_2 = \{0_E\}$. Cela entraîne l'unicité de la décomposition $y = y_1 + y_2$ d'un vecteur de $F_1 + F_2$ et l'existence des applications projections de $F_1 + F_2$ sur les deux facteurs. Dans le cas d'une somme directe, on note l'espace somme $F_1 \oplus F_2$.

La somme $F_1 + F_2$ est directe si et seulement si l'application $(y_1, y_2) \rightarrow y_1 + y_2$ de l'espace $F_1 \times F_2$ dans $F_1 + F_2$ est un isomorphisme.

Pour définir une somme directe $F_1 \oplus \dots \oplus F_p$ de plusieurs sous-espaces vectoriels F_1, \dots, F_p d'un espace vectoriel E , le plus simple est de partir de la propriété précédente : on dit que la somme $F_1 + \dots + F_p$ est une *somme directe* lorsque l'application linéaire $(y_1, \dots, y_p) \in F_1 \times \dots \times F_p \rightarrow y_1 + \dots + y_p$ est un isomorphisme de $F_1 \times \dots \times F_p$ sur $F_1 + \dots + F_p$. Comme cette application est surjective par définition de la somme, dire que la somme est directe signifie donc que l'application est injective, donc

la somme $F_1 + \dots + F_p$ est une somme directe si et seulement si pour tout choix de vecteurs $(y_1, \dots, y_p) \in F_1 \times \dots \times F_p$, la condition $y_1 + \dots + y_p = 0_E$ implique les égalités $y_1 = y_2 = \dots = y_p = 0_E$.

Remarque. Le système (x_1, \dots, x_n) est une base de E si et seulement si

$$E = \mathbb{K}x_1 \oplus \dots \oplus \mathbb{K}x_n.$$

Caractérisation par conditions successives

La somme $F_1 + \dots + F_p$ est directe si et seulement si pour tout $j = 1, 2, \dots, p-1$, l'espace F_{j+1} rencontre $F_1 + \dots + F_j$ seulement en 0_E ,

$$F_{j+1} \cap (F_1 + \dots + F_j) = \{0_E\}.$$

Base d'une somme directe. Soit $F_1 \oplus \dots \oplus F_p$ une somme directe de sous-espaces de dimension finie ; pour obtenir une base de $F_1 \oplus \dots \oplus F_p$, il suffit de prendre un système de vecteurs formé d'une base de F_1 , suivie d'une base de F_2 , ainsi de suite jusqu'à une base de F_p .

Lemme. Soient F_1, \dots, F_p des sous-espaces de dimension finie d'un espace vectoriel E ; la somme $F_1 + \dots + F_p$ est une somme directe si et seulement si

$$\dim(F_1 + \dots + F_p) = \dim F_1 + \dots + \dim F_p.$$

Matrice d'une application linéaire

On suppose donnée $u : E \rightarrow F$ une application linéaire, $\mathbf{e} = (e_1, \dots, e_m)$ une base de E et $\mathbf{f} = (f_1, \dots, f_n)$ une base de F . Pour chaque $j = 1, \dots, m$ soit Y_j le vecteur colonne de \mathbb{K}^n formé des coordonnées **dans la base \mathbf{f}** (à l'arrivée) de **l'image** $u(e_j)$ du j ème vecteur e_j de la base de départ \mathbf{e} ,

$$Y_j = \begin{pmatrix} a_{1,j} \\ \vdots \\ a_{n,j} \end{pmatrix}, \quad \text{avec } u(e_j) = a_{1,j}f_1 + \dots + a_{n,j}f_n.$$

La matrice $M = \text{mat}(u, \mathbf{e}, \mathbf{f})$ de u par rapport aux bases \mathbf{e} et \mathbf{f} est alors

$$M = (Y_1 \quad Y_2 \quad \dots \quad Y_m)$$

matrice de taille $n \times m$ (c'est-à-dire n lignes, m colonnes) dont les colonnes successives sont Y_1, Y_2, \dots, Y_m . On notera $M_{n,m}(\mathbb{K})$ l'espace vectoriel des matrices de taille $n \times m$ à coefficients dans \mathbb{K} (et $M_n(\mathbb{K})$ si $m = n$). Si X est le vecteur colonne des coordonnées de $x \in E$ dans la base \mathbf{e} , le vecteur colonne $Y = MX$ obtenu par produit avec la matrice M est le vecteur des coordonnées de $y = u(x) \in F$ dans la base \mathbf{f} .

Exemple. Rotation d'angle θ dans \mathbb{R}^2 . Sa matrice dans la base canonique est

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

Soit $\mathbf{e} = (e_1, \dots, e_n)$ une base d'un espace vectoriel E et soit Id_E l'application identique de E ; la matrice de Id_E en prenant la base \mathbf{e} au départ et à l'arrivée est la *matrice identité* de dimension n

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}, \quad I_n = \text{mat}(\text{Id}_E, \mathbf{e}, \mathbf{e}).$$

Rappel important. Produit d'applications linéaires et produit de matrices. Si $u : E \rightarrow F$ et $v : F \rightarrow G$ sont deux applications linéaires, \mathbf{e} une base de E , \mathbf{f} une base de F et \mathbf{g} une base de G , la matrice de l'application composée $v \circ u$ par rapport aux bases \mathbf{e} et \mathbf{g} est égale au produit des matrices de v et de u ,

$$\text{mat}(v \circ u, \mathbf{e}, \mathbf{g}) = \text{mat}(v, \mathbf{f}, \mathbf{g}) \text{mat}(u, \mathbf{e}, \mathbf{f}).$$

Si u est inversible, $\text{mat}(u^{-1}, \mathbf{f}, \mathbf{e})$ est la matrice inverse de $\text{mat}(u, \mathbf{e}, \mathbf{f})$.

Rappel. Changement de base.

On considère deux bases $\mathbf{e} = (e_1, \dots, e_n)$ et $\mathbf{f} = (f_1, \dots, f_n)$ pour le même espace vectoriel E de dimension n . Pour un endomorphisme u de E , désignons par $A_{\mathbf{e}}$ et $A_{\mathbf{f}}$ les matrices de u relatives aux bases \mathbf{e} et \mathbf{f} respectivement, $A_{\mathbf{e}} = \text{mat}(u, \mathbf{e}, \mathbf{e})$ et $A_{\mathbf{f}} = \text{mat}(u, \mathbf{f}, \mathbf{f})$. Considérons par ailleurs la matrice $V = \text{mat}(\text{Id}_E, \mathbf{f}, \mathbf{e})$ de l'application identique de l'espace E , avec la base \mathbf{f} au départ et la base \mathbf{e} à l'arrivée. La matrice V est inversible et son inverse V^{-1} est la matrice de l'application inverse, qui est encore l'application identique, mais avec la base \mathbf{e} au départ et \mathbf{f} à l'arrivée. On a d'après les principes généraux sur la matrice d'une composition d'applications

$$A_{\mathbf{f}} = V^{-1} A_{\mathbf{e}} V.$$

Par définition de la matrice d'une application linéaire par rapport à deux bases, la *matrice de passage* V contient dans sa colonne j les coordonnées du vecteur f_j de la base \mathbf{f} calculées par rapport à la base \mathbf{e} .

Calcul par blocs

On va essayer dans ce paragraphe de convaincre le lecteur de l'utilité de se servir du langage matriciel dans des situations plus abstraites, en introduisant des matrices dont les coefficients ne seront plus nécessairement des nombres, mais des applications linéaires, ou des vecteurs, ou bien d'autres matrices. Expliquons le principe de l'opération. Supposons donnés quatre espaces vectoriels E_1, E_2, F_1 et F_2 . Les deux espaces E joueront le rôle d'espaces de départ, et les F seront les espaces d'arrivée. Supposons données quatre applications linéaires $u_{j,k} : E_k \rightarrow F_j, j, k = 1, 2$. Elles permettent de définir une application linéaire u du produit $E_1 \times E_2$ dans le produit $F_1 \times F_2$, en posant pour tout couple $(x_1, x_2) \in E_1 \times E_2$

$$u(x_1, x_2) = (y_1, y_2) = (u_{1,1}(x_1) + u_{1,2}(x_2), u_{2,1}(x_1) + u_{2,2}(x_2)) \in F_1 \times F_2.$$

Il est assez naturel d'introduire la matrice abstraite

$$\begin{pmatrix} u_{1,1} & u_{1,2} \\ u_{2,1} & u_{2,2} \end{pmatrix}$$

et de remarquer que le calcul de $u(x_1, x_2)$ peut se comprendre en langage matriciel étendu, en effectuant avec les règles presque usuelles le produit

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} u_{1,1} & u_{1,2} \\ u_{2,1} & u_{2,2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} u_{1,1}(x_1) + u_{1,2}(x_2) \\ u_{2,1}(x_1) + u_{2,2}(x_2) \end{pmatrix}.$$

Si on se donne maintenant deux autres espaces G_1, G_2 et quatre applications linéaires $v_{i,j} : F_j \rightarrow G_i$, il leur correspond une application linéaire $v : F_1 \times F_2 \rightarrow G_1 \times G_2$, que l'on peut représenter par une matrice 2×2 dont les coefficients sont les $(v_{i,j})$. Ce qui est très sympathique, c'est que la composition $v \circ u$ peut alors se représenter par le produit de ces deux matrices abstraites, en écrivant simplement $v_{i,j} u_{j,k}$ pour la composition $v_{i,j} \circ u_{j,k}$

$$v \circ u = \begin{pmatrix} v_{1,1} & v_{1,2} \\ v_{2,1} & v_{2,2} \end{pmatrix} \begin{pmatrix} u_{1,1} & u_{1,2} \\ u_{2,1} & u_{2,2} \end{pmatrix} = \begin{pmatrix} v_{1,1}u_{1,1} + v_{1,2}u_{2,1} & v_{1,1}u_{1,2} + v_{1,2}u_{2,2} \\ v_{2,1}u_{1,1} + v_{2,2}u_{2,1} & v_{2,1}u_{1,2} + v_{2,2}u_{2,2} \end{pmatrix}$$

toujours avec les règles presque habituelles ; il faut seulement faire attention à l'ordre des facteurs, parce que par exemple $v_{1,1}u_{1,1}$ a un sens alors que $u_{1,1}v_{1,1}$ n'a pas de sens en général (et même si les deux ont un sens — quand $E_1 = F_1 = G_1$ — les deux produits sont différents en général).

Supposons que l'espace vectoriel E soit égal à la somme directe $E_1 \oplus E_2$ de deux sous-espaces E_1 et E_2 . L'espace E est alors naturellement isomorphe au produit $E_1 \times E_2$, et on peut appliquer ce qui a été dit précédemment. Tout vecteur $x \in E$ se décompose alors de façon unique en $x = x_1 + x_2$, avec $x_i = p_i(x) \in E_i$, $i = 1, 2$, les applications p_1 et p_2 étant les deux projections de E sur les facteurs de la somme directe. Pour chaque endomorphisme u de E , on peut définir quatre applications linéaires $u_{i,j}$, où $i, j = 1, 2$, et où $u_{i,j}$ est l'application linéaire de E_j dans E_i définie par $u_{i,j}(y) = p_i(u(y))$ pour tout $y \in E_j$. On peut alors représenter u par la matrice

$$\begin{pmatrix} u_{1,1} & u_{1,2} \\ u_{2,1} & u_{2,2} \end{pmatrix}.$$

On dira qu'on a effectué la *décomposition en blocs* de l'endomorphisme u par rapport à la décomposition $E = E_1 \oplus E_2$. On imagine facilement comment généraliser au cas où E est décomposé en une somme de plus de deux facteurs directs.

Si on se donne une base de chacun des deux espaces E_i de la somme directe, on pourra remplacer les matrices « abstraites » qui précèdent par de vraies matrices, mais considérées comme des matrices découpées en blocs bien identifiés. Par exemple, si $\mathbf{e}_1 = (z_1, \dots, z_p)$ et $\mathbf{e}_2 = (z_{p+1}, \dots, z_{p+q})$ sont des bases de E_1 et E_2 respectivement, si on décompose un vecteur $x \in E$ sous la forme $x = x_1 + x_2$ avec $x_i \in E_i$, $i = 1, 2$, on aura deux matrices colonne X_1 et X_2 des coordonnées de x_1 et de x_2 dans les bases correspondantes \mathbf{e}_1 et \mathbf{e}_2 . On notera les coordonnées de x dans la base $(z_1, \dots, z_p, z_{p+1}, \dots, z_{p+q})$ de E obtenue en prenant la base de E_1 suivie de la base de E_2 par une notation en blocs,

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{ représenté par } X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

Le vecteur colonne X a $n = p + q$ lignes, réparties en p lignes pour X_1 et q lignes pour X_2 . De même, chaque application $u_{i,j}$ sera représentée par sa matrice $A_{i,j}$ par rapport aux bases de E_j (au départ) et de E_i (à l'arrivée),

$$u = \begin{pmatrix} u_{1,1} & u_{1,2} \\ u_{2,1} & u_{2,2} \end{pmatrix} \text{ représenté par } A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}.$$

On dira qu'on a effectué la *décomposition par blocs* de la matrice A , décomposition associée à la décomposition $E = [z_1, \dots, z_p] \oplus [z_{p+1}, \dots, z_{p+q}]$. Les deux matrices $A_{1,1}$ et $A_{2,2}$ sont carrées, respectivement de tailles $p \times p$ et $q \times q$, mais les deux autres sont en général des matrices rectangulaires, de tailles $p \times q$ et $q \times p$. Si $Y \in \mathbb{K}^n$ représente les coordonnées de $u(x)$, on écrira

$$Y = AX = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} A_{1,1}X_1 + A_{1,2}X_2 \\ A_{2,1}X_1 + A_{2,2}X_2 \end{pmatrix}.$$

Supposons maintenant que v soit un autre endomorphisme de E , lui aussi représenté par une matrice en blocs $B = (B_{i,j})$ pour les mêmes bases. La composition $v \circ u$ sera représentée par la matrice produit BA , produit qu'on pourra calculer par blocs (en respectant bien l'ordre des facteurs matriciels),

$$BA = \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix} \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} = \begin{pmatrix} B_{1,1}A_{1,1} + B_{1,2}A_{2,1} & B_{1,1}A_{1,2} + B_{1,2}A_{2,2} \\ B_{2,1}A_{1,1} + B_{2,2}A_{2,1} & B_{2,1}A_{1,2} + B_{2,2}A_{2,2} \end{pmatrix}.$$

Il est bien sûr possible de généraliser ce qui précède à une somme directe d'un nombre fini quelconque de facteurs.

Exemple. On peut voir le produit usuel AB de deux matrices comme un produit par blocs, en groupant A par lignes et B par colonnes.

Matrices diagonales par blocs, triangulaires par blocs

Supposons que l'ensemble d'indices $\{1, \dots, n\}$ soit découpé en p intervalles disjoints I_1, \dots, I_p . On aura alors une décomposition du carré $\{1, \dots, n\}^2$ en p^2 rectangles $I_i \times I_j$, qui permettra de décomposer une matrice carrée A de taille $n \times n$ en p^2 blocs $(A_{i,j})$ où i et j varient tous les deux de 1 à p , et où chaque bloc diagonal $A_{i,i}$ est une matrice carrée ; on dit que A est *diagonale par blocs* si on a $A_{i,j} = 0$ pour tout couple (i, j) tel que $i \neq j$. Sous les mêmes conditions de décomposition on dit que A est *triangulaire (supérieure) par blocs* si $A_{i,j} = 0$ chaque fois que $i > j$. Le produit AB de deux matrices A et B diagonales par blocs, pour des décompositions correspondant à la même partition de $\{1, \dots, n\}$, est facile à exprimer : c'est la matrice diagonale par blocs dont les blocs diagonaux sont les produits $A_{i,i}B_{i,i}$ (dans le bon ordre !). On a donc dans ce cas $AB = BA$ si et seulement si $A_{i,i}B_{i,i} = B_{i,i}A_{i,i}$ pour tout $i = 1, \dots, n$. On voit aussi que le produit de deux matrices triangulaires par blocs est triangulaire par blocs. Les blocs diagonaux du produit sont les produits des blocs diagonaux.

Supposons que F soit un sous-espace vectoriel de E , et choisissons une base de E formée d'une base (f_1, \dots, f_p) de F complétée par des vecteurs (g_1, \dots, g_q) . Posons $G = [g_1, \dots, g_q]$. On a $u(F) \subset F$ (on dit dans ce cas que F est *stable* par u) si et seulement si la matrice de u est triangulaire par blocs dans la décomposition par blocs associée à la décomposition $E = [f_1, \dots, f_p] \oplus [g_1, \dots, g_q]$.

Un exemple de décomposition en blocs

Soit a un endomorphisme d'un espace vectoriel de dimension finie E . On a toujours

$$\{0_E\} \subset \ker a \subset \ker a^2 \subset \dots \subset \ker a^p \dots$$

Comme E est de dimension finie, cette chaîne de sous-espaces ne peut pas rester strictement croissante, donc il existe un plus petit entier $r \geq 0$ tel que $\ker a^r = \ker a^{r+1}$ (si $r = 0$, on a $\ker a = \{0_E\}$, donc a est injective). Supposons pour fixer les idées que $r = 2$, c'est-à-dire que $\{0_E\} \neq \ker a \neq \ker a^2 = \ker a^3$. On vérifie alors que $\ker a^p = \ker a^2$ pour tout $p \geq 2$ (par récurrence sur p). Ensuite,

$$\ker a^2 \cap a^2(E) = \{0_E\}.$$

En effet, si $x \in \ker a^2 \cap a^2(E)$, on a $a^2(x) = 0_E$ et on peut écrire $x = a^2(y)$ pour un certain $y \in E$. Alors $0_E = a^2(x) = a^4(y)$, donc $y \in \ker a^4 = \ker a^2$, ce qui donne $a^2(y) = x = 0_E$. Par ailleurs, d'après la proposition 4.2.1, on sait que $\dim \ker a^2 + \dim a^2(E) = \dim E$. On a donc

$$E = \ker a^2 \oplus a^2(E),$$

et cette décomposition est formée de deux sous-espaces stables par a . On a *a fortiori* $\ker a \cap a^2(E) = \{0_E\}$, donc la restriction de a au sous-espace $a^2(E)$ est injective. Si a_1 désigne la restriction de a à $\ker a^2$ et a_2 la restriction à $a^2(E)$, on voit que $a_1^2 = 0$ et a_2 est inversible.

Écrivons $\ker a^2 = \ker a \oplus F$, et soit (f_1, \dots, f_p) une base de F . Pour tout $j = 1, \dots, p$, on voit que $a(f_j) \in \ker a$, et a est injective sur F , donc $a(F) \subset \ker a$ est de même dimension que F , et les vecteurs $a(f_1), \dots, a(f_p)$ sont linéairement indépendants. On peut choisir un supplémentaire G de $a(F)$ dans $\ker a$, muni d'une base (g_1, \dots, g_q) et par rapport à la décomposition $\ker a^2 = [g_1, \dots, g_q] \oplus [a(f_1), f_1, \dots, a(f_p), f_p]$, la matrice de u_1 aura une décomposition par blocs de la forme

$$\begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix} \quad \text{où} \quad B = \begin{pmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & & & \ddots \end{pmatrix}.$$

Si on choisit une base (y_1, \dots, y_s) de $a^2(E)$, la matrice de a aura par rapport à la décomposition $E = [g_1, \dots, g_q] \oplus [a(f_1), f_1, \dots, f_p] \oplus [y_1, \dots, y_s]$ la décomposition en blocs

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix}$$

où C est une matrice inversible de taille $s \times s$ (c'est la matrice de a_2).

4.4. Déterminants

On considère un espace vectoriel E sur \mathbb{K} .

Définition 4.4.1. Une *forme p -linéaire alternée* sur E est une application φ de E^p dans \mathbb{K} telle que

1. Pour tout $j = 1, \dots, p$ et pour tous vecteurs (x_1, \dots, x_p) fixés dans E , l'application $x \in E \rightarrow \varphi(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_p)$ est linéaire de E dans \mathbb{K} .

2. S'il existe $i \neq j$ tel que $x_i = x_j$, alors $\varphi(x_1, \dots, x_p) = 0$.

Il en résulte que si $i \neq j$, la fonction φ change de signe quand on échange les variables x_i et x_j ,

$$\varphi(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_j, \dots, x_p) = -\varphi(x_1, \dots, x_{i-1}, x_j, x_{i+1}, \dots, x_i, \dots, x_p).$$

Proposition 4.4.1. Formes alternées et indépendance. *Si un système (x_1, \dots, x_p) de vecteurs de E est lié, toute fonction p -linéaire alternée sur E s'annule sur (x_1, \dots, x_p) .*

En effet, si par exemple $x_p = a_1 x_1 + \dots + a_{p-1} x_{p-1}$ on a en décomposant x_p

$$\varphi(x_1, \dots, x_p) = \varphi(x_1, \dots, x_{p-1}, \sum_{j=1}^{p-1} a_j x_j) = \sum_{j=1}^{p-1} a_j \varphi(x_1, \dots, x_j, \dots, x_{p-1}, x_j) = 0.$$

Construction de formes multilinéaires alternées

Supposons pour commencer que l'on s'est donné deux formes linéaires f_1, f_2 sur E . On pose alors pour tous $x_1, x_2 \in E$

$$(f_1 \wedge f_2)(x_1, x_2) = f_1(x_1)f_2(x_2) - f_1(x_2)f_2(x_1).$$

On peut voir que $f_1 \wedge f_2$ est linéaire par rapport à chaque variable et qu'elle est alternée. Si $f_1 = \mathbf{e}_1^*$, $f_2 = \mathbf{e}_2^*$, où $(\mathbf{e}_1, \mathbf{e}_2)$ est la base canonique de \mathbb{K}^2 , et si $x = (a, b)$ et $y = (c, d)$, on aura

$$(\mathbf{e}_1^* \wedge \mathbf{e}_2^*)(x, y) = ad - bc.$$

Il est facile de voir que la seule forme 2-linéaire alternée sur \mathbb{K}^2 telle que $\varphi(\mathbf{e}_1, \mathbf{e}_2) = 1$ est donnée par

$$\varphi(x, y) = ad - bc$$

si $x = (a, b)$ et $y = (c, d)$. En effet, on aura $\varphi(\mathbf{e}_1, \mathbf{e}_1) = \varphi(\mathbf{e}_2, \mathbf{e}_2) = 0$ et $\varphi(\mathbf{e}_2, \mathbf{e}_1) = -\varphi(\mathbf{e}_1, \mathbf{e}_2) = -1$, d'où le résultat en développant l'expression

$$\varphi(x, y) = \varphi(a\mathbf{e}_1 + b\mathbf{e}_2, c\mathbf{e}_1 + d\mathbf{e}_2).$$

Passage aux ordres supérieurs

Expliquons pour commencer le passage de deux à trois variables. Supposons données une forme linéaire f et une forme 2-linéaire alternée g sur E . Posons

$$(f \wedge g)(x_1, x_2, x_3) = f(x_1)g(x_2, x_3) - f(x_2)g(x_1, x_3) + f(x_3)g(x_1, x_2).$$

Il est facile de vérifier que $f \wedge g$ est 3-linéaire alternée. Si $f = e_1^*$ et $g = e_2^* \wedge e_3^*$, on posera

$$e_1^* \wedge e_2^* \wedge e_3^* = e_1^* \wedge (e_2^* \wedge e_3^*).$$

Dans le cas de \mathbb{K}^3 , si on prend $f = e_1^*$ et $g = e_2^* \wedge e_3^*$, on obtient, en développant g dans l'expression de la forme $e_1^* \wedge g$ ci-dessus, une expression contenant six produits de trois termes, et compte-tenu de l'expression de g sur deux vecteurs $x = (x_1, x_2, x_3)$ et $y = (y_1, y_2, y_3) \in \mathbb{K}^3$

$$g(x, y) = x_2 y_3 - x_3 y_2,$$

on aura si $x_i = (x_{i,1}, x_{i,2}, x_{i,3}) \in \mathbb{K}^3$ pour $i = 1, 2, 3$ l'expression

$$x_{1,1}(x_{2,2} x_{3,3} - x_{2,3} x_{3,2}) - x_{2,1}(x_{1,2} x_{3,3} - x_{1,3} x_{3,2}) + x_{3,1}(x_{1,2} x_{2,3} - x_{1,3} x_{2,3}).$$

C'est le déterminant de la matrice 3×3 des coordonnées des trois vecteurs (x_1, x_2, x_3) .

Expliquons maintenant le pas général pour passer de $p - 1$ variables à p variables. Supposons données une forme linéaire f et une forme $(p - 1)$ -linéaire alternée g sur E . Posons

$$(f \wedge g)(x_1, x_2, \dots, x_p) = f(x_1)g(x_2, x_3, \dots, x_p) - f(x_2)g(x_1, x_3, x_4, \dots, x_p) + f(x_3)g(x_1, x_2, x_4, \dots, x_p) + \dots + (-1)^{p-1} f(x_p)g(x_1, \dots, x_{p-2}, x_{p-1}).$$

Il est facile de vérifier que $f \wedge g$ est p -linéaire alternée : en effet, il est clair que chacun des p termes de la somme est linéaire par rapport à chacune des variables x_i , et d'autre part, si par exemple on suppose $x_1 = x_2$, on aura $g(x_1, x_2, \dots) = 0$ pour tous les termes où g contient les deux vecteurs x_1 et x_2 (parce que g est alternée) ; de la somme ci-dessus, il restera donc seulement les deux premiers termes

$$(f \wedge g)(x_1, x_1, x_3, \dots, x_p) = f(x_1)g(x_1, x_3, \dots, x_p) - f(x_1)g(x_1, x_3, \dots, x_p) = 0.$$

Supposons que (e_1, \dots, e_n) soit une base de E et calculons $(e_1^* \wedge \dots \wedge e_n^*)(e_1, \dots, e_n)$. Posons $g = e_2^* \wedge \dots \wedge e_n^*$, on aura

$$(e_1^* \wedge g)(e_1, \dots, e_n) = e_1^*(e_1)g(e_2, \dots, e_n) + 0 + \dots + 0 = (e_2^* \wedge \dots \wedge e_n^*)(e_2, \dots, e_n)$$

parce que $e_1^*(e_j)$ est nul pour tout $j \neq 1$. En continuant de proche en proche on arrive à

$$(e_1^* \wedge \dots \wedge e_n^*)(e_1, \dots, e_n) = \dots = e_1^*(e_1) = 1.$$

On a donc montré que sur tout espace de dimension $n \geq 1$ existe une forme n -linéaire alternée non nulle.

Notons S_n l'ensemble des permutations de $\{1, \dots, n\}$, et soit $\sigma \in S_n$; on sait que σ peut être obtenue comme produit d'un certain nombre k de transpositions. Si on considère une fonction n -linéaire alternée φ , l'expression $\varphi(x_{\sigma(1)}, \dots, x_{\sigma(n)})$ est donc obtenue à partir de $\varphi(x_1, \dots, x_n)$ par k transpositions des vecteurs, donc

$$\varphi(x_{\sigma(1)}, \dots, x_{\sigma(n)}) = (-1)^k \varphi(x_1, \dots, x_n).$$

Supposons maintenant que (e_1, \dots, e_n) soit une base de E et que φ_e soit une fonction n -linéaire alternée sur E telle que $\varphi_e(e) = 1$. On aura $\varphi_e(e_{\sigma(1)}, \dots, e_{\sigma(n)}) = (-1)^k$, ce qui montre que l'expression $(-1)^k$ ne dépend que de σ . C'est la *signature* de la permutation σ , signature qui sera notée $\varepsilon(\sigma)$. On a donc

$$\varphi(x_{\sigma(1)}, \dots, x_{\sigma(n)}) = \varepsilon(\sigma) \varphi(x_1, \dots, x_n)$$

pour toute fonction alternée φ sur l'espace E , toute permutation σ et tout système de vecteurs (x_1, \dots, x_n) .

Théorème 4.4.1. Soit E un espace vectoriel de dimension $n \geq 1$, muni d'une base $\mathbf{e} = (e_1, \dots, e_n)$; soient x_1, \dots, x_n des vecteurs de E , et $x_j = \sum_{i=1}^n a_{i,j} e_i$ leur décomposition dans la base \mathbf{e} , pour $j = 1, \dots, n$. Pour toute forme n -linéaire alternée φ sur E , on a

$$\varphi(x_1, \dots, x_n) = \left(\sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1),1} \dots a_{\sigma(n),n} \right) \varphi(e_1, \dots, e_n).$$

Il existe une et une seule forme n -linéaire alternée $\varphi_{\mathbf{e}}$ sur E telle que $\varphi_{\mathbf{e}}(e_1, \dots, e_n) = 1$. Toute autre forme n -linéaire alternée ψ sur E est proportionnelle à $\varphi_{\mathbf{e}}$. L'espace vectoriel des formes n -linéaires alternées sur un espace de dimension n (le même n aux deux endroits!) est donc de dimension 1.

Expliquons la démonstration dans le cas $n = 3$ pour simplifier. En utilisant la multilinéarité, on obtient en développant

$$\varphi(x_1, x_2, x_3) = \sum a_{i_1,1} a_{i_2,2} a_{i_3,3} \varphi(e_{i_1}, e_{i_2}, e_{i_3}),$$

où la somme porte sur les 27 choix possibles de $(i_1, i_2, i_3) \in \{1, 2, 3\}^3$. Mais $\varphi(e_{i_1}, e_{i_2}, e_{i_3})$ est nul chaque fois que deux des indices i_j sont égaux. Ils ne reste plus alors que les six choix correspondant à une permutation σ de l'ensemble $\{1, 2, 3\}$, c'est-à-dire $(i_1, i_2, i_3) = (\sigma(1), \sigma(2), \sigma(3))$, donc

$$\varphi(x_1, x_2, x_3) = \sum_{\sigma \in S_3} a_{\sigma(1),1} a_{\sigma(2),2} a_{\sigma(3),3} \varphi(e_{\sigma(1)}, e_{\sigma(2)}, e_{\sigma(3)}),$$

et on termine en utilisant $\varphi(e_{\sigma(1)}, e_{\sigma(2)}, e_{\sigma(3)}) = \varepsilon(\sigma) \varphi(e_1, e_2, e_3)$. On a déjà vu l'existence et l'unicité de $\varphi_{\mathbf{e}}$,

$$\varphi_{\mathbf{e}} = e_1^* \wedge \dots \wedge e_n^*, \quad \varphi_{\mathbf{e}}(x_1, \dots, x_n) = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1),1} \dots a_{\sigma(n),n}.$$

D'après les formules précédentes il est clair que pour toute forme n -linéaire alternée ψ sur E on a $\psi = \psi(\mathbf{e}) \varphi_{\mathbf{e}}$, ce qui montre que ψ est proportionnelle à $\varphi_{\mathbf{e}}$.

Remarques.

1. Si ψ_1 et ψ_2 sont deux formes n -linéaires alternées sur E et si $\psi_1(e_1, \dots, e_n) = \psi_2(e_1, \dots, e_n)$, alors $\psi_1 = \psi_2$.

2. L'unique forme n -linéaire alternée $\varphi_{\mathbf{e}}$ sur E telle que $\varphi_{\mathbf{e}}(e_1, \dots, e_n) = 1$ est donc donnée par la formule

$$\varphi_{\mathbf{e}}(x_1, \dots, x_n) = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1),1} \dots a_{\sigma(n),n}.$$

On aurait pu partir directement de cette formule pour définir $e_1^* \wedge \dots \wedge e_n^*$, mais l'autre méthode est peut-être un peu moins féroce!

Déterminant dans une base

Soit $\mathbf{e} = (e_1, \dots, e_n)$ une base de E ; on appelle *déterminant par rapport à la base \mathbf{e}* l'unique forme n -linéaire alternée $\varphi_{\mathbf{e}}$ telle que $\varphi_{\mathbf{e}}(e_1, \dots, e_n) = 1$. On le notera $\det_{\mathbf{e}}$.

Déterminant et indépendance. Si un système $\mathbf{x} = (x_1, \dots, x_n)$ est lié, on a vu dans la proposition 4.4.1 que $\det_{\mathbf{e}}(\mathbf{x}) = 0$. Réciproquement, si $\mathbf{x} = (x_1, \dots, x_n)$ est une base de

E, la fonction $\det_{\mathbf{x}}$ est proportionnelle à $\det_{\mathbf{e}}$, donc il existe λ tel que $\det_{\mathbf{x}} = \lambda \det_{\mathbf{e}}$; en appliquant au système \mathbf{x} on obtient $\det_{\mathbf{x}}(\mathbf{x}) = 1 = \lambda \det_{\mathbf{e}}(\mathbf{x})$, ce qui montre en particulier que $\det_{\mathbf{e}}(\mathbf{x}) \neq 0$. En appliquant au système \mathbf{e} on voit que $\lambda = \det_{\mathbf{x}}(\mathbf{e})$, donc

$$\det_{\mathbf{x}} = \det_{\mathbf{x}}(\mathbf{e}) \det_{\mathbf{e}}.$$

Exemple d'application du principe d'unicité. La surface orientée $S(x_1, x_2)$ du parallélogramme construit sur deux vecteurs x_1, x_2 de \mathbb{R}^2 est une forme 2-linéaire alternée, égale à 1 pour la base canonique, donc elle est égale au déterminant dans la base canonique. On peut généraliser aux volumes dans \mathbb{R}^3 .

Déterminant des matrices

Considérons une matrice $n \times n$ comme formée de n vecteurs colonne. L'unique forme n -linéaire alternée sur \mathbb{K}^n qui vaut 1 sur la base canonique donne le déterminant des matrices $n \times n$. Autrement dit :

Le déterminant d'une matrice $n \times n$ est l'unique forme n -linéaire des vecteurs colonne, qui est nulle quand deux colonnes sont égales, et vaut 1 pour la matrice unité I_n de taille $n \times n$.

Le déterminant de la matrice $A = (a_{i,j})$ de taille $n \times n$ est donc donné par la « grosse formule »

$$\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1),1} \dots a_{\sigma(n),n}.$$

On remarquera qu'on pourrait remplacer les nombres $a_{i,j}$ dans l'expression ci-dessus par des objets plus généraux, en particulier des polynômes à coefficients dans \mathbb{K} . Si A est une matrice dont les coefficients sont des polynômes $P_{i,j}$, son déterminant sera un polynôme $D = \det A$. Il est intéressant de noter que pour tout $a \in \mathbb{K}$, on aura alors $D(a) = \det(P_{i,j}(a))$: la valeur au point a du déterminant-polynôme est égale au déterminant de la matrice des valeurs au point a des coefficients-polynômes.

Développement du déterminant d'une matrice suivant une ligne

Soit $A = (a_{i,j})$ une matrice de taille $n \times n$. Pour tous $i, j = 1, \dots, n$, désignons par $A^{(i,j)}$ la matrice $(n-1) \times (n-1)$ obtenue en effaçant dans A la ligne i et la colonne j , et considérons l'expression

$$\varphi_i(A) = \sum_{j=1}^n (-1)^{i+j} a_{i,j} \det A^{(i,j)}.$$

On peut déduire de la caractérisation précédente que $\varphi_i(A) = \det A$ pour tout indice $i = 1, \dots, n$: en effet, la formule obtenue par ce développement suivant une ligne définit une forme n -linéaire alternée des vecteurs colonne d'une matrice, égale à 1 pour la matrice identité I_n : si $A = I_n$, on aura $a_{i,j} = 0$ si $i \neq j$, et $a_{i,i} = 1$, $A^{(i,i)} = I_{n-1}$, donc

$$\varphi_i(I_n) = (-1)^{2i} \det I_{n-1} = 1.$$

Transposition. Déterminant et lignes de la matrice. La « grosse formule » du théorème 4.4.1 permet de voir que le déterminant de la transposée tA de la matrice A est le même que le déterminant de la matrice A . Il en résulte que le déterminant est aussi une fonction alternée des lignes de la matrice, et qu'on peut caractériser la fonction qui

associe à chaque matrice son déterminant comme l'unique fonction définie sur l'espace des matrices qui est une forme n -linéaire alternée des lignes des matrices et qui est égale à 1 pour la matrice unité I_n . On déduit de ce qui précède la formule de développement suivant une colonne (qui revient à appliquer à la transposée le développement suivant une ligne).

Remarque. Soient \mathbf{e} une base de E , $\mathbf{x} = (x_1, \dots, x_n)$ un système de vecteurs de E et A la matrice carrée dont les vecteurs colonne sont formés par les coordonnées des vecteurs x_j dans la base \mathbf{e} . Le déterminant de la matrice A est égal à $\det_{\mathbf{e}}(x_1, \dots, x_n)$.

En effet, on obtient quand le système \mathbf{x} varie deux fonctions n -linéaires alternées sur E qui coïncident sur \mathbf{e} , donc sont égales sur E^n .

Cette remarque permet de montrer que le déterminant d'un produit de matrices est égal au produit des déterminants. On voit aussi que le déterminant d'une matrice A est non nul si et seulement si elle est inversible.

Déterminant d'une matrice triangulaire. La « grosse formule », ou bien plus simplement le développement par rapport à la première colonne et une récurrence, permettent de voir que le déterminant d'une matrice triangulaire (supérieure) est égal au produit des coefficients diagonaux.

Proposition 4.4.2. *Considérons une matrice carrée M de la forme*

$$M = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$$

où A et D sont des matrices carrées. Le déterminant de M est égal à $\det A \det D$.

Démonstration. Supposons que A soit de taille $p \times p$ et D de taille $q \times q$. Considérons lorsque B et D sont fixées la fonction

$$A \rightarrow \varphi(A) = \det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}.$$

C'est une forme p -linéaire alternée des colonnes de A , qui est donc proportionnelle à la fonction $A \rightarrow \det A$, donc

$$\varphi(A) = \varphi(I_p) \det A = \det \begin{pmatrix} I_p & B \\ 0 & D \end{pmatrix} \det A.$$

Il est facile de terminer en montrant que

$$\psi(D) = \det \begin{pmatrix} I_p & B \\ 0 & D \end{pmatrix} = \det D$$

soit par le calcul (développer suivant la première colonne, puis récurrence sur p), soit en considérant ce déterminant $\psi(D)$ comme une fonction alternée des lignes de D et en raisonnant comme pour φ .

Bien entendu la formule se généralise dans le cas où la matrice M admet une décomposition en blocs qui est triangulaire supérieure, avec plus de deux blocs diagonaux carrés. Le déterminant de M dans ce cas est égal au produit des déterminants des blocs diagonaux.

Exercice. Démontrer ce qui précède avec la « grosse formule », ou bien par développement suivant la première colonne et récurrence sur p .

ATTENTION!! La formule qui pourrait sembler « naturelle » pour calculer le déterminant d'une matrice carrée décomposée en blocs (tous carrés) de la forme

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

n'est pas vraie : le déterminant de M n'est pas en général égal à $\det A \det D - \det B \det C$.

Exercice.

a. Si A et D sont carrées et si A est inversible, le déterminant de la matrice M précédente est égal à $\det M = \det A \det(D - CA^{-1}B)$.

b. Si A et D sont de même dimension et si A commute avec C (c'est-à-dire que $AC = CA$), on a $\det M = \det(AD - CB)$.

Indépendance linéaire et changement de corps de base

Soient x_1, \dots, x_p des vecteurs de \mathbb{R}^n , linéairement indépendants. Si on les considère maintenant comme des vecteurs du \mathbb{C} -espace vectoriel \mathbb{C}^n , on aura beaucoup plus de choix de scalaires pour former des combinaisons linéaires (complexes), donc il n'est pas tout à fait évident que les vecteurs proposés restent \mathbb{C} -indépendants, c'est-à-dire que les conditions

$$\lambda_1, \dots, \lambda_p \in \mathbb{C}, \quad \sum_{j=1}^p \lambda_j x_j = 0_{\mathbb{C}^n}$$

impliquent $\lambda_1 = \dots = \lambda_p = 0$. C'est pourtant vrai. Dans le cas de \mathbb{R} et \mathbb{C} , on peut y arriver assez facilement en décomposant les coefficients λ_j en partie réelle et partie imaginaire, mais nous allons voir un résultat plus général valable pour tout sous-corps L de \mathbb{C} .

Proposition 4.4.3. *Soient L un sous-corps de \mathbb{C} et x_1, \dots, x_p des vecteurs de L^n ; si les vecteurs x_1, \dots, x_p sont linéairement indépendants dans le L -espace vectoriel L^n , ils restent \mathbb{C} -indépendants quand on les considère comme vecteurs de \mathbb{C}^n .*

Démonstration. Complétons le système (x_1, \dots, x_p) en une base (x_1, \dots, x_n) de L^n . Soit A la matrice dont les colonnes sont x_1, \dots, x_n . Alors $\det A \neq 0$. Mais quand on travaille dans \mathbb{C} , le calcul du déterminant est le même, donc les vecteurs sont aussi \mathbb{C} -indépendants.

Symétrisation et antisymétrisation

Considérons la fonction de trois variables réelles $f(x_1, x_2, x_3) = x_1 x_2 x_3$. Cette fonction est certainement *symétrique* par rapport à l'ensemble des variables x_1, x_2, x_3 . En revanche, la fonction $f(x_1, x_2, x_3) = x_1 x_2^2$ n'est visiblement pas symétrique. Comment lui associer de façon « naturelle » une fonction symétrique? Essayons la fonction

$$g(x_1, x_2, x_3) = x_1 x_2^2 + x_2 x_3^2 + x_3 x_1^2$$

qui a déjà l'air un peu plus symétrique. Mais elle n'est pas vraiment symétrique, mais seulement invariante par permutation circulaire $1 \rightarrow 2, 2 \rightarrow 3$ et $3 \rightarrow 1$; si on effectue la transposition $(1, 2)$ on obtient en reclassant les termes

$$x_1^2 x_2 + x_2^2 x_3 + x_3^2 x_1$$

qui n'est pas la même fonction que g !

Pour comprendre le procédé de symétrisation, introduisons une famille d'applications bijectives T_σ de $Z = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ dans Z , définies pour toute permutation σ de $\{1, 2, 3\}$ par

$$T_\sigma(z) = (x_{\sigma^{-1}(1)}, x_{\sigma^{-1}(2)}, x_{\sigma^{-1}(3)}) \text{ pour tout } z = (x_1, x_2, x_3) \in Z.$$

On note que $T_{id} = id_Z$ et $T_{\sigma_1} \circ T_{\sigma_2} = T_{\sigma_1 \circ \sigma_2}$ pour tout couple σ_1, σ_2 de permutations de $\{1, 2, 3\}$. On associera alors à toute fonction f de trois variables la nouvelle fonction f_s définie pour tout $z = (x_1, x_2, x_3)$ par la formule

$$f_s(z) = \frac{1}{3!} \sum_{\sigma \in S_3} f(T_\sigma(z)) = \frac{1}{6} \left(f(x_1, x_2, x_3) + f(x_1, x_3, x_2) + f(x_2, x_1, x_3) + f(x_2, x_3, x_1) + f(x_3, x_1, x_2) + f(x_3, x_2, x_1) \right).$$

Cette procédure se généralise à tout ensemble Z sur lequel est définie une *action* T du groupe S_n . On se limitera ici au cas où l'ensemble Z est un produit X^n , et où l'action d'une permutation $\sigma \in S_n$ sur un élément $z = (x_1, \dots, x_n) \in Z$ est définie par

$$T_\sigma(z) = (x_{\sigma^{-1}(1)}, \dots, x_{\sigma^{-1}(n)}),$$

c'est-à-dire que l'action consiste simplement à permuter les coordonnées du produit X^n . Il est encore clair que $T_{id} = id_Z$ et que $T_{\sigma_1} \circ T_{\sigma_2} = T_{\sigma_1 \circ \sigma_2}$ pour tout couple $\sigma_1, \sigma_2 \in S_n$.

Définition. La fonction réelle f définie sur $Z = X^n$ est *symétrique* si on a $f(z) = f(T_\sigma(z))$ pour tout $z \in Z$ et pour toute permutation $\sigma \in S_n$.

On associera encore à toute fonction réelle f définie sur Z la nouvelle fonction f_s définie pour tout $z = (x_1, \dots, x_n)$ par la formule

$$f_s(z) = \frac{1}{n!} \sum_{\sigma \in S_n} f(T_\sigma(z)).$$

Si f est symétrique, il est clair que $f = f_s$, parce que tous les termes $f(T_\sigma(z))$ de la somme sont égaux à $f(z)$. Mais en fait, cette fonction f_s est symétrique pour toute fonction f . En effet

$$f(T_\sigma(z)) = \frac{1}{n!} \sum_{\pi \in S_n} f(T_\pi(T_\sigma(z))) = \frac{1}{n!} \sum_{\pi \in S_n} f(T_{\pi \circ \sigma}(z)).$$

La permutation σ étant fixée, on vérifie que l'application $\pi \in S_n \rightarrow \pi \circ \sigma$ est une bijection de S_n . Par conséquent, si on effectue le « changement de variable » $\rho = \pi \circ \sigma$ dans l'expression ci-dessus, on verra que

$$f(T_\sigma(z)) = \frac{1}{n!} \sum_{\rho \in S_n} f(T_\rho(z)) = f_s(z).$$

Exemples.

1. Symétriser en trois variables les fonctions $f(x_1, x_2, x_3) = x_1 x_2$, $f(x_1, x_2, x_3) = x_1 x_2^2$.

2. Prenons un cas un peu plus compliqué. Prenons $X = \mathbb{R}^3$, puis $Z = X^3$, et pour éviter les doubles indices écrivons chaque $x \in X$ comme $x = (a, b, c)$, ce qui donnera $x_j = (a_j, b_j, c_j)$ pour chaque vecteur x_j composante de $z = (x_1, x_2, x_3) \in Z$. Symétriser la fonction $f(x_1, x_2, x_3) = a_1 b_2 c_3$.

Passons maintenant à l'antisymétrie. Supposons comme avant que $Z = X^n$, et faisons encore agir S_n par permutation des coordonnées. Une fonction réelle g définie sur Z est *antisymétrique* si

$$g(T_\tau(z)) = -g(z)$$

pour toute *transposition* $\tau \in S_n$. Il en résulte alors

$$g(T_\sigma(z)) = \varepsilon(\sigma)g(z)$$

pour toute permutation σ (où $\varepsilon(\sigma)$ désigne la signature de la permutation σ).

Exemple. La fonction $g(x_1, x_2) = x_1 - x_2$ est antisymétrique sur $Z = \mathbb{R} \times \mathbb{R}$.

Exercice. Montrer que

$$\varepsilon(\sigma) = \prod_{i < j} \frac{\sigma(j) - \sigma(i)}{j - i}.$$

Pour antisymétriser une fonction f , on posera pour tout $z \in Z = X^n$

$$f_a(z) = \frac{1}{n!} \sum_{\sigma \in S_n} \varepsilon(\sigma) f(T_\sigma(z)).$$

Si f est déjà antisymétrique, on a $f_a = f$. Pour montrer que f_a est antisymétrique pour toute fonction f , on écrit, en désignant par τ une transposition quelconque de $\{1, \dots, n\}$

$$f(T_\tau(z)) = \frac{1}{n!} \sum_{\pi \in S_n} \varepsilon(\pi) f(T_\pi(T_\tau(z))) = -\frac{1}{n!} \sum_{\pi \in S_n} \varepsilon(\pi \circ \tau) f(T_{\pi \circ \tau}(z)),$$

et on termine comme dans le cas symétrique.

Exercice. Antisymétriser en trois variables les deux fonctions

$$f(x_1, x_2, x_3) = x_1 x_2, \quad f(x_1, x_2, x_3) = x_1 x_2^2.$$

Revenons au cas linéaire. Soient E un espace de dimension n , $\mathbf{e} = (e_1, \dots, e_n)$ une base de E et e_1^*, \dots, e_n^* les fonctions coordonnées dans cette base ; posons sur $Z = E^n$

$$f(x_1, \dots, x_n) = e_1^*(x_1) \dots e_n^*(x_n).$$

Cette fonction est linéaire par rapport à chacune des variables $x_1, \dots, x_n \in E$. Il en résulte que l'antisymétrisée f_a est elle aussi linéaire par rapport à chaque variable. Considérons

$$n! f_a(e_1, \dots, e_n) = \sum_{\sigma \in S_n} \varepsilon(\sigma) e_1^*(e_{\sigma(1)}) \dots e_n^*(e_{\sigma(n)}).$$

Chacun des produits $e_1^*(e_{\sigma(1)}) \dots e_n^*(e_{\sigma(n)})$ est nul si $\sigma \neq id$, et égal à 1 pour $\sigma = id$, donc

$$n! f_a(e_1, \dots, e_n) = 1.$$

On a donc trouvé une forme n -linéaire sur E , non identiquement nulle, à savoir $n! f_a$. C'est le déterminant dans la base \mathbf{e} .

4.5. Formes linéaires. Espace dual

Dans cette section, on désigne par E un espace vectoriel sur un corps \mathbb{K} ; pour l'essentiel, le lecteur pourra considérer que \mathbb{K} sera égal à \mathbb{R} ou à \mathbb{C} ; l'amateur de mathématiques pourra néanmoins observer avec intérêt que les résultats sont valables pour un corps \mathbb{K} quelconque.

Une *forme linéaire* sur E est une application \mathbb{K} -linéaire de E dans \mathbb{K} . Étant données deux formes linéaires f et g sur E , on définit leur somme $f + g$ par l'opération usuelle de somme de deux fonctions,

$$\forall x \in E, \quad (f + g)(x) = f(x) + g(x).$$

Il est facile de vérifier que $f + g$ est encore une forme linéaire. Pour tout $\lambda \in \mathbb{K}$, on définit aussi la fonction λf par la formule $(\lambda f)(x) = \lambda f(x)$. Cette fonction λf est linéaire, et muni de ces deux opérations, l'ensemble des formes linéaires sur E est un espace vectoriel sur \mathbb{K} . C'est l'*espace dual* de E , noté E^* .

L'image $f(E)$ d'une forme linéaire f est un \mathbb{K} -sous-espace vectoriel de \mathbb{K} , donc elle est égale à $\{0\}$ ou à \mathbb{K} . Une forme linéaire f sur E est non nulle si et seulement s'il existe un vecteur $x \in E$ tel que $f(x) = 1$. Si E est de dimension n et si f est une forme linéaire non nulle sur E , le noyau $\ker f$ est un sous-espace vectoriel de dimension $n - 1$ de E . On appelle *hyperplan* (vectoriel) de E tout sous-espace vectoriel F de E qui est de *codimension* 1 c'est-à-dire tel que $F \neq E$ et que $E = F + \mathbb{K}x$ pour tout vecteur $x \notin F$ (autrement dit, quand on a F , il ne manque plus qu'une dimension pour arriver à E tout entier). Si E est de dimension finie $n > 0$, un sous-espace F de E est un hyperplan de E si et seulement si $\dim F = n - 1$. Un *hyperplan affine* est un sous-ensemble obtenu en translatant un hyperplan vectoriel. Le noyau d'une forme linéaire f sur E non nulle est donc un hyperplan de E , et les ensembles de la forme $H_c = \{x \in E : f(x) = c\}$ sont des hyperplans affines. Dans le cas réel (lorsque $\mathbb{K} = \mathbb{R}$), cet hyperplan affine H_c sépare l'espace E en deux *demi-espaces*, à savoir $\{x \in E : f(x) \leq c\}$ et $\{x \in E : f(x) \geq c\}$.

Exercice facile. Montrer sans supposer la dimension de E finie que le noyau d'une forme linéaire non nulle est de codimension 1.

Base duale d'une base de E

Supposons que l'espace vectoriel E soit de dimension finie $n > 0$, et supposons donnée une base $\mathbf{e} = (e_1, \dots, e_n)$ de l'espace E ; on définit un système \mathbf{e}^* de formes linéaires sur E à partir de cette base, de la façon suivante : pour chaque $i = 1, \dots, n$, on désigne par e_i^* la fonction scalaire définie sur E , qui associe à chaque vecteur x de E sa i ème coordonnée dans la base \mathbf{e} , et on pose $\mathbf{e}^* = (e_1^*, \dots, e_n^*)$. On peut définir toutes ces fonctions (e_i^*) par une formule (implicite) unique,

$$(*) \quad \forall x \in E, \quad x = \sum_{i=1}^n e_i^*(x) e_i.$$

On remarque que

$$e_i^*(e_j) = \delta_{i,j},$$

où $\delta_{i,j}$, appelé *symbole de Kronecker*, est égal à 1 si $i = j$ et à 0 sinon (en d'autres termes, la matrice des coefficients $(\delta_{i,j})$ est la matrice unité I_n).

Proposition 4.5.1. *Le système \mathbf{e}^* est une base de l'espace dual E^* . En conséquence, lorsque E est de dimension finie, on a $\dim E^* = \dim E$.*

On dit que \mathbf{e}^* est la *base duale* de la base \mathbf{e} de E .

Démonstration. Soit x^* une forme linéaire sur E ; en appliquant x^* à la décomposition d'un vecteur $x \in E$ quelconque donnée par la formule (*), on obtient

$$\forall x \in E, \quad x^*(x) = \sum_{i=1}^n e_i^*(x) x^*(e_i) = \left(\sum_{i=1}^n x^*(e_i) e_i^* \right)(x).$$

En termes de fonctions sur E , ceci signifie que $x^* = \sum_{i=1}^n x^*(e_i) e_i^*$, et montre que le système de formes linéaires \mathbf{e}^* est générateur pour l'espace vectoriel dual E^* . Si on écrit une combinaison linéaire $x^* = \sum_{i=1}^n c_i e_i^*$, on trouve, en appliquant la forme linéaire x^* au vecteur e_j , la relation $c_j = x^*(e_j)$ qui montre que les coefficients (c_j) de la combinaison linéaire sont uniquement déterminés, donc \mathbf{e}^* est une base du dual E^* .

Calcul des coordonnées d'une forme linéaire dans une base duale

Si \mathbf{e}^* est la base duale d'une base $\mathbf{e} = (e_1, \dots, e_n)$ de E , et si $x^* = c_1 e_1^* + \dots + c_n e_n^*$, on vient de dire qu'on trouve les coefficients (c_i) au moyen de la formule

$$c_i = x^*(e_i), \quad i = 1, \dots, n;$$

cette formule évidente sera utilisée à plusieurs reprises dans ce chapitre.

Remarque. Étant donné un système de vecteurs $\mathbf{x} = (x_1, \dots, x_n)$ dans E (espace vectoriel de dimension n) et un système de formes linéaires (y_1^*, \dots, y_n^*) dans E^* , les relations

$$\forall i, j = 1, \dots, n, \quad y_i^*(x_j) = \delta_{i,j}$$

impliquent que le système $\mathbf{x} = (x_1, \dots, x_n)$ est une base de E et que (y_1^*, \dots, y_n^*) est la base duale de la base \mathbf{x} de E .

Proposition 4.5.2. *Soit f une forme linéaire non nulle sur E , et soit $H = \ker f$ son noyau ; si g est une forme linéaire sur E qui est nulle en tout point de H , il existe un scalaire $\lambda \in \mathbb{K}$ tel que $g = \lambda f$ (autrement dit, g est proportionnelle à f).*

Démonstration. Puisque f est non nulle, on peut trouver un vecteur $x_0 \in E$ tel que $f(x_0) = 1$. Supposons que g soit une forme linéaire nulle sur le noyau H de f , et posons $\lambda = g(x_0)$. Soit y un vecteur quelconque de E ; écrivons

$$y = (y - f(y)x_0) + f(y)x_0.$$

Posons $z = y - f(y)x_0$. On a $f(z) = f(y) - f(y)f(x_0) = 0$, donc $z \in H$ et $g(z) = 0$ par hypothèse. On a donc $g(y) = g(z) + f(y)g(x_0) = f(y)g(x_0) = \lambda f(y)$, ce qui montre bien que $g = \lambda f$.

On peut résumer la démonstration ainsi : la forme linéaire $g - \lambda f$ est nulle sur E , parce qu'elle est nulle sur H et sur $\mathbb{K}x_0$, et que $E = H \oplus \mathbb{K}x_0$; elle est nulle sur H puisque pour tout $y \in H$, on a $(g - \lambda f)(y) = g(y) - \lambda g(y) = 0 - 0 = 0$, et nulle sur $\mathbb{K}x_0$ puisque $(g - \lambda f)(x_0) = \lambda - \lambda = 0$.

Corollaire 4.5.1. *Deux formes linéaires qui ont le même noyau sont proportionnelles.*

Proposition 4.5.3. *Soit E un espace vectoriel de dimension finie ; pour tout vecteur non nul $x \in E$, il existe une forme linéaire $x^* \in E^*$ telle que $x^*(x) \neq 0$.*

Démonstration. Puisque x est non nul, on peut trouver d'après le théorème de la base incomplète une base $\mathbf{e} = (e_1, \dots, e_n)$ de E telle que $e_1 = x$. La forme linéaire e_1^* de la base duale \mathbf{e}^* convient, puisque $e_1^*(x) = e_1^*(e_1) = 1 \neq 0$.

Proposition 4.5.4. *Soient f_1, \dots, f_k des formes linéaires indépendantes sur un espace vectoriel E de dimension finie ; posons*

$$M = \{x \in E : \forall i = 1, \dots, k, f_i(x) = 0\}$$

(c'est l'intersection des noyaux $\ker f_j$ des formes linéaires considérées). La dimension du sous-espace vectoriel M est égale à $\dim E - k$.

Démonstration. Posons $n = \dim E$, et complétons le système (f_1, \dots, f_k) en une base (f_1, \dots, f_n) du dual E^* . Considérons l'application linéaire $u : E \rightarrow \mathbb{K}^n$ définie par

$$u(x) = (f_1(x), \dots, f_n(x)) \in \mathbb{K}^n.$$

Si $x \in \ker u$, on a $f_i(x) = 0$ pour tout $i = 1, \dots, n$, donc $f(x) = 0$ pour toute forme linéaire $f \in E^*$ (micro-exercice), donc $x = 0_E$ d'après la proposition précédente 4.5.3. Il en résulte que u est injective, donc u est un isomorphisme puisque $\dim E^* = \dim E = n = \dim \mathbb{K}^n$. On voit alors que M est l'image par l'isomorphisme inverse u^{-1} du sous-espace M' de \mathbb{K}^n de dimension $n - k$ défini par

$$M' = \{(t_1, \dots, t_n) \in \mathbb{K}^n : t_1 = t_2 = \dots = t_k = 0\}$$

donc $\dim M = \dim M' = n - k$.

Remarque. Si on ne suppose pas que les formes (f_1, \dots, f_k) sont indépendantes, on verra que $\dim M = \dim E - \dim[f_1, \dots, f_k]$ (où la notation $[f_1, \dots, f_k]$ désigne le sous-espace vectoriel de E^* engendré par f_1, \dots, f_k). En particulier, on a toujours $\dim M \geq \dim E - k$.

Corollaire 4.5.2. *Si (f_1, \dots, f_n) est une base de E^* , il existe une base x_1, \dots, x_n de E telle que $f_i(x_j) = \delta_{i,j}$ pour tous $i, j = 1, \dots, n$ (toute base de E^* est donc la base duale d'une base de E).*

Démonstration. Comme dans la démonstration de la proposition précédente, on introduit l'isomorphisme u de E sur \mathbb{K}^n défini par $u(x) = (f_1(x), \dots, f_n(x)) \in \mathbb{K}^n$ pour tout $x \in E$. Pour trouver les vecteurs (x_j) il suffit de considérer les images par l'inverse u^{-1} des vecteurs de la base canonique de \mathbb{K}^n .

Application linéaire transposée

Soit $a : E \rightarrow F$ une application linéaire ; on définit une application ${}^t a : F^* \rightarrow E^*$ par la formule

$$\forall y^* \in F^*, \quad {}^t a(y^*) = y^* \circ a \in E^*.$$

On vérifie facilement que cette application ${}^t a$ est linéaire de F^* dans E^* . Si on se donne une deuxième application linéaire $b : F \rightarrow G$, on voit que

$${}^t(b \circ a) = {}^t a \circ {}^t b$$

(bien noter l'interversion de a et b). La transposée de l'identité de E est l'identité de E^* . Il résulte des deux propriétés précédentes que la transposée d'un isomorphisme est un isomorphisme, et que l'on a dans ce cas $({}^t a)^{-1} = {}^t(a^{-1})$.

Supposons que E et F soient de dimensions finies m et n , que \mathbf{e} soit une base de E et \mathbf{f} une base de F . Considérons la matrice (en général rectangulaire) $A = \text{mat}(a, \mathbf{e}, \mathbf{f})$ de l'application linéaire a par rapport à ces deux bases. La matrice $B = \text{mat}({}^t a, \mathbf{f}^*, \mathbf{e}^*)$ est alors égale à la matrice transposée de A . En effet, la colonne i de la matrice B contient les coordonnées de la forme linéaire ${}^t a(f_i^*)$ dans la base \mathbf{e}^* . On a vu que la coordonnée j d'une forme linéaire x^* dans cette base \mathbf{e}^* est égale à $x^*(e_j)$, donc en appliquant ceci à $x^* = {}^t a(f_i^*)$ on obtient

$$B_{j,i} = {}^t a(f_i^*)(e_j) = f_i^*(a(e_j)) = A_{i,j}.$$

Transposition d'un changement de base : si V est la matrice de changement de base de \mathbf{e} vers \mathbf{f} , la matrice transposée ${}^t V$ est la matrice de changement de \mathbf{f}^* vers \mathbf{e}^* (attention au changement de l'ordre des bases !).

Bidual d'un espace vectoriel

À tout vecteur $x \in E$, on peut associer une fonction scalaire $j_E(x)$, définie sur l'espace dual E^* au moyen de la formule

$$\forall x^* \in E^*, (j_E(x))(x^*) = x^*(x) \in \mathbb{K}.$$

Il est facile de vérifier que $j_E(x)$ est une fonction linéaire de E^* dans \mathbb{K} , c'est-à-dire un élément du dual $(E^*)^*$ de E^* , qu'on appelle le *bidual* de E , et que l'on note E^{**} . De plus, l'application $j_E : x \rightarrow j_E(x)$ est linéaire de E dans E^{**} .

Théorème 4.5.1. *Si E est de dimension finie, l'application j_E est un isomorphisme de E sur son bidual E^{**} .*

Le caractère « naturel » de la définition justifie que l'on appelle j_E l'isomorphisme *canonique* de E sur E^{**} (la définition de j_E ne dépend que de E et de la définition du dual E^* , et ne dépend d'aucun choix auxiliaire).

Démonstration. On sait déjà que $\dim E^{**} = \dim E^* = \dim E$ d'après la proposition 4.5.1. Pour savoir que j_E est un isomorphisme, il suffit donc de vérifier que j_E est injective, c'est-à-dire de voir que pour tout vecteur $x \in E$ non nul, l'image $j_E(x)$ est non nulle. Mais d'après la proposition 4.5.3, si $x \in E$ est non nul il existe une forme linéaire x_0^* telle que $x_0^*(x) \neq 0$, donc $(j_E(x))(x_0^*) = x_0^*(x) \neq 0$, ce qui montre que la fonction $j_E(x)$ n'est pas identiquement nulle sur E^* , ce qui signifie que $j_E(x) \neq 0_{E^{**}}$.

Remarque. Retrouvons le fait que toute base $\mathbf{b} = (x_1^*, \dots, x_n^*)$ de E^* est la base duale d'une base \mathbf{e} de E . Considérons la base duale $(x_1^{**}, \dots, x_n^{**})$ de la base \mathbf{b} ; c'est une base du bidual E^{**} . Pour chaque $i = 1, \dots, n$ il existe un vecteur $x_i \in E$ tel que $x_i^{**} = j_E(x_i)$. On a alors pour tous $i, j = 1, \dots, n$

$$x_i^*(x_j) = x_j^{**}(x_i^*) = \delta_{i,j},$$

donc la base \mathbf{b} est la base duale de la base (x_1, \dots, x_n) de E .

Chapitre 5. Réduction des endomorphismes d'un espace vectoriel de dimension finie sur un sous-corps de \mathbb{C}

Dans ce chapitre, E est un espace vectoriel sur \mathbb{K} , où \mathbb{K} sera le plus souvent un sous-corps de \mathbb{C} (par exemple : \mathbb{Q} , \mathbb{R} , \mathbb{C}). Considérons deux exemples. Soit u l'application linéaire de \mathbb{R}^2 dans \mathbb{R}^2 définie par $u(x_1, x_2) = (2x_1 + x_2, x_1 + 2x_2)$; on voit que dans la base (f_1, f_2) de \mathbb{R}^2 définie par $f_1 = (1, 1)$, $f_2 = (1, -1)$, la matrice de u est diagonale. En effet on a $u(f_1) = 3f_1$ et $u(f_2) = f_2$. On dit que f_1 et f_2 sont des *vecteurs propres* de u , et que les scalaires 3 et 1 sont des *valeurs propres* correspondant à ces vecteurs propres. Une des questions fondamentales de ce chapitre est d'étudier la possibilité de trouver une base de E formée de vecteurs propres d'un endomorphisme u donné. On dira dans ce cas que u est *diagonalisable*.

Si on définit l'endomorphisme v de \mathbb{R}^2 par la formule $v(x_1, x_2) = (x_2, 0)$, on voit que la seule valeur propre possible pour v est 0, et les seuls vecteurs propres possibles sont de la forme $(x_1, 0)$, donc on ne peut pas trouver une base de \mathbb{R}^2 formée de vecteurs propres de v . Il existe donc des endomorphismes non diagonalisables, et nous donnerons des critères de diagonalisabilité.

L'une des motivations pour l'étude de la diagonalisation est la suivante : dans un certain nombre de situations, on étudie un système qui évolue dans le temps par étapes successives, et on décrit son état au temps n par un vecteur $X_n \in \mathbb{R}^d$; dans certains modèles simples (les chaînes de Markov par exemple), l'état X_{n+1} est obtenu en multipliant X_n par une matrice carrée M de taille $d \times d$, c'est-à-dire que $X_{n+1} = MX_n$, donc $X_n = M^n X_0$. Pour pouvoir prévoir l'évolution du système lorsque le temps n devient grand, on est donc conduit à étudier la suite des matrices M^n . Cette étude est évidente si M est diagonale, et encore assez facile quand M est diagonalisable. On rencontre une situation analogue lorsqu'on veut résoudre les *systèmes différentiels linéaires*, qui sont des systèmes d'équations différentielles linéaires que l'on peut écrire sous la forme matricielle $X'(t) = AX(t)$, où $X(t)$ est un vecteur dépendant du temps $t \in \mathbb{R}$ et A une matrice carrée de taille $d \times d$; si on sait diagonaliser A , le problème se décompose immédiatement en d équations d'une seule variable.

5.1. Polynôme caractéristique ; vecteurs propres et valeurs propres

Soit E un espace vectoriel sur \mathbb{K} ; rappelons qu'un *endomorphisme* de E est une application linéaire de E dans E , et que nous avons noté Id_E l'*endomorphisme identité* de E , qui est défini par $\text{Id}_E(x) = x$ pour tout $x \in E$. Le produit de deux endomorphismes est obtenu par la composition des applications, $uv = u \circ v$. On désigne par $\mathcal{L}(E)$ l'espace vectoriel des endomorphismes de E ; muni de l'opération de produit, c'est une *algèbre*.

Déterminant d'un endomorphisme

Soit E un espace vectoriel de dimension finie > 0 sur le corps \mathbb{K} ; on va montrer qu'on peut définir le déterminant d'un endomorphisme $u \in \mathcal{L}(E)$, indépendamment du choix d'une base de E .

Montrons d'abord une méthode « à la main ». Soit \mathbf{e} une base de E , et notons $A_{\mathbf{e}} = \text{mat}(u, \mathbf{e}, \mathbf{e}) = \text{mat}_{\mathbf{e}}(u)$ la matrice de u par rapport à la base \mathbf{e} ; considérons le déterminant $\det A_{\mathbf{e}}$ de cette matrice. Si on change de base, on aura $A_{\mathbf{f}} = V^{-1}A_{\mathbf{e}}V$, donc

$$\det A_{\mathbf{f}} = \det V^{-1} \det A_{\mathbf{e}} \det V = \det A_{\mathbf{e}},$$

ce qui montre que le déterminant de la matrice de u ne dépend pas de la base choisie, donc il serait raisonnable de dire que c'est le *déterminant de u* . Indiquons ensuite une méthode super-conceptuelle. Posons $n = \dim E$; l'espace vectoriel $\wedge^n E^*$ des formes n -linéaires alternées sur E est de dimension 1, et l'application qui associe à chaque $\varphi \in \wedge^n E^*$ la forme $T\varphi$ définie par

$$(T\varphi)(x_1, \dots, x_n) = \varphi(u(x_1), \dots, u(x_n))$$

est un endomorphisme de $\wedge^n E^*$, donc c'est une homothétie. Le coefficient de cette homothétie sera appelé $\det u$.

Pour finir, donnons la méthode que nous emploierons dans la suite de la section; c'est une variante adoucie de la précédente. Soit $u \in \mathcal{L}(E)$ et soit $\mathbf{e} = (e_1, \dots, e_n)$ une base de E ; la fonction

$$\varphi_{\mathbf{e}}(x_1, \dots, x_n) = \det_{\mathbf{e}}(u(x_1), \dots, u(x_n))$$

est une forme n -linéaire alternée sur E , donc elle est proportionnelle à la fonction $\det_{\mathbf{e}}$. Il existe donc $\lambda \in \mathbb{K}$ tel que $\varphi_{\mathbf{e}} = \lambda \det_{\mathbf{e}}$. Il est facile de vérifier que λ ne dépend que de u et pas de la base \mathbf{e} choisie: si on a une deuxième base \mathbf{f} de E , on sait qu'il existe $\mu \neq 0$ tel que $\det_{\mathbf{f}} = \mu \det_{\mathbf{e}}$, donc

$$\begin{aligned} \varphi_{\mathbf{f}}(x_1, \dots, x_n) &= \det_{\mathbf{f}}(u(x_1), \dots, u(x_n)) = \\ &= \mu \det_{\mathbf{e}}(u(x_1), \dots, u(x_n)) = \mu \lambda \det_{\mathbf{e}}(x_1, \dots, x_n) = \lambda \det_{\mathbf{f}}(x_1, \dots, x_n). \end{aligned}$$

On va appeler *déterminant de u* cette constante λ qui ne dépend que de u .

Définition 5.1.1. Soit E un espace vectoriel de dimension $n > 0$; le déterminant d'un endomorphisme $u \in \mathcal{L}(E)$ est donné par

$$\det u = \det_{\mathbf{e}}(u(e_1), \dots, u(e_n)),$$

où \mathbf{e} est une base quelconque de E . On a alors pour tout système (x_1, \dots, x_n) de n vecteurs de E

$$\det_{\mathbf{e}}(u(x_1), \dots, u(x_n)) = \det(u) \det_{\mathbf{e}}(x_1, \dots, x_n).$$

Proposition 5.1.1. *Le déterminant des endomorphismes vérifie les propriétés suivantes :*

1. $\det(\text{Id}_E) = 1$.
2. *Le déterminant d'un produit d'endomorphismes est égal au produit des déterminants,*

$$\det(u \circ v) = \det u \det v.$$

3. *L'endomorphisme u est inversible si et seulement si $\det(u) \neq 0$. Lorsque l'inverse de u existe, $\det(u^{-1}) = (\det(u))^{-1}$.*

4. *Si \mathbf{e} est une base de E , le déterminant d'un endomorphisme u de E est égal au déterminant de la matrice de u par rapport à la base \mathbf{e} .*

Démonstration. Commençons par la fin. On a remarqué que pour tout système $\mathbf{x} = (x_1, \dots, x_n)$ de vecteurs de E , le déterminant $\det_{\mathbf{e}}(x_1, \dots, x_n)$ est égal au déterminant

de la matrice M_x dont les colonnes sont les vecteurs colonne successifs des coordonnées de x_1, \dots, x_n par rapport à la base \mathbf{e} . On voit donc que $\det(u) = \det_{\mathbf{e}}(u(e_1), \dots, u(e_n))$ est égal au déterminant de la matrice de u dans la base \mathbf{e} . Revenons au point 1. Il est clair d'après la définition que $\det(\text{Id}_E) = 1$. Soit \mathbf{e} une base de E ; on aura

$$\det(u \circ v) = \det_{\mathbf{e}}(u(v(e_1)), \dots, u(v(e_n))) = \det u \det_{\mathbf{e}}(v(e_1), \dots, v(e_n)) = \det u \det v.$$

Passons au point trois. Si u est inversible, on aura $1 = \det(u u^{-1}) = \det(u) \det(u^{-1})$, ce qui prouve que $\det(u) \neq 0$. Réciproquement, si u n'est pas inversible, il existe un vecteur $x \neq 0$ dans E tel que $u(x) = 0$. On complète en une base (x, e_2, \dots, e_n) de E , et dans cette base la matrice M de u a une première colonne nulle, donc $\det u = \det M = 0$.

Interprétation physique dans \mathbb{R}^3 . Le déterminant du système de vecteurs (X_1, X_2, X_3) est égal au volume orienté du parallélépipède construit sur ces trois vecteurs. Si u est un endomorphisme de \mathbb{R}^3 , le déterminant de u est le facteur de transformation des volumes, quand on prend l'image par u .

Valeurs propres, vecteurs propres

Définitions 5.1.2. Soient E un espace vectoriel de dimension finie sur \mathbb{K} et $u \in \mathcal{L}(E)$; un vecteur $x \in E$ est *vecteur propre* de u si $x \neq 0_E$ et si $u(x) \in \mathbb{K}x$.

Si x est vecteur propre de u il existe donc $\lambda \in \mathbb{K}$ tel que $u(x) = \lambda x$. Un élément $\lambda \in \mathbb{K}$ est *valeur propre* de u s'il existe un vecteur $x \neq 0_E$ dans E tel que $ux = \lambda x$.

Exemple. L'application linéaire u de \mathbb{R}^2 dans \mathbb{R}^2 dont la matrice dans la base canonique est

$$\begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}$$

admet $(1, 0)$ pour vecteur propre, avec valeur propre $\lambda = 2$.

L'application linéaire de rotation d'angle $\pi/2$ dans \mathbb{R}^2 n'admet aucun vecteur propre.

Proposition 5.1.2. Soient E un espace vectoriel de dimension $n > 0$ sur \mathbb{K} et u un endomorphisme de E ; le nombre λ est valeur propre de u si et seulement si $\lambda \in \mathbb{K}$ et $\det(u - \lambda \text{Id}_E) = 0$. Si A est la matrice de u dans une base de E , ce qui précède équivaut à dire que λ est valeur propre de u si et seulement si $\lambda \in \mathbb{K}$ et $\det(A - \lambda I_n) = 0$.

Démonstration. Supposons que $\lambda \in \mathbb{K}$ soit valeur propre de u . Alors il existe un vecteur $x \neq 0_E$ tel que $(u - \lambda \text{Id}_E)(x) = 0_E$, donc $u - \lambda \text{Id}_E$ n'est pas injectif, donc pas inversible, ce qui implique $\det(u - \lambda \text{Id}_E) = 0$. Inversement si $\det(u - \lambda \text{Id}_E) = 0$, l'endomorphisme $u - \lambda \text{Id}_E$ n'est pas inversible donc pas injectif d'après la proposition 4.2.1, donc il existe $x \neq 0_E$ tel que $ux = \lambda x$.

Lemme. Soient φ une forme p -linéaire sur E et $y_1, \dots, y_p, z_1, \dots, z_p$ des vecteurs de E fixés; la fonction

$$\lambda \in \mathbb{K} \rightarrow \varphi(y_1 + \lambda z_1, \dots, y_p + \lambda z_p)$$

est une fonction polynomiale de la variable λ , de degré $\leq p$, à coefficients dans \mathbb{K} : il existe un polynôme $P \in \mathbb{K}[X]$ de degré $\leq p$, de la forme

$$P = \varphi(y_1, \dots, y_p) + c_1 X + \dots + c_{p-1} X^{p-1} + \varphi(z_1, \dots, z_p) X^p$$

et tel que $\varphi(y_1 + \lambda z_1, \dots, y_p + \lambda z_p) = P(\lambda)$ pour tout $\lambda \in \mathbb{K}$.

Démonstration. Par récurrence sur p ; on peut en fait démontrer la formule plus explicite

$$\varphi(y_1 + \lambda z_1, \dots, y_p + \lambda z_p) = \sum_{L \subset \{1, \dots, p\}} \lambda^{|L|} \varphi(w_1^L, \dots, w_p^L)$$

où $|L|$ désigne le cardinal du sous-ensemble L , $w_j^L = z_j$ si $j \in L$ et $w_j^L = y_j$ si $j \notin L$. Il y a 2^p termes dans la somme, correspondant aux 2^p sous-ensembles de l'ensemble $\{1, \dots, p\}$.

Si E est de dimension n , le déterminant de $u - \lambda \text{Id}_E$ est une fonction polynomiale de degré n en λ d'après le lemme précédent, puisque dans toute base \mathbf{e} de E on a

$$\det(u - \lambda \text{Id}_E) = \det_{\mathbf{e}}(u(e_1) - \lambda e_1, \dots, u(e_n) - \lambda e_n),$$

et il existe un polynôme $\chi_u \in \mathbb{K}[X]$, de la forme

$$\chi_u = \det(u) + c_1 X + \dots + c_{n-1} X^{n-1} + (-1)^n X^n$$

tel que $\det(u - \lambda \text{Id}_E) = \chi_u(\lambda)$ pour tout $\lambda \in \mathbb{K}$. C'est le *polynôme caractéristique* de u . On note que $\deg \chi_u = \dim E = n$.

Exercices.

1. Trace d'un endomorphisme. Montrer que le coefficient c_{n-1} est égal à

$$(-1)^{n-1} \sum_{i=1}^n a_{i,i},$$

si $A = (a_{i,j})$ est la matrice de u dans une base de E . En déduire que la quantité $\sum_{i=1}^n a_{i,i}$ ne dépend que de u . On l'appelle la *trace* de l'endomorphisme u (et aussi la trace de la matrice). On notera $\text{tr } A$, $\text{tr } u$ pour la trace de A et de u .

2. Montrer que le coefficient c_k du polynôme caractéristique peut s'exprimer à partir des mineurs d'ordre $n - k$ de la matrice de u dans une base \mathbf{e} .

Remarque. Si \mathbf{e} est une base de E et si $A = (a_{i,j})$ est la matrice de u dans cette base, on peut considérer la matrice B à coefficients polynômes dont les termes diagonaux sont les polynômes de degré un $a_{i,i} - X$ et dont les coefficients non-diagonaux sont les polynômes constants $a_{i,j}$. On peut écrire symboliquement $B = A - X I_n$ et calculer le déterminant $\det(A - X I_n)$ qui est un polynôme. On vérifiera comme au début de la section que ce déterminant-polynôme ne dépend pas de la base choisie, mais seulement de u : c'est le polynôme caractéristique χ_u .

Corollaire 5.1.1. Soient E un espace vectoriel de dimension finie > 0 sur \mathbb{K} et $u \in \mathcal{L}(E)$; les valeurs propres de u sont les racines de χ_u qui sont dans \mathbb{K} .

Exemple. Prenons $\mathbb{K} = \mathbb{R}$. L'endomorphisme r_θ de rotation d'angle θ dans \mathbb{R}^2 a pour matrice

$$R_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

dans la base canonique. Le polynôme caractéristique est $\chi = X^2 - 2X \cos \theta + 1$. Lorsque $\sin \theta \neq 0$, ce trinôme n'a pas de racine réelle, donc l'endomorphisme r_θ n'a aucune valeur propre et aucun vecteur propre.

En particulier, lorsque $\theta = \pi/2$, l'endomorphisme $r_{\pi/2}$ de \mathbb{R}^2 n'a pas de valeur propre. En revanche, l'endomorphisme de \mathbb{C}^2 défini par $u(z_1, z_2) = (-z_2, z_1)$ admet les vecteurs $(1, i)$ et $(1, -i)$ comme vecteurs propres, avec $-i$ et i comme valeurs propres. Dans ces deux exemples la matrice (dans la base canonique de \mathbb{R}^2 ou \mathbb{C}^2) est la même, la différence vient du changement de corps de base.

Polynôme caractéristique d'une matrice

Si A est une matrice $n \times n$ à coefficients dans \mathbb{K} , on désigne par χ_A le polynôme caractéristique de l'endomorphisme de \mathbb{K}^n défini par $Y \in \mathbb{K}^n \rightarrow AY \in \mathbb{K}^n$. On a pour tout $\lambda \in \mathbb{K}$

$$\chi_A(\lambda) = \det(A - \lambda I_n).$$

Si E est un espace vectoriel de dimension n sur \mathbb{K} , si on choisit une base de E et si A est la matrice d'un endomorphisme u de E dans cette base, on aura $\chi_u = \chi_A$.

Exemples de polynômes caractéristiques

1. Cas d'une matrice triangulaire $A = (a_{i,j})$: on voit directement dans ce cas que $\chi_A = \prod_{i=1}^n (a_{i,i} - X)$ puisque le déterminant caractéristique est triangulaire. Les valeurs propres de A sont donc les coefficients diagonaux de la matrice triangulaire A .

2. Calcul par blocs ; le polynôme caractéristique d'une matrice triangulaire par blocs est le produit des polynômes caractéristiques des blocs diagonaux.

3. Étant donné un polynôme $P = a_0 + a_1X + \dots + a_{n-1}X^{n-1} + X^n$ de degré $n \geq 1$ à coefficients dans \mathbb{K} , on lui associe une matrice $n \times n$ appelée *matrice compagnon* de P ,

$$M(P) = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & -a_{n-2} \\ 0 & \dots & 0 & 1 & -a_{n-1} \end{pmatrix}.$$

Le calcul du polynôme caractéristique de la matrice $M(P)$ se fait par récurrence sur le degré n de P , en développant par rapport à la première ligne. On montre que

$$\chi_{M(P)} = (-1)^n P.$$

Définition 5.1.3. Soit u un endomorphisme de E ; on dit qu'un sous-espace vectoriel F de E est *stable par u* si $u(F) \subset F$, c'est-à-dire

$$\forall x \in F, \quad u(x) \in F.$$

Exemples. Les sous-espaces $\{0\}$ et E sont toujours stables. Si $u \in \mathcal{L}(E)$, les deux sous-espaces $\ker u$ et $u(E)$ sont stables par u . Si x est un vecteur propre de u , la droite $\mathbb{K}x$ est stable par u .

Remarques.

1. Si F est engendré par (y_1, \dots, y_q) , il suffit que $u(y_j) \in F$ pour tout $j = 1, \dots, q$ pour que F soit stable par u .

2. Si F est stable par u , si $\mathbf{f} = (f_1, \dots, f_p)$ est une base de F et si

$$\mathbf{b} = (f_1, \dots, f_p, x_{p+1}, \dots, x_n)$$

est une base de E obtenue en complétant la base \mathbf{f} de F , la matrice de u dans la base \mathbf{b} est triangulaire par blocs pour la décomposition $E = [f_1, \dots, f_p] \oplus [x_{p+1}, \dots, x_n]$.

Proposition 5.1.3. Soient $u, v \in \mathcal{L}(E)$; si $u \circ v = v \circ u$ alors $\ker v$ et $v(E)$ sont stables par u .

Démonstration. Supposons que $v(x) = 0_E$. On voit que $v(u(x)) = u(v(x)) = 0_E$, donc $u(x) \in \ker v$, ce qui montre que $\ker v$ est stable par u . Si $y = v(x)$ appartient à l'image de v , on aura $u(y) = u(v(x)) = v(u(x)) \in v(E)$, donc l'image $v(E)$ est stable par u .

Si un sous-espace $F \subset E$ est stable par $u \in \mathcal{L}(E)$, on peut définir un endomorphisme v de F en posant $v(y) = u(y) \in F$ pour tout $y \in F$ (on aurait pu aussi définir une application linéaire w de F dans E en posant $w(y) = u(y) \in E$ pour tout $y \in F$! cela peut sembler du pinaillage, mais il est important de comprendre que ces trois objets sont différents, parce que l'ensemble de définition ou bien l'ensemble d'arrivée n'est pas le même. Dans ce chapitre on ne se servira pas de w , mais de v).

Définition 5.1.4. Soient u un endomorphisme de E et F un sous-espace vectoriel de E stable par u ; l'endomorphisme restriction de u au sous-espace F est l'endomorphisme $v = u|_F : F \rightarrow F$ défini par

$$\forall y \in F, \quad v(y) = u(y) \in F.$$

Exemples. La restriction de Id_E à F est Id_F . Si F est stable par u , il est stable par toutes les puissances u^n et la restriction de u^n à F est la puissance v^n de la restriction v de u . Si F est stable par u_1 et u_2 , il est stable par $\lambda u_1 + u_2$ et la restriction de $\lambda u_1 + u_2$ est $\lambda v_1 + v_2$; par exemple, la restriction de $u - \lambda \text{Id}_E$ est $v - \lambda \text{Id}_F$.

Si on aime bien les grands mots, on peut traduire le paragraphe précédent en disant que l'ensemble $\mathcal{L}_F(E)$ des endomorphismes u de E qui laissent stable un sous-espace donné F est une *sous-algèbre* de $\mathcal{L}(E)$, c'est-à-dire un sous-espace vectoriel de $\mathcal{L}(E)$, contenant l'identité et tel que $u_1 u_2 \in \mathcal{L}_F(E)$ chaque fois que $u_1, u_2 \in \mathcal{L}_F(E)$, et en disant ensuite que l'application de restriction $u \rightarrow u|_F$ est un *homomorphisme d'algèbre* de $\mathcal{L}_F(E)$ dans $\mathcal{L}(F)$.

Polynôme caractéristique de la restriction à un sous-espace stable

Supposons que $u \in \mathcal{L}(E)$ admette un sous-espace stable F , et désignons par v la restriction de u à F . On a dit qu'en formant une base de E en prenant d'abord une base \mathbf{f} de F , suivie d'une base d'un supplémentaire quelconque G de F dans E , on obtient pour u une matrice M triangulaire par blocs pour la décomposition $E = F \oplus G$, dont le coin « nord-ouest » est la matrice A de v dans la base \mathbf{f} . Cela implique d'après la proposition 4.4.2 que le polynôme caractéristique de v divise χ_u . Dans le cas où l'espace est somme directe de deux sous-espaces stables F_1 et F_2 , on choisit une base de E formée d'une base \mathbf{f}_1 de F_1 suivie d'une base \mathbf{f}_2 de F_2 . Dans cette base de E , la matrice M est diagonale par blocs, avec deux blocs diagonaux A_1 et A_2 qui sont les matrices des restrictions v_1 et v_2 de u par rapport aux bases \mathbf{f}_1 et \mathbf{f}_2 . Il en résulte que le polynôme caractéristique de u est le produit des polynômes caractéristiques des restrictions v_1 et v_2 de u aux sous-espaces F_1 et F_2 .

Proposition 5.1.4. Si F est un sous-espace stable pour un endomorphisme $u \in \mathcal{L}(E)$ et si v désigne la restriction de u à F , le polynôme χ_v divise χ_u . Si $E = F_1 \oplus F_2$, où F_1 et F_2 sont stables par u , on a

$$\chi_u = \chi_{v_1} \chi_{v_2}$$

où v_1 et v_2 désignent les restrictions de u à F_1 et F_2 .

Valeurs propres, vecteurs propres : le cas complexe

Lorsque E est un espace vectoriel de dimension n sur \mathbb{C} , toutes les racines du polynôme caractéristique sont automatiquement dans $\mathbb{K} = \mathbb{C}$, donc toutes les racines de χ_u sont des valeurs propres de u dans ce cas. De plus, on sait d'après le théorème de d'Alembert que tout polynôme de degré ≥ 1 à coefficients complexes a au moins une racine complexe (et en fait, exactement n quand on les compte avec leur ordre de multiplicité) donc tout endomorphisme de E admet au moins un vecteur propre.

Donc : sur \mathbb{C} , tout endomorphisme a une droite stable.

Exemple des matrices de Jordan. Il est possible qu'il n'existe qu'une seule direction propre pour une matrice donnée. C'est le cas pour une matrice de la forme

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

dont le polynôme caractéristique est $\chi = X^4$, la seule valeur propre est $\lambda = 0$ et les seuls vecteurs propres sont de la forme $(a, 0, 0, 0)$.

Cas réel en dimension impaire

On sait que tout polynôme à coefficients réels de degré impair admet une racine réelle (théorème des valeurs intermédiaires). Un endomorphisme d'un espace réel de dimension impaire a donc toujours au moins un vecteur propre.

Exemple. Une isométrie u de \mathbb{R}^3 admet toujours un vecteur $x \neq 0$ tel que $u(x) = \pm x$.

Valeurs propres d'une matrice à coefficients dans un sous-corps \mathbb{K} de \mathbb{C}

Une matrice A de taille $n \times n$ à coefficients dans \mathbb{K} est en particulier une matrice à coefficients dans \mathbb{C} et définit donc un endomorphisme $A^{\mathbb{C}}$ de \mathbb{C}^n auquel la discussion du paragraphe précédent s'applique. L'endomorphisme $A^{\mathbb{C}}$ est défini par le calcul matriciel,

$$\forall Z \in \mathbb{C}^n, \quad A^{\mathbb{C}}(Z) = AZ.$$

Exemple. Matrice R_θ de la rotation d'angle θ : pour cette matrice on a le polynôme caractéristique $\chi = X^2 - 2(\cos \theta)X + 1$, qui admet les deux racines complexes (conjuguées) $\lambda = \cos \theta \pm i \sin \theta$. On trouve une base de \mathbb{C}^2 formée de deux vecteurs propres de l'endomorphisme $R_\theta^{\mathbb{C}}$, les vecteurs $(1, i)$ et $(1, -i)$ de \mathbb{C}^2 . Il faut donc faire bien attention au changement de corps. La rotation r_θ d'angle θ (avec $\sin \theta \neq 0$) dans le plan \mathbb{R}^2 n'a pas de vecteur propre, mais sa matrice a des valeurs propres complexes, qui donnent des vecteurs propres pour l'endomorphisme $R_\theta^{\mathbb{C}}$ de \mathbb{C}^2 qui a la même matrice.

D'après le théorème de d'Alembert, on peut factoriser dans $\mathbb{C}[X]$ le polynôme caractéristique χ_A en facteurs de degré 1,

$$\chi_A = (\lambda_1 - X) \dots (\lambda_n - X),$$

où chaque λ_i est une des valeurs propres de A . Chaque valeur propre peut apparaître plusieurs fois, c'est-à-dire que certaines valeurs propres peuvent être racines multiples du polynôme caractéristique. Si on compte les valeurs propres avec leur ordre de multiplicité comme racines de χ_A , on trouve par identification que $\det A$ est égal au produit des valeurs propres.

Proposition 5.1.5. *Soit E un espace vectoriel sur \mathbb{R} , de dimension finie > 0 ; tout endomorphisme u de l'espace E admet une droite ou un plan stable.*

Démonstration. En passant à la matrice (réelle) A de u dans une base $\mathbf{e} = (e_1, \dots, e_n)$ de E on trouve une valeur propre complexe $\lambda = a + ib$ de la matrice, $a, b \in \mathbb{R}$. Si $b = 0$ (c'est-à-dire que $\lambda \in \mathbb{R}$), λ est valeur propre de u , et il lui correspond une droite vectorielle stable par u . Sinon, on peut trouver un vecteur propre $Z = X + iY \in \mathbb{C}^n$ de la matrice A , où $X, Y \in \mathbb{R}^n$. On voit que X et Y sont \mathbb{R} -indépendants (sinon, si $Y = \mu X$, μ réel,

on aurait $(1 + i\mu)AX = (1 + i\mu)(a + ib)X$, donc $AX = aX + ibX$, ce qui est impossible puisque AX est réel et $b \neq 0, X \neq 0$, donc X et Y engendrent un plan passant par 0 . La relation

$$A(X + iY) = (a + ib)(X + iY) = (aX - bY) + i(bX + aY)$$

montre par identification que $AX = aX - bY$ et $AY = bX + aY$. Si on écrit $X = (c_1, \dots, c_n) \in \mathbb{R}^n$ et $Y = (d_1, \dots, d_n) \in \mathbb{R}^n$ et si on pose

$$x = \sum_{i=1}^n c_i e_i \in E; \quad y = \sum_{i=1}^n d_i e_i \in E,$$

on aura aussi $u(x) = ax - by$ et $u(y) = bx + ay$, donc le plan $[x, y]$ est stable par u .

Exemple. Réduction sur \mathbb{R} de la matrice

$$\begin{pmatrix} 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Le polynôme caractéristique est $X^4 + 1$, on trouve quatre racines complexes $\lambda = \gamma(\pm 1 \pm i)$ où $\gamma = \sqrt{2}/2$. Pour chaque racine λ le vecteur $(\lambda^3, \lambda^2, \lambda, 1) \in \mathbb{C}^4$ est vecteur propre et on obtient un plan stable en séparant partie réelle et imaginaire.

5.2. Sous-espaces propres d'un endomorphisme

Si $u \in \mathcal{L}(E)$ et si $\mu \in \mathbb{K}$ est une valeur propre de u , le sous-espace vectoriel

$$E_\mu = \ker(u - \mu \text{Id}_E) = \{x \in E : ux = \mu x\}$$

est différent de $\{0_E\}$. Tout vecteur $x \neq 0_E$ de E_μ est un vecteur propre de u de valeur propre μ . On appelle E_μ le *sous-espace propre* de u associé à la valeur propre μ . Il est clair que E_μ est un sous-espace stable pour u . Si $v \in \mathcal{L}(E)$ commute avec u , il commute avec $u - \lambda \text{Id}_E$ donc les sous-espaces propres de u sont stables par v d'après la proposition 5.1.3.

Proposition 5.2.1. *Les sous-espaces propres d'un endomorphisme u forment une somme directe.*

Démonstration. On démontre cette propriété par récurrence sur le nombre de sous-espaces propres considérés. Supposons donc que toute famille de $k - 1$ sous-espaces propres forme une somme directe, et montrons que cela reste vrai pour k . Soient $\lambda_1, \dots, \lambda_k$ des valeurs propres de u , deux à deux distinctes, et soient E_1, \dots, E_k les sous-espaces propres correspondants; supposons que $x_j \in E_j$ pour $j = 1, \dots, k$ et que

$$x_1 + \dots + x_k = 0_E.$$

Nous devons montrer que $x_1 = x_2 = \dots = x_k = 0_E$. On déduit en appliquant u , puisque $u(x_j) = \lambda_j x_j$ pour $j = 1, \dots, k$

$$\lambda_1 x_1 + \dots + \lambda_k x_k = 0_E,$$

donc en retranchant $\lambda_k(x_1 + \dots + x_k) = 0_E$ on obtient

$$(\lambda_1 - \lambda_k)x_1 + \dots + (\lambda_{k-1} - \lambda_k)x_{k-1} = 0_E,$$

relation qui fait intervenir seulement $k - 1$ des sous-espaces propres, ce qui entraîne que $(\lambda_j - \lambda_k)x_j = 0_E$ pour $j = 1, \dots, k - 1$ par l'hypothèse de récurrence, donc $x_1 = \dots = x_{k-1} = 0_E$ puisque $\lambda_j \neq \lambda_k$ pour $j < k$, donc $x_k = 0_E$ aussi.

Définition 5.2.1. Soit u un endomorphisme d'un espace vectoriel E de dimension finie sur \mathbb{K} ; on dit que u est *diagonalisable* s'il existe une base de E formée de vecteurs propres de u .

Proposition 5.2.2. Si $\dim E = n$ et si $u \in \mathcal{L}(E)$ admet n valeurs propres distinctes, u est diagonalisable.

Démonstration. Soient $\lambda_1, \dots, \lambda_n$ les valeurs propres (supposées distinctes) ; pour chaque $j = 1, \dots, n$ on peut trouver un vecteur propre f_j tel que $u(f_j) = \lambda_j f_j$. On va montrer que le système (f_1, \dots, f_n) est libre (donc ce sera une base de E puisque $n = \dim E$) : si on a une relation

$$c_1 f_1 + \dots + c_n f_n = 0_E$$

on aura en posant $x_j = c_j f_j$ pour tout $j = 1, \dots, n$, d'une part $x_j \in E_{\lambda_j}$ pour tout j et $x_1 + \dots + x_n = 0_E$. Puisque les sous-espaces propres forment une somme directe, il en résulte que $x_1 = \dots = x_n = 0_E$, donc $c_j f_j = 0_E$ pour tout $j = 1, \dots, n$, donc $c_j = 0$ puisque $f_j \neq 0_E$ par définition d'un vecteur propre, et on a montré que (f_1, \dots, f_n) est libre. On a donc trouvé une base de E formée de vecteurs propres de u .

Dans une base formée de vecteurs propres la matrice de u est diagonale, et les coefficients diagonaux sont les valeurs propres, qui apparaissent éventuellement plusieurs fois. Chaque valeur propre λ apparaît un nombre de fois égal à la dimension du sous-espace propre E_λ correspondant. Si u est diagonalisable, il existe donc une base de E dans laquelle la matrice A de u est diagonale, avec diagonale $\lambda_1, \dots, \lambda_n$ formée d'éléments de \mathbb{K} , et alors

$$\chi_u = \chi_A = \prod_{i=1}^n (\lambda_i - X)$$

montre que toutes les racines de χ_u sont dans \mathbb{K} .

Pour que u soit diagonalisable, il est nécessaire que toutes les racines de χ_u soient dans \mathbb{K} .

Dire que u est diagonalisable revient à dire que E est la somme des sous-espaces propres de u . En effet, si $\mathbf{f} = (f_1, f_2, \dots, f_n)$ est une base de E formée de vecteurs propres de u , et si μ_1, \dots, μ_p est la liste des valeurs propres (sans répétition), chaque vecteur de base f_j appartient à l'un des sous-espaces propres $E_{\mu_1}, \dots, E_{\mu_p}$, donc *a fortiori* f_j appartient au sous-espace $F = E_{\mu_1} + \dots + E_{\mu_p}$, donc $F = E$ puisque F contient tous les vecteurs de la base \mathbf{f} . Inversement, si $F = E$, on aura

$$E = E_{\mu_1} \oplus \dots \oplus E_{\mu_p},$$

et on obtiendra une base de E en prenant une base de chaque sous-espace E_{μ_j} , puis en rassemblant tous les vecteurs ainsi obtenus ; une telle base de E est formée de vecteurs propres de u .

Puisque la somme des sous-espaces propres est une somme directe, la dimension de $E_{\mu_1} \oplus \dots \oplus E_{\mu_p}$ est la somme des dimensions, donc

l'endomorphisme u est diagonalisable si et seulement si la somme des dimensions des sous-espaces propres est égale à $n = \dim E$.

Soit $\mu \in \mathbb{K}$ une valeur propre de u , et soit r l'ordre de multiplicité de μ comme racine de χ_u ; alors $\dim E_\mu \leq r$. En effet, si on écrit la matrice A de u dans une base de E qui commence par une base de E_μ , le bloc carré A_μ correspondant au sous-espace E_μ

est égal à μI_p , où p est la dimension de E_μ , et la matrice A est triangulaire par blocs, donc le polynôme $\chi_{A_\mu} = (\mu - X)^p$ divise le polynôme χ_u , donc $\dim E_\mu = p \leq r$.

Supposons que toutes les racines de χ_u soient dans \mathbb{K} , et soit $\mu_1, \dots, \mu_k \in \mathbb{K}$ la liste des racines (deux à deux distinctes) du polynôme χ_u ; soient r_1, \dots, r_k leurs multiplicités comme racines de χ_u ; on a $\dim E = n = r_1 + \dots + r_k$. On sait que $\dim E_{\mu_j} \leq r_j$, donc pour que u soit diagonalisable il faut que pour tout $j = 1, \dots, k$, on ait $\dim E_{\mu_j} = r_j$.

Pour que u soit diagonalisable, il faut et il suffit que toutes les racines de χ_u soient dans \mathbb{K} et que pour chaque racine μ de χ_u de multiplicité r on ait $\dim E_\mu = r$.

Puisque E_μ est le noyau de $u - \mu \text{Id}_E$, on sait que sa dimension est donnée par la relation $\dim E_\mu = \dim E - \dim(u - \mu \text{Id}_E)(E)$, c'est-à-dire $\dim E_\mu = n - \text{rang}(A - \mu I_n)$ si A est la matrice de u dans une base de E . Cette remarque fournit une méthode pratique pour appliquer le critère précédent.

Exercices.

1. Montrer que la matrice

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

n'est pas diagonalisable.

2. Montrer que la matrice compagnon $M(P)$ d'un polynôme P est diagonalisable sur \mathbb{C} si et seulement si le polynôme P a toutes ses racines simples.

Réduction réelle d'une matrice qui est diagonalisable sur \mathbb{C}

Soit A une matrice réelle, qui définit un endomorphisme de \mathbb{R}^n par $X \in \mathbb{R}^n \rightarrow AX \in \mathbb{R}^n$; supposons que l'endomorphisme $A^{\mathbb{C}}$ de \mathbb{C}^n défini par la même matrice soit diagonalisable. Pour chaque vecteur $Z \in \mathbb{C}^n$ ou matrice B à coefficients complexes on peut définir le vecteur complexe conjugué \bar{Z} et la matrice \bar{B} , et les règles habituelles pour le calcul des conjugués des sommes et des produits restent valables. Puisque A est réelle, l'équation $AZ = \mu Z$ équivaut à $A\bar{Z} = \bar{\mu}\bar{Z}$. Par conséquent, μ est valeur propre de $A^{\mathbb{C}}$ si et seulement si $\bar{\mu}$ est valeur propre, et la correspondance $Z \rightarrow \bar{Z}$ (qui n'est pas \mathbb{C} -linéaire) transforme les éléments de E_μ en éléments de $E_{\bar{\mu}}$. Soit $\mu = a + ib$, $a, b \in \mathbb{R}$ et $b \neq 0$ une valeur propre non réelle de $A^{\mathbb{C}}$; soit (Z_1, \dots, Z_p) une base de E_μ , et décomposons $Z_j = X_j + iY_j$, $X_j, Y_j \in \mathbb{R}^n$, pour tout $j = 1, \dots, p$; alors $(\bar{Z}_1, \dots, \bar{Z}_p)$ est une base de $E_{\bar{\mu}}$ et $((Z_1 + \bar{Z}_1), \dots, (Z_p + \bar{Z}_p), -i(Z_1 - \bar{Z}_1), \dots, -i(Z_p - \bar{Z}_p))$ est une base (sur \mathbb{C}) de $E_\mu \oplus E_{\bar{\mu}}$. Mais ces vecteurs sont les vecteurs réels $(2X_1, \dots, 2X_p, 2Y_1, \dots, 2Y_p)$, qui sont *a fortiori* \mathbb{R} -indépendants. De plus, comme on l'a déjà vu, on a pour chaque $j = 1, \dots, k$

$$AX_j = aX_j - bY_j; \quad AY_j = bX_j + aY_j.$$

Revenant au problème réel, on obtiendra pour la partie d'espace \mathbb{R}^n engendrée par le système libre $(X_1, Y_1, \dots, X_p, Y_p)$ une matrice diagonale par blocs, avec k blocs de taille 2×2 tous égaux à

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}.$$

Sous-espaces caractéristiques

Soient E un espace vectoriel de dimension finie $n > 0$ sur \mathbb{K} et $u \in \mathcal{L}(E)$; soit μ une valeur propre de u , racine d'ordre r du polynôme caractéristique,

$$\chi_u = (\mu - X)^r Q,$$

avec $Q(\mu) \neq 0$. Le sous-espace propre $E_\mu = \ker(u - \mu \text{Id}_E)$ vérifie $0 < \dim E_\mu \leq r$. Dans les bons cas, on a $\dim E_\mu = r$. Que faire dans les cas moins bons où $\dim E_\mu < r$? (rappelons qu'on est déjà sûr que u n'est pas diagonalisable dans ces cas moins bons).

Posons $v = u - \mu \text{Id}_E$. On a

$$\{0\} \subset \ker v \subset \ker v^2 \subset \dots \subset \ker v^q \dots$$

La suite croissante des entiers $d_q = \dim \ker v^q$ est majorée par $n = \dim E$, donc elle finit par être constante : il existe un entier q_0 tel que $d_q = d_{q_0}$ pour tout $q \geq q_0$, ce qui entraîne $\ker v^q = \ker v^{q_0}$ pour tout $q \geq q_0$. On appelle *sous-espace caractéristique de u pour la valeur propre μ* ce « noyau limite » $F_\mu = \ker(u - \mu \text{Id}_E)^q$ lorsque $q \geq q_0$. Il contient le sous-espace propre E_μ . Dans le cas où u est diagonalisable, on vérifie que ces deux sous-espaces E_μ et F_μ sont égaux (petit exercice).

On peut en fait déterminer l'entier q_0 de la discussion précédente très simplement : désignons par p le plus petit entier $q \geq 0$ vérifiant l'égalité $\ker v^q = \ker v^{q+1}$; on a $p > 0$ car $E_\mu = \ker v \neq \{0\} = \ker v^0$. On vérifie alors que $\ker v^q = \ker v^p$ pour tout entier $q \geq p$ par récurrence sur $q \geq p$ (exercice), donc $\ker v^p$ est déjà égal au sous-espace caractéristique F_μ (on verra plus loin que p est inférieur ou égal à l'ordre de multiplicité r de la racine μ , donc on a aussi $F_\mu = \ker v^r = \ker(u - \mu \text{Id}_E)^r$, ce qui donne une formule plus directe); on va montrer que

$$E = \ker v^p \oplus v^p(E).$$

Tout d'abord, $\ker v^p \cap v^p(E) = \{0\}$; en effet, si $x \in \ker v^p \cap v^p(E)$, on a $v^p(x) = 0$ et on peut écrire $x = v^p(y)$ pour un certain $y \in E$; alors $0 = v^p(x) = v^{2p}(y)$, donc $y \in \ker v^{2p} = \ker v^p$ (puisque $2p \geq p$), ce qui donne $v^p(y) = x = 0$. Ensuite, la relation $\dim E = \dim \ker v^p + \dim v^p(E)$ finit de montrer que $E = \ker v^p \oplus v^p(E)$; cette décomposition est formée de deux sous-espaces stables par v , donc par u .

Théorème 5.2.1. *On suppose que $\mu \in \mathbb{K}$ est valeur propre de $u \in \mathcal{L}(E)$ et racine d'ordre r du polynôme caractéristique χ_u . Le sous-espace caractéristique F_μ de u correspondant à la valeur propre μ est égal à*

$$F_\mu = \ker(u - \mu \text{Id}_E)^r \quad \text{et on a} \quad \dim F_\mu = r.$$

Si on pose $v = u - \mu \text{Id}_E$, $G_1 = F_\mu = \ker v^r$ et $G_2 = v^r(E)$, on a $E = G_1 \oplus G_2$, et G_1, G_2 sont stables par u ; si on désigne par u_1 et u_2 les deux endomorphismes restriction de u à G_1 et à G_2 , on a $\chi_u = \chi_{u_1} \chi_{u_2}$, avec $\chi_{u_1} = (\mu - X)^r$ et $\chi_{u_2}(\mu) \neq 0$.

Démonstration. On pose avec les notations établies avant l'énoncé

$$G_1 = \ker v^p; \quad G_2 = v^p(E).$$

On a vu que $E = G_1 \oplus G_2$, et cette décomposition est formée de deux sous-espaces stables par u . On a *a fortiori* $\ker v \cap v^p(E) = \{0\}$, donc la restriction $v_2 = u_2 - \mu \text{Id}_{G_2}$ de v au sous-espace G_2 est injective, ce qui signifie que la restriction u_2 de u à G_2 n'admet plus la valeur

propre μ , ou encore que $\chi_{u_2}(\mu) \neq 0$. On sait que $\chi_u = \chi_{u_1} \chi_{u_2}$. Puisque $(X - \mu)^r$ divise χ_u et $\chi_{u_2}(\mu) \neq 0$, on déduit que $(X - \mu)^r$ divise χ_{u_1} . Il en résulte que $r \leq \deg \chi_{u_1} = \dim G_1$.

On va montrer maintenant que μ est la seule racine complexe de χ_{u_1} , ce qui entraînera que $\chi_{u_1} = (\mu - X)^r$ et $\dim G_1 = r$. Pour tout $x \in G_1$, on a $v^p(x) = 0$ par définition de G_1 , ce qui signifie que $v_1^p = 0$. Si on choisit une base \mathbf{f} de G_1 et si A est la matrice de u_1 dans cette base, la matrice de v_1 sera $A - \mu I_d$ (en posant $d = \dim G_1$) et on aura $(A - \mu I_d)^p = 0$. Si λ est une valeur propre complexe de A et $Z \in \mathbb{C}^d$ un vecteur non nul tel que $AZ = \lambda Z$, on aura $(A - \mu I_d)Z = (\lambda - \mu)Z$ et $0 = (A - \mu I_d)^p Z = (\lambda - \mu)^p Z$, donc $\lambda - \mu = 0$.

On sait maintenant que $\chi_{u_1} = (\mu - X)^r$ et $\dim G_1 = r$. Pour finir, on remarque que $0 < \dim \ker v < \dim \ker v^2 < \dots < \dim \ker v^p = \dim G_1$, ce qui montre que $p \leq \dim G_1 = r$. On sait alors que $G_1 = \ker v^p = \ker v^r$. Enfin, $v^r(E) = v^p(E)$ (on a $v^r(E) \subset v^p(E)$ et ils ont la même dimension $n - \dim \ker v^r = n - \dim \ker v^p$).

Remarque. Le lecteur aura peut-être remarqué que nous avons renoncé à notre manie de toujours écrire 0_E pour le vecteur nul de E , manie qui finissait par être insupportable. Ca nous reprendra encore de temps en temps, mais rarement...

Dans le cas d'un endomorphisme $u \in \mathcal{L}(E)$ qui n'est pas diagonalisable, mais tel que toutes les racines de χ_u soient dans \mathbb{K} , on obtient une décomposition intéressante en sous-espaces stables par u , plus grands que les sous-espaces propres. En effet, si toutes les racines de χ_u sont dans \mathbb{K} , on peut continuer la décomposition en sous-espaces caractéristiques que nous avons commencée avec la valeur propre μ : on a pour l'instant

$$E = \ker(u - \mu \text{Id}_E)^r \oplus G,$$

et on continue la décomposition de E en somme directe en raisonnant sur la restriction de u à G , qui n'admet plus la valeur propre μ .

Théorème 5.2.2. *Si toutes les racines de χ_u sont dans \mathbb{K} , l'espace E est somme directe des sous-espaces caractéristiques,*

$$E = F_{\mu_1} \oplus \dots \oplus F_{\mu_k},$$

où μ_1, \dots, μ_k est la liste des racines de χ_u (sans répétition).

Triangulation

Définition 5.2.2. On dit que $u \in \mathcal{L}(E)$ est *triangulable* s'il existe une base \mathbf{f} de E dans laquelle la matrice $A = \text{mat}_{\mathbf{f}}(u)$ de u est triangulaire supérieure.

Si u est triangulable, on voit immédiatement que les coefficients diagonaux de la matrice triangulaire A qui représente u dans la base \mathbf{f} sont les racines de χ_u , et ces coefficients sont dans \mathbb{K} . Pour pouvoir trianguler u , il est donc nécessaire que toutes les racines de χ_u soient dans \mathbb{K} .

Proposition 5.2.3. *Soit E un espace vectoriel de dimension finie sur \mathbb{K} ; l'endomorphisme $u \in \mathcal{L}(E)$ est triangulable si et seulement si le polynôme caractéristique χ_u se décompose en facteurs de degré 1 dans $\mathbb{K}[X]$ (en d'autres termes si et seulement si χ_u a toutes ses racines dans \mathbb{K}).*

Démonstration. Si u est triangulable, le polynôme caractéristique a toutes ses racines dans \mathbb{K} . On montre le résultat inverse par récurrence sur la dimension de E . C'est trivial quand $n = \dim E = 1$. Supposons le résultat vrai pour les espaces de dimension $< n$, et soit $u \in \mathcal{L}(E)$ un endomorphisme d'un espace E de dimension n , dont le polynôme caractéristique ait toutes ses racines dans \mathbb{K} . Soit $\mu \in \mathbb{K}$ l'une de ces racines; soit x un

vecteur propre tel que $u(x) = \mu x$ et soit F un supplémentaire de $\mathbb{K}x$ dans E ; désignons par p la projection de $E = \mathbb{K}x \oplus F$ sur F et soit v l'endomorphisme de F défini par

$$\forall y \in F, \quad v(y) = p(u(y)).$$

En écrivant la matrice A de u dans une base de E formée de x suivi d'une base de F , on voit que la matrice A est triangulaire par blocs, avec un bloc diagonal de taille 1×1 égal à (μ) , et un autre bloc diagonal D de taille $(n-1) \times (n-1)$, et D est la matrice de v dans la base choisie pour F . On a alors $\chi_A = (\mu - X)\chi_D$, donc χ_D a toutes ses racines dans \mathbb{K} , puisque toutes les racines de χ_D sont racines de χ_A . D'après l'hypothèse de récurrence, il existe une base (f_1, \dots, f_{n-1}) de F dans laquelle la matrice de v est triangulaire supérieure. On vérifie que dans la base (x, f_1, \dots, f_{n-1}) de E , la matrice de u est triangulaire supérieure.

Corollaire 5.2.1. *Sur \mathbb{C} , tout endomorphisme est triangulable.*

Triangulation d'une matrice

Définition 5.2.3. On dit que $A \in M_n(\mathbb{R})$ est *triangulable sur \mathbb{R}* s'il existe une base \mathbf{f} de \mathbb{R}^n dans laquelle la nouvelle matrice $B = A_{\mathbf{f}} = V^{-1}AV$ obtenue par changement de base soit triangulaire supérieure. Autrement dit, s'il existe une matrice **réelle** V inversible telle que $V^{-1}AV$ soit triangulaire. On dira que A , matrice carrée complexe, est triangulable sur \mathbb{C} s'il existe une matrice **complexe** V inversible telle que $V^{-1}AV$ soit triangulaire supérieure.

Corollaire 5.2.2. *Sur \mathbb{C} , toute matrice carrée est triangulable.*

Proposition 5.2.4. *Une matrice $A \in M_n(\mathbb{R})$ est triangulable sur \mathbb{R} si et seulement si toutes les racines du polynôme caractéristique χ_A sont réelles.*

Combinaison du théorème des sous-espaces caractéristiques et de la triangulation

Examinons le cas particulier où le polynôme caractéristique de la matrice A est $\chi_A = (\mu - X)^n$, avec $\mu \in \mathbb{K}$. Dans ce cas la seule valeur propre est μ , donc la forme triangulaire sera

$$T = \begin{pmatrix} \mu & * & \dots & * \\ 0 & \mu & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & \mu \end{pmatrix},$$

où les étoiles représentent des coefficients inconnus. On peut écrire $T = \mu I_n + N$, où la matrice N est *strictement triangulaire*,

$$N = \begin{pmatrix} 0 & * & \dots & * \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & 0 \end{pmatrix}.$$

On montre par le calcul que $N^n = 0$, c'est-à-dire que N est *nilpotente* (on pourra aussi le voir par le théorème de Cayley-Hamilton 5.3.1, sachant que $\chi_n = (-1)^n X^n$). Puisque μI_n commute avec N , on a obtenu le résultat suivant : si $\chi_A = (\mu - X)^n$, avec $\mu \in \mathbb{K}$, il

existe trois matrices V, D, N à coefficients dans \mathbb{K} telles que V est inversible, D diagonale et N strictement triangulaire, et

$$A = V(D + N)V^{-1}, \quad DN = ND.$$

Soit $A \in M_n(\mathbb{K})$ telle que toutes les racines de χ_A soient dans \mathbb{K} (ce qui est automatique si $\mathbb{K} = \mathbb{C}$) ; soit μ_1, \dots, μ_p la liste des racines du polynôme χ_A (deux à deux distinctes) ; il existe une matrice inversible $V \in M_n(\mathbb{K})$ telle que $V^{-1}AV$ soit diagonale par blocs,

$$T = V^{-1}AV = \begin{pmatrix} T_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & T_p \end{pmatrix},$$

où chaque bloc diagonal T_j est une matrice carrée triangulaire supérieure de la forme

$$T_j = \begin{pmatrix} \mu_j & * & \dots & * \\ 0 & \mu_j & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & \mu_j \end{pmatrix}.$$

Au total, T est triangulaire et peut s'écrire $T = D + N$ avec D diagonale et N strictement triangulaire, et $DN = ND$.

Cette forme est très utile pour calculer les puissance de A . En effet, la formule du binôme s'applique parce que $DN = ND$, les puissances de D sont évidentes à calculer et celles de N s'arrêtent à n . On a donc pour tout $k \geq 0$

$$T^k = D^k + kD^{k-1}N + \mathbf{C}_k^2 D^{k-2}N^2 + \dots + \mathbf{C}_k^{n-1} D^{k-n+1}N^{n-1},$$

et pour finir $A^k = VT^kV^{-1}$.

Exercice. Réduire sous cette forme la matrice compagnon du polynôme $X^2(X + 1)^2$.

5.3. Polynômes d'endomorphismes

Soit E un espace vectoriel de dimension finie > 0 sur \mathbb{K} ; on rappelle que le produit de deux endomorphismes est défini par la composition des applications, $uv = u \circ v$.

Puissances et polynômes d'un endomorphisme

On pose $u^0 = \text{Id}_E$ pour tout $u \in \mathcal{L}(E)$, et on définit par récurrence $u^{m+1} = u^m u$ pour tout entier $m \geq 0$. On vérifie par récurrence sur k entier ≥ 0 que

$$u^{m+k} = u^m \circ u^k = u^k \circ u^m.$$

Si $P = c_0 + c_1X + \dots + c_kX^k \in \mathbb{K}[X]$ est un polynôme à coefficients dans \mathbb{K} , on définit un endomorphisme $P(u)$ par la formule

$$P(u) = c_0 \text{Id}_E + c_1 u + \dots + c_k u^k \in \mathcal{L}(E).$$

Pour bien comprendre le remplacement il vaut mieux écrire $P = c_0X^0 + \dots + c_kX^k \in \mathbb{K}[X]$, où X^0 n'est pas simplement le nombre 1 mais le polynôme constant égal à 1. Il faut faire attention à cet aspect un peu particulier de l'opération : il faut remplacer le polynôme

constant X^0 par $u^0 = \text{Id}_E$. Si P est un polynôme constant, disons $P = \lambda$, on pose $P(u) = \lambda \text{Id}_E$, en particulier lorsque P est le polynôme constant égal à 1, on a posé

$$1(u) = \text{Id}_E.$$

On voit facilement que $(P+Q)(u) = P(u)+Q(u)$ et $(\lambda P)(u) = \lambda P(u)$ pour tous polynômes $P, Q \in \mathbb{K}[X]$ et tout $\lambda \in \mathbb{K}$.

Exemples.

1. Si u est la rotation d'angle $+\pi/2$ dans \mathbb{R}^2 et si $P = X^2+1$, on trouve que $P(u) = 0$.
2. Application de $P(u)$ à un vecteur propre : si $u(x) = \lambda x$, on a $P(u)(x) = P(\lambda)x$ pour tout polynôme $P \in \mathbb{K}[X]$.
3. Supposons que u soit diagonalisable, et soit (e_1, \dots, e_n) une base de E formée de vecteurs propres de u , avec valeurs propres $\lambda_1, \dots, \lambda_n$; soit x un vecteur quelconque de l'espace E ; on écrit

$$x = a_1 e_1 + a_2 e_2 + \dots + a_n e_n,$$

ce qui donne en additionnant les égalités $P(u)(e_j) = P(\lambda_j)e_j$

$$P(u)(x) = a_1 P(\lambda_1)e_1 + a_2 P(\lambda_2)e_2 + \dots + a_n P(\lambda_n)e_n.$$

4. Avec les mêmes hypothèses (u diagonalisable), supposons que $P(\lambda_j) = 0$ pour tout $j = 1, \dots, n$. Il en résulte que $P(u)(x) = 0$ pour tout x , ce qui veut dire que $P(u)$ est l'endomorphisme nul. En particulier, si $P = \chi_u$ on trouve $\chi_u(u) = 0$. C'est le cas facile du théorème de Cayley-Hamilton, dont le cas général suit. En fait on peut faire un peu mieux : si μ_1, \dots, μ_p est la liste des valeurs propres sans répétition et si

$$P = (X - \mu_1) \dots (X - \mu_p),$$

le même argument montre que $P(u) = 0$. Cette remarque anticipe la notion de polynôme minimal qui sera vue plus loin.

La proposition qui suit est facile, mais elle joue un rôle important dans ce qui suit.

Proposition 5.3.1. *Si P et Q sont deux polynômes à coefficients dans \mathbb{K} et u un endomorphisme de E , on a*

$$P(u) \circ Q(u) = (PQ)(u) = Q(u) \circ P(u).$$

Démonstration. On le vérifie d'abord pour un monôme de la forme $Q = d_k X^k$. Dans ce cas, si $P = c_0 + c_1 X + \dots + c_p X^p$, on aura

$$(PQ)(u) = c_0 d_k u^k + \dots + c_p d_k u^{p+k} = (c_0 \text{Id}_E + \dots + c_p u^p)(d_k u^k) = P(u) \circ Q(u).$$

Ensuite, on décompose un polynôme quelconque Q en somme de monômes $Q = \sum_k Q_k$ et on utilise l'associativité,

$$\begin{aligned} (PQ)(u) &= \left(\sum_k PQ_k \right)(u) = \sum_k (PQ_k)(u) = \\ &= \sum_k P(u) \circ Q_k(u) = P(u) \circ \left(\sum_k Q_k(u) \right) = P(u) \circ Q(u). \end{aligned}$$

Corollaire 5.3.1. *Deux endomorphismes de la forme $P(u)$ et $Q(u)$ (pour le même u) commutent.*

On définit de même l'application d'un polynôme $P \in \mathbb{K}[X]$ à une matrice carrée A de taille $n \times n$ à coefficients dans \mathbb{K} ,

$$P(A) = c_0 I_n + c_1 A + \cdots + c_k A^k \in M_n(\mathbb{K}),$$

où I_n désigne la matrice identité de taille $n \times n$. Si A est la matrice d'un endomorphisme u par rapport à une base \mathbf{e} de E , la matrice $P(A)$ est la matrice de l'endomorphisme $P(u)$ dans la même base \mathbf{e} ,

$$\text{mat}_{\mathbf{e}}(P(u)) = P(\text{mat}_{\mathbf{e}}(u)).$$

On remarque dans le même ordre d'idées que $P(VAV^{-1}) = VP(A)V^{-1}$.

Si A est une matrice diagonale de taille $n \times n$, avec éléments diagonaux $\lambda_1, \dots, \lambda_n$, on voit facilement que pour tout entier $k \geq 0$, la matrice A^k est la matrice diagonale de diagonale $\lambda_1^k, \dots, \lambda_n^k$. Il en résulte immédiatement que pour tout polynôme P , la matrice $P(A)$ est la matrice diagonale de diagonale $P(\lambda_1), \dots, P(\lambda_n)$. Cette remarque se généralise facilement aux matrices diagonales par blocs.

Si A est une matrice triangulaire supérieure de taille $n \times n$, avec éléments diagonaux $\lambda_1, \dots, \lambda_n$, on voit aussi que pour tout entier $k \geq 0$, la matrice A^k est triangulaire supérieure de diagonale $\lambda_1^k, \dots, \lambda_n^k$ (mais les coefficients au dessus de la diagonale sont en général difficiles à calculer). Il en résulte immédiatement que pour tout polynôme P , la matrice $P(A)$ est triangulaire supérieure, de diagonale $P(\lambda_1), \dots, P(\lambda_n)$. Cette remarque se généralise également aux matrices triangulaires par blocs.

Remarque. Si F est stable par u , il est stable par tout endomorphisme de la forme $P(u)$, où P est un polynôme quelconque à coefficients dans \mathbb{K} . Par exemple, $\ker u$ et $u(E)$ sont stables par $P(u)$ pour tout polynôme P . Rappelons que si $u \circ v = v \circ u$ alors les deux sous-espaces $\ker v$ et $v(E)$ sont stables par u ; ils sont alors stables par $P(u)$ pour tout polynôme $P \in \mathbb{K}[X]$.

Sous-espace stable engendré par un vecteur non nul

Soit x un vecteur non nul d'un espace vectoriel E de dimension finie n sur \mathbb{K} , et soit $u \in \mathcal{L}(E)$; on va décrire le plus petit sous-espace stable par u contenant le vecteur x , et qui sera noté $F_{x,u}$, ou simplement F_x si l'endomorphisme u est clairement identifié par le contexte. Puisque E est de dimension n , il n'est pas possible que les $n + 1$ vecteurs du système $(x, u(x), \dots, u^n(x))$ soient linéairement indépendants; il existe donc un entier $k \leq n$ qui est le plus petit entier $\ell > 0$ tel que $u^\ell(x) \in [x, u(x), \dots, u^{\ell-1}(x)]$. Posons $x_j = u^j(x)$ pour $j = 0, \dots, k - 1$. Ces k vecteurs sont linéairement indépendants, puisque $x_0 = x \neq 0$ et que $x_i \notin [x_0, \dots, x_{i-1}]$ pour $i < k$. Posons $G = [x, u(x), u^2(x), \dots, u^{k-1}(x)]$ et montrons que G est stable par u . Pour cela, il suffit de vérifier que les images par u des générateurs x_0, \dots, x_{k-1} de G restent dans G ; pour $j < k - 1$ on a $u(x_j) = x_{j+1} \in G$, et pour le dernier générateur $x_{k-1} = u^{k-1}(x)$ on a $u(x_{k-1}) = u^k(x) \in [x, u(x), \dots, u^{k-1}(x)] = G$ d'après la définition de k . Remarquons ensuite que tout sous-espace F stable par u et qui contient x contient aussi $u(x), u^2(x), \dots$, donc $G = [x, u(x), u^2(x), \dots, u^{k-1}(x)]$ est contenu dans tout sous-espace stable F contenant x . Tout ceci montre que $F_x = G$ est le plus petit sous-espace stable contenant x , et que les vecteurs $(x, u(x), \dots, u^{k-1}(x))$ forment une base de F_x . Par ailleurs, il existe par définition de k des coefficients a_0, \dots, a_{k-1} dans \mathbb{K} tels que

$$u^k(x) + a_{k-1}u^{k-1}(x) + \cdots + a_1u(x) + a_0x = 0.$$

Si nous posons $P_x = P_{x,u} = a_0 + a_1X + \dots + a_{k-1}X^{k-1} + X^k$, nous voyons que

$$P_x(u)(x) = 0.$$

On dira que le polynôme $P_x = P_{x,u}$ est le *polynôme annulateur de x* associé à l'endomorphisme u . Dans la base $(x, u(x), \dots, u^{k-1}(x))$, la matrice de la restriction de u à F_x a la forme

$$M(P_x) = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ddots & 0 & -a_{k-2} \\ 0 & 0 & \dots & 1 & -a_{k-1} \end{pmatrix},$$

c'est-à-dire qu'elle est égale à la matrice compagnon $M(P_x)$ du polynôme P_x .

Remarquons que si R est un polynôme non nul de $\mathbb{K}[X]$ de degré $\leq k-1$, on a $R(u)(x) \neq 0$. En effet $R = c_0 + c_1X + \dots + c_{k-1}X^{k-1}$, donc

$$R(u)(x) = c_0x + c_1u(x) + \dots + c_{k-1}u^{k-1}(x) \neq 0$$

puisque $(x, u(x), \dots, u^{k-1}(x))$ est une base et que les coefficients (c_i) ne sont pas tous nuls. Le polynôme P_x est donc le polynôme unitaire P du plus petit degré tel que $P(u)(x) = 0$. Son degré est égal à la dimension de F_x .

Exemples.

Rotation de $\pi/2$ autour de l'axe Oz dans $E = \mathbb{R}^3$, avec $x = (1, 0, 1)$. Dans ce cas $F_x = \mathbb{R}^3$ et $P_x = X^3 - X^2 + X - 1$.

Projection $(y_1, y_2, y_3) \rightarrow (y_1, y_2, 0)$ de \mathbb{R}^3 sur le plan $y_3 = 0$, en partant de $x = (1, 0, 1)$. Dans ce cas F_x est le sous-espace de dimension deux formé de tous les vecteurs $(y_1, 0, y_3)$, et $P_x = X^2 - X$.

Partons de l'application linéaire de \mathbb{R}^3 définie par $u(y) = (y_1, 2y_2, 3y_3)$ pour tout $y = (y_1, y_2, y_3) \in \mathbb{R}^3$. Quand on part de $x = \mathbf{e}_2$ par exemple, on a affaire à un vecteur propre, donc $u(\mathbf{e}_2) = 2\mathbf{e}_2$, donc $F_x = \mathbb{R}\mathbf{e}_2$, $P_x = X - 2$. Pour $x = (1, 1, 1)$, les vecteurs successifs

$$x = (1, 1, 1), \quad u(x) = (1, 2, 3), \quad u^2(x) = (1, 4, 9)$$

forment une base de \mathbb{R}^3 , donc $F_x = \mathbb{R}^3$ et on voit que $P_x = (X - 1)(X - 2)(X - 3)$.

Il ne faut pas croire qu'on puisse toujours trouver un vecteur x qui donne un sous-espace F_x égal à E tout entier : exemple de $u = \text{Id}_E$, ou de la projection sur un plan vue ci-dessus.

Proposition 5.3.2. *Soit E un espace de dimension finie > 0 et soit $u \in \mathcal{L}(E)$; pour tout vecteur $x \neq 0$, le polynôme $P_{x,u}$ défini ci-dessus divise le polynôme caractéristique de u .*

Démonstration. Dans une base de E commençant par la base de F_x

$$(x, u(x), \dots, u^{k-1}(x)),$$

la matrice de u sera triangulaire par blocs pour la décomposition $E = F_x \oplus G$, avec un bloc diagonal « nord-ouest » égal à la matrice compagnon de $P_x = P_{x,u}$, ce qui donne le résultat d'après la proposition 4.4.2 et le fait que le polynôme caractéristique de $M(P_x)$ est égal à $\pm P_x$.

Théorème de Cayley-Hamilton

Le théorème de Cayley-Hamilton pour une matrice carrée A dit que

$$\chi_A(A) = 0.$$

Il existe de nombreuses démonstrations de ce théorème. Commençons par indiquer des cas particuliers faciles ; on peut faire une vérification directe en dimension deux par calcul matriciel. Dans le cas d'une matrice diagonale A de diagonale $\lambda_1, \dots, \lambda_n$, on sait que les valeurs propres sont $\lambda_1, \dots, \lambda_n$ et on sait que $P(A)$ est la matrice diagonale dont la diagonale est $P(\lambda_1), \dots, P(\lambda_n)$. Si $P = \chi_A$, les λ_i sont racines de P donc $\chi_A(A) = 0$ pour une matrice diagonale. Si B est seulement diagonalisable, on aura $B = V^{-1}AV$ pour une certaine matrice inversible V et une certaine matrice diagonale A . On aura alors $\chi_B = \chi_A$ et

$$\chi_B(B) = V^{-1}\chi_B(A)V = 0.$$

Le théorème de Cayley-Hamilton est donc facile dans le cas d'une matrice diagonalisable.

Théorème 5.3.1. (Théorème de Cayley-Hamilton) *Soit E un espace vectoriel de dimension finie > 0 sur \mathbb{K} ; pour tout endomorphisme $u \in \mathcal{L}(E)$, le polynôme caractéristique annule l'endomorphisme u ,*

$$\chi_u(u) = 0_{\mathcal{L}(E)}.$$

Démonstration. Soit x un vecteur de E ; si $x = 0_E$, on a bien $\chi_u(u)(x) = 0_E$. Sinon, soit $P_x = P_{x,u}$ le polynôme annulateur de x associé à l'endomorphisme u ; on sait que $P_x(u)(x) = 0_E$ et d'après la proposition 5.3.2, le polynôme P_x divise le polynôme χ_u , donc il existe un polynôme Q tel que $\chi_u = QP_x$. Alors $\chi_u(u) = Q(u)P_x(u)$ et $\chi_u(u)(x) = Q(u)(P_x(u)(x)) = 0_E$. Ainsi l'endomorphisme $\chi_u(u)$ annule tous les vecteurs de E , c'est-à-dire que

$$\chi_u(u) = 0_{\mathcal{L}(E)}.$$

5.4. Idéaux de polynômes. Polynôme minimal

On dit qu'un polynôme $P \in \mathbb{K}[X]$ est *unitaire* si le coefficient de son terme de plus haut degré est égal à 1,

$$P = X^n + a_{n-1}X^{n-1} + \dots + a_0.$$

Un polynôme unitaire est donc non nul par définition.

Définition 5.4.1. Un ensemble I de polynômes de $\mathbb{K}[X]$ est un *idéal* de $\mathbb{K}[X]$ si

1. L'ensemble I contient le polynôme nul 0 ,
2. I contient $P - Q$ chaque fois qu'il contient les polynômes P et Q ,
3. Pour tous $P \in I$ et pour tout polynôme $Q \in \mathbb{K}[X]$, on a $PQ \in I$.

Si $P \in I$, alors $\alpha P \in I$ pour tout $\alpha \in \mathbb{K}$. Si $P \neq 0$ appartient à I , on voit que I contient aussi le polynôme unitaire P_1 de même degré que P obtenu en multipliant P par l'inverse du coefficient du terme de plus haut degré de P .

Exemples d'idéaux.

1. $\{0\}$ et $\mathbb{K}[X]$ sont deux idéaux de $\mathbb{K}[X]$.
2. Si P est un polynôme de $\mathbb{K}[X]$, l'ensemble des multiples de P , noté

$$(P) = \{AP : A \in \mathbb{K}[X]\}$$

est un idéal de $\mathbb{K}[X]$.

3. Les polynômes P à coefficients réels tels que $P(i) = 0$ forment un idéal de $\mathbb{R}[X]$. Posons $\alpha = 2^{1/3}$. L'ensemble des polynômes à coefficients rationnels qui s'annulent en α est un idéal de $\mathbb{Q}[X]$; il contient le polynôme unitaire $X^3 - 2$.

4. Soit u un endomorphisme d'un espace vectoriel E de dimension finie > 0 ; l'ensemble

$$I(u) = \{P \in \mathbb{K}[X] : P(u) = 0\}$$

est un idéal de $\mathbb{K}[X]$. On sait que χ_u appartient à cet idéal, donc cet idéal est non nul.

5. Soit C une partie d'un espace vectoriel E de dimension finie > 0 et soit $u \in \mathcal{L}(E)$; l'ensemble

$$I_C(u) = \{P \in \mathbb{K}[X] : \forall x \in C, P(u)(x) = 0\}$$

est un idéal de $\mathbb{K}[X]$. En particulier, $I_E(u) = I(u)$ est l'idéal des polynômes tels que $P(u) = 0$. Si $x \in E$ est non nul et $C = \{x\}$, l'idéal $I_x(u)$ contient le polynôme $P_{x,u}$ de la proposition 5.3.2.

Théorème 5.4.1. *Soit I un idéal de polynômes de $\mathbb{K}[X]$, avec $I \neq \{0\}$; il existe dans I un (unique) polynôme unitaire P de plus petit degré, et l'idéal I est exactement l'ensemble (P) des multiples de P .*

On dit que P est le *générateur unitaire* de I .

Démonstration. Puisque $I \neq \{0\}$, il contient des polynômes non nuls. Soit k le plus petit degré d'un polynôme non nul de I , et soit P un polynôme non nul tel que $P \in I$ et $\deg P = k$; on peut supposer que P est unitaire (en multipliant P par une constante). Montrons que tous les éléments de I sont multiples de P . Soit $S \in I$, et effectuons la division euclidienne de S par P ,

$$S = PQ + R, \quad \deg R < \deg P$$

(avec la convention usuelle $\deg 0 = -\infty$). Puisque $P \in I$, on a $PQ \in I$ et puisque $S \in I$, $R = S - PQ \in I$. Si on avait $R \neq 0$, on aurait trouvé dans I un polynôme non nul de degré $< k$, ce qui est impossible par la définition de k . Donc $R = 0$ et on a montré que tous les éléments de I sont multiples de P . Inversement tout multiple de P est dans I par la définition d'un idéal. On a donc montré que $I = (P)$. Si $Q \in I$ est unitaire et $\deg Q = \deg P$, on a $R = P - Q \in I$ et $\deg R < \deg P$, donc $R = 0$, $Q = P$, donc P est le seul polynôme unitaire de degré minimal dans I .

Exemples.

1. Soit $u \in \mathcal{L}(E)$ et soit $x \neq 0$; le polynôme $P_x = P_{x,u}$ est le générateur unitaire de $I_x(u)$; en effet, on l'a introduit comme le polynôme unitaire de plus petit degré tel que $P_x(u)(x) = 0$. Si $y \in F_x$, le polynôme P_y divise P_x . Dans le cas où P_x est irréductible, on voit que $F_y = F_x$ pour tout y non nul dans F_x (parce que $\dim F_y = \deg P_y$).

2. Les polynômes P à coefficients réels tels que $P(i) = 0$ forment un idéal de $\mathbb{R}[X]$ dont le générateur unitaire est $X^2 + 1$. En effet, aucun polynôme unitaire à coefficients réels de degré < 2 ne s'annule au point i .

3. Posons $\alpha = 2^{1/3}$. L'idéal des polynômes à coefficients rationnels qui s'annulent en α contient le polynôme unitaire $X^3 - 2$. Soit P le générateur unitaire de cet idéal ; si $P \neq X^3 - 2$, on aura $X^3 - 2 = PQ$, avec $P, Q \in \mathbb{Q}[X]$, donc P ou Q sera de degré 1, donc aura une racine rationnelle, ce qui est impossible puisque la seule racine réelle de $X^3 - 2$ est α qui n'est pas rationnel.

Remarque. Soit I un idéal de $\mathbb{K}[X]$, avec $\{0\} \neq I \neq \mathbb{K}[X]$ et soit P un polynôme unitaire irréductible de $\mathbb{K}[X]$; si $P \in I$, alors $I = (P)$. En effet, puisque $I \neq \{0\}$ cet idéal admet un générateur unitaire G et $G \neq 1$ puisque $I \neq \mathbb{K}[X]$. Le polynôme G est alors un diviseur unitaire différent de 1 du polynôme unitaire irréductible P , donc $P = G$.

Théorème 5.4.2. (Bézout) *Soient P et Q deux polynômes de $\mathbb{K}[X]$ premiers entre eux. Il existe $U, V \in \mathbb{K}[X]$ tels que*

$$1 = UP + VQ.$$

Démonstration. On montre facilement que l'ensemble de tous les polynômes de la forme $UP + VQ$, lorsque U, V varient dans $\mathbb{K}[X]$ est un idéal I . Cet idéal n'est pas $\{0\}$ parce que P et Q ne sont pas tous les deux nuls. Soit G le générateur unitaire de l'idéal I . Puisque $P = 1P + 0Q$ et $Q = 0P + 1Q$, les deux polynômes P et Q sont dans I , ils sont par conséquent multiples de G , donc $G = 1$ puisque P et Q sont premiers entre eux, et $1 = G$ est bien de la forme $UP + VQ$ puisque $G \in I$.

Polynôme minimal d'un endomorphisme, d'une matrice

Soient E un \mathbb{K} -espace vectoriel de dimension finie > 0 et $u \in \mathcal{L}(E)$; d'après le théorème de Cayley-Hamilton (théorème 5.3.1), l'idéal $I(u)$ de $\mathbb{K}[X]$ formé des polynômes $P \in \mathbb{K}[X]$ tels que $P(u) = 0$ n'est pas réduit à 0. Le générateur unitaire de l'idéal $I(u)$ est le polynôme unitaire P de plus petit degré tel que $P(u) = 0$. On l'appelle le *polynôme minimal* de u , il sera noté m_u dans ces notes.

Le polynôme caractéristique χ_u est donc un multiple du polynôme minimal. En particulier, le degré du polynôme minimal est $\leq \dim E$. Il est clair aussi que $\deg m_u \geq 1$ (parce que $E \neq \{0\}$).

Si A est une matrice à coefficients dans \mathbb{K} , on peut considérer l'idéal $I(A)$ de $\mathbb{K}[X]$ formé de tous les polynômes P à coefficients dans \mathbb{K} tels que $P(A) = 0$. C'est un idéal non nul, dont le générateur unitaire est le *polynôme minimal* de la matrice A , noté m_A .

Soit e une base d'un \mathbb{K} -espace vectoriel E de dimension finie, soit $u \in \mathcal{L}(E)$ et $A = \text{mat}_e(u)$; alors $P(u) = 0$ si et seulement si $P(A) = 0$, donc $I(u) = I(A)$ et $m_u = m_A$.

Exemples de polynôme minimal.

- a. La rotation d'angle $\pi/2$ dans le plan \mathbb{R}^2 a pour polynôme minimal $X^2 + 1$.
- b. Si $m_u = X - \lambda$, alors $u = \lambda Id$ (et réciproquement).
- c. Si $m_u = X(X - 1)$, u est un projecteur et l'image $F = u(E)$ est telle que $\{0\} \neq F \neq E$.

d. Posons $\alpha = 2^{1/3}$. Soit $E \subset \mathbb{C}$ l'espace vectoriel de dimension ≤ 3 sur \mathbb{Q} engendré par $1, \alpha, \alpha^2$; la multiplication par α définit un endomorphisme u_α de E . Pour tout polynôme $P \in \mathbb{Q}[X]$ l'endomorphisme $P(u_\alpha)$ est la multiplication par $P(\alpha)$ dans \mathbb{C} , donc $P(u_\alpha) = 0$ si et seulement si $P(\alpha) = 0$, donc l'idéal $I(u_\alpha)$ est égal à l'idéal de $\mathbb{Q}[X]$ formé des polynômes à coefficients rationnels qui s'annulent en α , dont le générateur unitaire est $X^3 - 2$.

Valeurs propres et polynôme minimal

Proposition 5.4.1. Soient E un \mathbb{K} -espace vectoriel de dimension finie et $u \in \mathcal{L}(E)$; les polynômes χ_u et m_u ont les mêmes racines complexes.

Démonstration. Puisque m_u divise χ_u , chaque racine de m_u est une racine de χ_u . Réciproquement, soit e une base de E et soit A la matrice de u dans cette base ; soit $\lambda \in \mathbb{C}$ tel que $\chi_u(\lambda) = 0$; il existe alors un vecteur colonne non nul $Z \in \mathbb{C}^n$ tel que $AZ = \lambda Z$, et il en résulte que $P(A)Z = P(\lambda)Z$ pour tout polynôme $P \in \mathbb{K}[X]$. En particulier,

$$0 = m_u(A)Z = m_u(\lambda)Z,$$

donc $m_u(\lambda) = 0$.

Puisque les valeurs propres de u sont les racines de χ_u dans \mathbb{K} , on obtient :

Corollaire 5.4.1. Les valeurs propres de $u \in \mathcal{L}(E)$ sont celles des racines du polynôme minimal m_u qui appartiennent au corps \mathbb{K} .

Exemples.

1. Polynôme minimal d'un endomorphisme diagonalisable : soit $u \in \mathcal{L}(E)$ un endomorphisme diagonalisable et soient $\mu_1, \dots, \mu_p \in \mathbb{K}$ les valeurs propres de u (deux à deux distinctes) ; d'après ce qui précède, chaque μ_i est une racine de m_u , donc

$$P = (X - \mu_1) \dots (X - \mu_p)$$

divise m_u . Pour chaque i , écrivons $P = (X - \mu_i)Q_i$ avec $Q_i \in \mathbb{K}[X]$. Soit $x_i \in E_{\mu_i}$; on aura

$$P(u)(x_i) = Q_i(u)(u - \mu_i \text{Id}_E)(x_i) = 0$$

donc $P(u)$ annule chaque sous-espace propre E_{μ_i} de u , donc $P(u)$ annule E tout entier puisque $E = \sum_{i=1}^p E_{\mu_i}$, ce qui entraîne que m_u divise P , donc

$$m_u = (X - \mu_1) \dots (X - \mu_p).$$

Remarquons que m_u a toutes ses racines dans \mathbb{K} , et toutes ses racines sont simples.

2. Matrice de Jordan. Si la matrice de u dans une base de E est une matrice de taille $n \times n$ de la forme

$$\begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \ddots & \vdots \\ 0 & 0 & \lambda & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \dots & 0 & 0 & \lambda \end{pmatrix}$$

la seule racine possible de m_u est λ , donc m_u est de la forme $(X - \lambda)^r$, avec $r \leq n$. On vérifie qu'en posant $v = u - \lambda \text{Id}_E$ on a $v(e_j) = e_{j-1}$ pour $j \geq 2$, donc $v^{n-1}(e_n) = e_1 \neq 0$, donc $r = n$.

3. Calcul du polynôme minimal d'une matrice compagnon de la forme

$$M(P) = \begin{pmatrix} 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & \dots & 0 & -a_1 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ddots & 0 & -a_{n-2} \\ 0 & 0 & \dots & 1 & -a_{n-1} \end{pmatrix}.$$

Soit $(\mathbf{e}_1, \dots, \mathbf{e}_n)$ la base canonique de \mathbb{K}^n ; désignons par u l'endomorphisme de \mathbb{K}^n dont la matrice dans la base canonique est $M(P)$. En posant $x_0 = \mathbf{e}_1$, on voit que $u^j(x_0) = \mathbf{e}_{j+1}$ pour tout $j < n$, ce qui montre que tout vecteur de \mathbb{K}^n peut s'écrire $Q(u)(x_0)$ et que $R(u) \neq 0$ lorsque $\deg R < n$. Le polynôme unitaire

$$P = X^n + a_{n-1}X^{n-1} + \dots + a_1X + a_0$$

est tel que $P(u)(x_0) = 0$, donc $P(u) = 0$, donc P est le polynôme minimal de la matrice $M(P)$.

4. Polynôme minimal d'un endomorphisme lorsque l'espace est somme directe de deux sous-espaces stables. Dans ce cas le polynôme minimal est le PPCM des polynômes minimaux des deux restrictions u_1 et u_2 . Cette remarque s'applique aux matrices diagonales par blocs.

Proposition 5.4.2. *Les polynômes minimal et caractéristique ont les mêmes facteurs irréductibles dans $\mathbb{K}[X]$.*

Démonstration. Si P est un facteur irréductible de m_u , il divise χ_u , puisque m_u divise χ_u d'après Cayley-Hamilton. Inversement, soit Q un facteur irréductible (et unitaire) de χ_u et soit λ une racine complexe de Q , donc de χ_u ; on a vu que λ est racine de m_u . Considérons l'idéal $I = (G)$ (avec G unitaire) des polynômes à coefficients dans \mathbb{K} qui s'annulent au point λ . Alors $G \neq 1$ et G divise Q , donc $G = Q$, et Q divise m_u puisque m_u appartient à l'idéal I .

Critère de diagonalisabilité

Théorème 5.4.3. *Soient E un espace vectoriel de dimension finie > 0 sur \mathbb{K} et u un endomorphisme de E ; l'endomorphisme u est diagonalisable si et seulement si son polynôme minimal m_u se décompose en facteurs du premier degré dans $\mathbb{K}[X]$, et si les racines sont deux à deux distinctes.*

Démonstration. On a vu que lorsque u est diagonalisable, on a

$$m_u = (X - \mu_1) \dots (X - \mu_p)$$

où μ_1, \dots, μ_p est la liste des valeurs propres de u (deux à deux distinctes ; par définition, $\mu_i \in \mathbb{K}$). Inversement, supposons que

$$m_u = (X - \mu_1) \dots (X - \mu_k)$$

avec μ_1, \dots, μ_k éléments deux à deux distincts de \mathbb{K} . Il est utile de dégager la remarque suivante, qu'il suffira d'appliquer au polynôme m_u pour terminer.

Lemme. *Soient μ_1, \dots, μ_k des éléments deux à deux distincts de \mathbb{K} , et considérons le polynôme*

$$P = (X - \mu_1) \dots (X - \mu_k).$$

Si $P(u) = 0$, l'endomorphisme u est diagonalisable. En résumé, si un endomorphisme u est annulé par un polynôme ayant toutes ses racines simples et toutes dans \mathbb{K} , il est diagonalisable.

Démonstration. Pour chaque $j = 1, \dots, k$ soit P_j le polynôme de degré $k - 1$ qui est le produit des facteurs $(X - \mu_i)$ pour $i \neq j$ et soit $c_j = P(\mu_j)^{-1}$; le polynôme

$$1 - \sum_{j=1}^k c_j P_j$$

est de degré $\leq k - 1$ et s'annule en k points distincts (il s'annule en chaque μ_j) donc ce polynôme est identiquement nul. Soit $v_j = c_j P_j(u)$; puisque $P = (X - \mu_j)P_j$ et $P(u) = 0$, il est clair que l'image de v_j est contenue dans le noyau de $u - \mu_j \text{Id}_E$. On a

$$\forall x \in E, \quad x = \sum_j v_j(x),$$

ce qui montre que E est la somme des sous-espaces propres de u , donc u est diagonalisable.

Exemple. La rotation d'angle $\pi/2$ dans \mathbb{R}^2 admet $(X^2 + 1)$ pour polynôme minimal. Les racines de ce polynôme sont $\pm i$, qui ne sont pas dans $\mathbb{K} = \mathbb{R}$, donc la rotation n'a pas de vecteur propre. En revanche, tout endomorphisme d'un espace vectoriel complexe tel que $m_u = X^2 + 1$ est diagonalisable (et plus généralement d'après le lemme, tout endomorphisme u d'un espace vectoriel complexe E tel que $u^2 = -\text{Id}_E$ est diagonalisable).

Sous-espaces stables d'un endomorphisme diagonalisable

Soit $u \in \mathcal{L}(E)$ diagonalisable et soit G un sous-espace stable par u ; désignons par $v \in \mathcal{L}(G)$ la restriction de u à G . Puisque m_u annule v , m_u est un multiple de m_v . Puisque m_u a toutes ses racines simples, on voit que m_v est le produit de certains des facteurs $(X - \mu_i)$ de m_u . L'espace G est donc de la forme

$$G = G_{\mu_1} + \dots + G_{\mu_p},$$

somme de sous-espaces G_{μ_i} des sous-espaces propres E_{μ_i} de u , et l'endomorphisme v est diagonalisable.

Invariance du polynôme minimal

Théorème 5.4.4. *Le polynôme minimal d'une matrice $A \in M_n(\mathbb{K})$ ne dépend pas du corps (pour tous les sous-corps L de \mathbb{C} contenant les coefficients de la matrice A , le polynôme de $L[X]$ de degré minimal qui annule A est le même).*

Démonstration. Soit L un sous-corps de \mathbb{K} contenant les coefficients de la matrice A , et soit m_L le polynôme minimal calculé dans $M_n(L)$. *A fortiori*, m_L est un polynôme à coefficients dans \mathbb{K} qui annule A , donc m_L est un multiple de m_A . On a donc $p = \deg m_L \geq \deg m_A = q$. Mais dire que m_L est de degré $\geq p$ équivaut à dire que les matrices

$$I_n, A, A^2, \dots, A^{p-1}$$

sont des vecteurs L -linéairement indépendants du L -espace $M_n(L)$, que l'on peut considérer comme étant l'espace vectoriel L^{n^2} . D'après la proposition 4.4.3, ces matrices restent \mathbb{K} -linéairement indépendantes dans $\mathbb{K}^{n^2} \simeq M_n(\mathbb{K})$, ce qui implique que le degré de m_A ne peut pas être $< p$. On a donc $p = q$ et $m_A = m_L$.

Sous-espaces caractéristiques et polynôme minimal

Rappels sur les projecteurs. Un endomorphisme p de E est appelé *projecteur* si $p^2 = p$. Cela équivaut à dire que $p(y) = y$ pour tout $y \in p(E)$. Si p est un projecteur, $\text{Id}_E - p$ est aussi un projecteur. Si p est un projecteur, on obtient une décomposition de E en somme directe $E = F \oplus G$ en posant $F = p(E)$ et $G = (\text{Id}_E - p)(E)$. En effet, pour tout $x \in E$,

$$x = (p + (\text{Id}_E - p))(x) = p(x) + (\text{Id}_E - p)(x) \in F + G$$

et si $x \in F \cap G$, on a $x = p(x) = x - p(x)$ donc $p(x) = x = 0$.

Lemme des noyaux. Soient E un espace vectoriel de dimension finie > 0 sur \mathbb{K} et $u \in \mathcal{L}(E)$.

- a. Pour tout polynôme $P \in \mathbb{K}[X]$, le noyau $F = \ker P(u)$ est stable par u .
- b. Si $P = P_1 P_2$, avec $P_1, P_2 \in \mathbb{K}[X]$ premiers entre eux, on a

$$\ker P(u) = \ker P_1(u) \oplus \ker P_2(u).$$

Démonstration. Le point a est un rappel. Passons au second point. D'après Bézout il existe $V_1, V_2 \in \mathbb{K}[X]$ tels que

$$1 = V_1 P_1 + V_2 P_2.$$

Posons $p = P(u)$, $F = \ker p$, et $p_i = P_i(u)$, $v_i = V_i(u)$, $F_i = \ker p_i$ pour $i = 1, 2$. On a $p = p_1 p_2$ et $\text{Id}_E = v_1 p_1 + v_2 p_2$. Il est clair que $F_i \subset F$ puisque $p = p_1 p_2$. Montrons que $F = F_1 + F_2$. Pour tout $y \in F$, on a $y = (v_1 p_1)(y) + (v_2 p_2)(y)$, et $p_2(v_1 p_1)(y) = v_1(p(y)) = 0$ puisque $y \in \ker p$, donc $(v_1 p_1)(y) \in F_2$ et de même $(v_2 p_2)(y) \in F_1$, donc on a bien montré que $F = F_1 + F_2$. Montrons que la somme est directe : si $y \in F_1 \cap F_2$, on aura $y = v_1(p_1(y)) + v_2(p_2(y)) = 0$.

Corollaire 5.4.2. Soient E un espace vectoriel de dimension finie > 0 sur \mathbb{K} , $u \in \mathcal{L}(E)$ et P un polynôme à coefficients dans \mathbb{K} ; si $P(u) = 0$ et si $P = P_1 \dots P_k$ avec des polynômes (P_j) deux à deux premiers entre eux, et tels que $\ker P_j(u) \neq \{0\}$ pour tout $j = 1, \dots, k$, on a en posant $F_j = \ker P_j(u)$, pour $j = 1, \dots, k$

- chaque espace F_j est stable par u .
- on a $E = F_1 \oplus \dots \oplus F_k$.

Démonstration. On montre par récurrence sur $j = 1, \dots, k$ que

$$\ker(P_1 \dots P_j)(u) = F_1 \oplus \dots \oplus F_j.$$

Remarque. On retrouve le critère de diagonalisation.

Considérons un endomorphisme u . Supposons que m_u ait toutes ses racines dans \mathbb{K} (cette hypothèse est automatique si $\mathbb{K} = \mathbb{C}$). Soit μ_1, \dots, μ_k la liste des racines du polynôme minimal m_u (deux à deux distinctes) ; on a

$$m_u = (X - \mu_1)^{p_1} \dots (X - \mu_k)^{p_k},$$

où p_1, \dots, p_k sont des entiers ≥ 1 . D'après les corollaires précédents, on voit que E est la somme directe des $E_i = \ker(u - \mu_i \text{Id}_E)^{p_i}$. On voit aussi que le polynôme minimal de chaque restriction u_i de u à E_i est égal à $(X - \mu_i)^{p_i}$.

Exercice. Soit μ l'une des racines de m_u , et soit p sa multiplicité comme racine de m_u (par exemple, $\mu = \mu_1$ et $p = p_1$). Montrer que le plus petit entier ℓ tel que $\ker(u - \mu \text{Id}_E)^\ell = \ker(u - \mu \text{Id}_E)^{\ell+1}$ est égal à p .

Le noyau de l'endomorphisme $(u - \mu_j \text{Id}_E)^{p_j}$ est donc le sous-espace caractéristique relatif à la valeur propre μ_j

$$F_{\mu_j} = \ker(u - \mu_j \text{Id}_E)^{p_j} = \ker(u - \mu_j \text{Id}_E)^{r_j}.$$

Il est clair que F_{μ_j} contient le sous-espace propre E_{μ_j} . Nous avons retrouvé le théorème 5.2.2 : les sous-espaces caractéristiques forment une somme directe, et cette somme est égale à E ,

$$E = F_{\mu_1} \oplus \dots \oplus F_{\mu_k}.$$

6. Équations différentielles linéaires

On va étudier dans ce chapitre les équations différentielles linéaires, mais il peut être utile d'introduire la notion plus générale d'équation différentielle ; une équation différentielle du premier ordre s'écrit (symboliquement) sous la forme

$$(1) \quad y' = F(x, y)$$

où x est une variable réelle, $y \in \mathbb{R}^d$, F est une fonction à valeurs dans \mathbb{R}^d , définie sur une partie $\Gamma \subset \mathbb{R} \times \mathbb{R}^d$. L'idée est que y représente une fonction dérivable $x \rightarrow y(x)$, dont la dérivée y' doit vérifier la relation (1). Il faut cependant prendre quelques précautions pour définir la notion de solution de l'équation (1).

Exemple. Cherchons les solutions de l'équation $y' = y^2$ telles que $y(0) = a$, avec $a > 0$. On voit que « la » solution « explose » à un point $x_1 > 0$ dépendant de la valeur de a . En effet, on trouve

$$y(x) = \frac{a}{1 - ax}$$

qui tend vers $+\infty$ lorsque $x \rightarrow 1/a$ (par valeurs inférieures). On voit donc qu'on ne peut pas imposer en même temps et *a priori* l'intervalle sur lequel on cherche une solution et la valeur de la solution en un point de cet intervalle ; c'est ce qui conduira à faire figurer un intervalle I dans la définition d'une solution. La variable, que nous avons appelée x jusqu'ici, peut aussi bien s'appeler t (si on pense qu'elle représente le *temps* par exemple), mais on réservera la lettre y pour représenter la fonction.

Définition 6.0.1. Une *solution de l'équation* (1) est un couple (I, φ) , où I est un intervalle ouvert de \mathbb{R} , φ une fonction dérivable de I dans \mathbb{R}^d , telle que $(x, \varphi(x)) \in \Gamma$ pour tout $x \in I$ et telle que

$$\forall x \in I, \quad \varphi'(x) = F(x, \varphi(x)).$$

L'une des façons de poser un problème d'équation différentielle est la suivante : on se donne une valeur initiale $x_0 \in \mathbb{R}$ et une donnée initiale $y^{(0)} \in \mathbb{R}^d$, telles que $(x_0, y^{(0)}) \in \Gamma$, et on cherche la ou les solutions (I, φ) telles que $x_0 \in I$ et $\varphi(x_0) = y^{(0)}$ (problème avec *donnée initiale*). On cherche ensuite à trouver le plus grand intervalle possible sur lequel on puisse définir chaque solution (notion de solution maximale).

Dans certains cas, l'équation est de la forme $y' = F(y)$, c'est-à-dire que F ne dépend pas de la variable x . On dit dans ce cas qu'il s'agit d'une équation *autonome*. On verra qu'on peut toujours ramener théoriquement le problème général avec F définie sur $\mathbb{R} \times \mathbb{R}^d$ à un problème autonome, mais où l'espace des valeurs sera \mathbb{R}^{d+1} au lieu de \mathbb{R}^d .

Pour une équation autonome, l'évolution d'une solution φ entre x_0 initial et x ne dépend que de la différence $x - x_0$. Pour être plus précis, disons que lorsque (I, φ) est une solution de $y' = F(y)$, avec $0 \in I$ et telle que $\varphi(0) = y^{(0)}$, on obtiendra une solution ψ telle que $\psi(x_0) = y^{(0)}$, définie sur un intervalle contenant x_0 en faisant tout simplement une translation,

$$\psi(x) = \varphi(x - x_0).$$

On peut aussi étudier des équations différentielles d'ordre plus élevé, par exemple les équations du second ordre, de la forme $y'' = F(x, y, y')$. Ces équations se ramènent aussi, au moins du point de vue théorique, à des équations du premier ordre, mais au prix du remplacement de l'espace des valeurs \mathbb{R}^d par un espace de dimension plus grande.

6.1. Équations différentielles linéaires à coefficients constants

On va étudier pour commencer l'une des équations différentielles les plus simples, et dans le cas $d = 1$

$$(L1) \quad y' = ay$$

avec $a \in \mathbb{R}$. Si $a = 0$ on obtient l'équation simple mais importante $y' = 0$: ses seules solutions sont de la forme (I, c) , où c est une fonction constante sur l'intervalle I .

Les solutions de (L1) sont bien connues mais on va les introduire par le procédé des approximations successives qui interviendra à nouveau plus loin pour résoudre des équations différentielles très générales. On cherche à résoudre (L1) avec la condition $y(0) = y_0$, où $y_0 \in \mathbb{R}$ est donné. L'idée de base est que l'opération d'intégration est « plus stable » que l'opération de dérivation. Si on a une approximation raisonnable ψ_1 d'une solution φ , on peut espérer améliorer cette approximation en résolvant $\psi_2' = a\psi_1$, puis en remplaçant ψ_1 par ψ_2 ; voici une façon d'explicitier cette idée d'amélioration : supposons que ψ_1 soit une approximation de la solution idéale φ , approximation suffisamment bonne pour avoir, par exemple, le même DL d'ordre 2 au voisinage de 0 que la solution idéale. On voit qu'en définissant ψ_2 par $\psi_2(0) = y_0$ et $\psi_2' = a\psi_1$, c'est maintenant ψ_2' qui a le même DL d'ordre 2 que $\varphi' = a\varphi$, donc par intégration de DL on conclut que le DL d'ordre 3 de ψ_2 au voisinage de 0 coïncide avec celui de φ , et on a bien obtenu une certaine amélioration. On peut ensuite envisager de recommencer cette opération une infinité de fois et essayer d'obtenir une vraie solution à la limite.

On définit donc une suite de fonctions $(\varphi_n)_{n \geq 0}$ en posant $\varphi_0(x) = y_0$ (fonction constante ; c'est le choix le plus simple mais on pourrait prendre à peu près n'importe quoi d'autre) et en définissant φ_{n+1} par $\varphi_{n+1}'(x) = a\varphi_n(x)$ et $\varphi_{n+1}(0) = y_0$. On obtient

$$\varphi_n(x) = y_0 \left(1 + ax + \frac{a^2 x^2}{2!} + \cdots + \frac{a^n x^n}{n!} \right).$$

On voit que $\varphi_n(x)$ converge vers $y_0 e^{ax}$ pour tout $x \in \mathbb{R}$. On pourrait en déduire que le couple $(\mathbb{R}, x \rightarrow e^{ax})$ est solution, mais ce serait idiot tellement cette affirmation est évidente par le calcul (mais cet argument de passage à la limite devra être utilisé plus loin pour le cas général).

Unicité des solutions de L1

Soit (I, φ) une solution de l'équation L1 ; posons $\psi(x) = e^{-ax} \varphi(x)$ pour tout $x \in I$. On voit que $\psi'(x) = e^{-ax}(\varphi'(x) - a\varphi(x)) = 0$, donc ψ est constante sur l'intervalle I , ce qui montre qu'il existe une constante C telle que pour tout $x \in I$, on ait $\varphi(x) = C e^{ax}$. On a donc montré que les seules solutions de l'équation (L1) sont de la forme $(I, x \rightarrow C e^{ax})$.

Équation avec second membre

On suppose maintenant donnée une fonction continue g définie sur un intervalle I , et on cherche à résoudre l'équation, dite *équation avec second membre*

$$(L1S) \quad y' - ay = g(x).$$

Supposons que (I, φ) soit solution de (L1S) et posons $\psi(x) = e^{-ax} \varphi(x)$ comme avant. On aura ici

$$\psi'(x) = e^{-ax}(\varphi'(x) - a\varphi(x)) = e^{-ax} g(x).$$

Supposons que x_0 soit un point de l'intervalle I. On obtiendra en calculant $\psi(x) = \psi(x_0) + \int_{x_0}^x \psi'(u) du$ l'expression de φ sous la forme

$$\varphi(x) = e^{a(x-x_0)} \varphi(x_0) + e^{ax} \int_{x_0}^x e^{-au} g(u) du.$$

Linéarité

Il est clair que si φ_1 et φ_2 sont deux solutions de $y' = ay$ définies sur le même intervalle I, les combinaisons linéaires $\lambda_1\varphi_1 + \lambda_2\varphi_2$ sont également solutions. Cette remarque très simple s'étend immédiatement aux équations linéaires à coefficients variables, $y'(x) = a(x)y(x)$, où a est une fonction continue définie sur un intervalle I. Il n'est pas difficile de résoudre cette équation par calcul de primitive. Si $x \rightarrow b(x)$ est une primitive de a sur I, on vérifie que $\varphi(x) = e^{b(x)}$ est solution. La transformation $\psi(x) = e^{-b(x)} \varphi(x)$ permet encore de montrer l'unicité et de résoudre l'équation avec second membre $y'(x) = a(x)y(x) + g(x)$.

Remarquons que si φ_1 et φ_2 sont deux solutions de l'équation avec second membre $y'(x) = a(x)y(x) + g(x)$, la différence $\varphi_1 - \varphi_2$ est solution de l'équation sans second membre $y'(x) = a(x)y(x)$. D'où le principe suivant : pour résoudre l'équation avec second membre, il suffit de trouver une solution particulière de cette équation avec second membre, et de lui ajouter la « solution générale » de l'équation sans second membre.

Cas complexe

Si on suppose maintenant que $a \in \mathbb{C}$, la fonction $y(x)$ doit être supposée à valeurs complexes. À part ça tout est pareil, y compris le cas avec second membre (et avec une fonction g elle aussi à valeurs complexes).

6.2. Systèmes différentiels linéaires

On considère maintenant l'équation différentielle linéaire du premier ordre générale, qui s'écrit sous la forme

$$(L) \quad y' = Ay$$

où A est une matrice réelle $d \times d$, et où y, y' sont aussi des fonctions à valeurs dans \mathbb{R}^d . Si on considère les fonctions coordonnées y_1, \dots, y_d de la fonction vectorielle y , on voit que l'équation (L) est équivalente au système différentiel suivant

$$\begin{aligned} y_1'(t) &= a_{1,1} y_1(t) + \dots + a_{1,d} y_d(t) \\ y_2'(t) &= a_{2,1} y_1(t) + \dots + a_{2,d} y_d(t) \\ &\dots \\ y_d'(t) &= a_{d,1} y_1(t) + \dots + a_{d,d} y_d(t), \end{aligned}$$

où les $(a_{i,j})$ sont les coefficients de la matrice A .

Avant de développer la théorie générale, étudions quelques exemples très simples, dans le cas $d = 2$. Pour aider l'intuition, on imagine que $y(t)$ est la position au temps t d'un point mobile dans \mathbb{R}^2 dont la vitesse $y'(t)$ vérifie à tout instant t la relation $y'(t) = Ay(t)$. On va s'intéresser à la trajectoire parcourue par le mobile, et trouver si possible une équation cartésienne pour celle-ci.

Quelques exemples simples avec description des trajectoires

Considérons par exemple le système différentiel

$$y_1' = y_1, \quad y_2' = -y_2.$$

On trouve $y_1(t) = C e^t$, $y_2(t) = D e^{-t}$. En supposant $C \neq 0$, on trouve l'équation cartésienne $y_2 = CD/y_1$ pour la trajectoire du mobile. C'est une trajectoire hyperbolique. Si on remplace la première équation par $y_1' = ay_1$, avec $a > 0$, on trouve des courbes de la forme $y_2 = C^{1/a} D y_1^{-1/a}$, qui ne sont plus des hyperboles mais ont un peu la même allure générale (les axes de coordonnées sont asymptotes).

Remarquons que dans tous les cas, si on change tous les signes dans la matrice du système (c'est-à-dire que l'on remplace la matrice A du système par la matrice $-A$), les trajectoires seront les mêmes mais parcourues dans le sens inverse. Dans tous ces exemples le point 0 est fixe, stable ou instable. Si on étudie le système

$$y_1' = -y_1, \quad y_2' = -2y_2$$

on trouve des trajectoires paraboliques qui tendent vers 0 lorsque $t \rightarrow +\infty$ (en fait le mouvement se fait sur une demi-parabole. Si on change les signes, le mouvement « part de 0 » lorsque $t = -\infty$ et s'éloigne à l'infini sur la même demi-parabole quand $t \rightarrow +\infty$).

Si la matrice A est diagonalisable sur \mathbb{R} , on obtiendra un exemple d'un des types précédents en se plaçant dans une base de \mathbb{R}^2 formée de vecteurs propres de A . Si la matrice A a la forme de Jordan,

$$y_1' = ay_1 + y_2, \quad y_2' = ay_2,$$

on trouve d'abord $y_2(t) = C e^{at}$, puis on reporte dans la première équation, en obtenant ainsi une équation avec second membre. Il vient

$$y_1(t) = D e^{at} + C t e^{at}.$$

La trajectoire a une équation moins sympathique, de la forme $y_1 = cy_2 + dy_2 \ln(y_2)$. Considérons pour finir le système

$$y_1' = -y_2, \quad y_2' = y_1.$$

Ici la matrice n'a pas de valeur propre réelle, mais on peut étendre le problème en système différentiel complexe, et revenir au problème initial par changement de base. On obtient des trajectoires circulaires. Si on remplace la première équation par $y_1' = -by_2$, avec $b > 0$ et $b \neq 1$, on trouve des trajectoires en spirale.

Cas général. Exponentielle de matrices

Commençons par quelques remarques sur l'équation différentielle linéaire à coefficients constants générale $y' = Ay$. On va s'intéresser plus particulièrement à la résolution de l'équation avec donnée initiale. On se donne un vecteur $y^{(0)} \in \mathbb{R}^d$, et on cherche une solution φ définie sur \mathbb{R} et telle que $\varphi(0) = y^{(0)}$ (par un changement de variable simple, on pourra traiter de même toute condition initiale de la forme $\psi(t_0) = y^{(0)}$, en posant simplement $\psi(t) = \varphi(t - t_0)$).

Remarquons avant toute chose que la fonction nulle est toujours solution de l'équation $y' = Ay$. C'est l'un des aspects de la *linéarité* de l'équation. Les remarques sur la linéarité du cas $d = 1$ sont encore valables ici. Il est clair que si φ_1 et φ_2 sont deux solutions de $y' = Ay$ définies sur le même intervalle I , les combinaisons linéaires $c_1\varphi_1 + c_2\varphi_2$ sont également solutions.

1. Si $y^{(0)}$ est un vecteur propre de A , tel que $Ay^{(0)} = \lambda y^{(0)}$, on voit facilement que

$$\varphi(t) = e^{\lambda t} y^{(0)}$$

est solution de l'équation. Si on a une base $(v^{(1)}, \dots, v^{(d)})$ de \mathbb{R}^d formée de vecteurs propres de A , avec $Av^{(j)} = \lambda_j v^{(j)}$ pour $j = 1, \dots, d$, il est facile de résoudre l'équation

avec n'importe quelle donnée initiale $y^{(0)}$. On peut décomposer $y^{(0)}$ dans la base de vecteurs propres,

$$y^{(0)} = \sum_{j=1}^d c_j v^{(j)}$$

et on constate que

$$\varphi(t) = \sum_{j=1}^d c_j e^{\lambda_j t} v^{(j)}$$

est solution de l'équation, avec $\varphi(0) = y^{(0)}$.

2. Si $\lambda = \alpha + i\beta$ est une valeur propre complexe de la matrice A , avec $\alpha, \beta \in \mathbb{R}$, $\beta \neq 0$, il est facile d'en déduire des solutions réelles de l'équation. Soit en effet $z = x + iy \in \mathbb{C}^d$ un vecteur propre complexe de A , avec $x, y \in \mathbb{R}^d$; on voit de la même façon que $\varphi(t) = e^{\lambda t} z$ définit une solution complexe de l'équation. En séparant partie réelle et partie imaginaire, on obtient deux solutions réelles,

$$t \rightarrow e^{\alpha t}(\cos(\beta t)x - \sin(\beta t)y), \quad t \rightarrow e^{\alpha t}(\sin(\beta t)x + \cos(\beta t)y).$$

La combinaison des remarques 1 et 2 permet de traiter le cas où la matrice A est diagonalisable sur \mathbb{C} .

3. On peut aussi utiliser la triangulation (sur \mathbb{R} ou sur \mathbb{C} , selon le cas). Illustrons ceci sur un exemple simple. Supposons que A soit écrite sous la forme $A = V^{-1}TV$, avec V inversible et T triangulaire supérieure, par exemple

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} a & 3 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

Supposons donnée la condition initiale $y^{(0)} = (4, 5)$. En posant $u(t) = Vy(t)$, on voit que l'équation $y' = Ay$ se ramène à l'équation triangulaire $u' = Tu$, avec condition initiale $u^{(0)} = Vy^{(0)} = (-1, 9)$, c'est-à-dire

$$u'_1 = au_1 + 3u_2, \quad u'_2 = u_2,$$

$u_1(0) = -1$, $u_2(0) = 9$. On obtient immédiatement avec la deuxième équation $u_2(t) = 9e^t$. En reportant dans la première, on a

$$u'_1 - au_1 = 27e^t,$$

équation avec second membre qui se résout comme on a dit plus haut. On observera une distinction selon que $a = 1$ ou $a \neq 1$. Pour finir, on écrira $y(t) = V^{-1}u(t)$. Cette méthode s'applique aussi si on utilise la triangulation complexe. Les calculs intermédiaires font appel à des fonctions complexes, mais la dernière étape (retour à y) doit les faire disparaître.

Approche théorique générale

Le même principe d'approximations successives que dans le cas $d = 1$ fournit ici une suite d'approximations $(\varphi_n)_{n \geq 0}$ à valeurs dans \mathbb{R}^d définies par

$$\varphi_n(t) = \left(I_d + tA + \frac{t^2}{2!}A^2 + \cdots + \frac{t^n}{n!}A^n \right) y^{(0)}$$

pour tout $n \geq 0$ et tout $t \in \mathbb{R}$, et conduit à introduire l'exponentielle de matrice e^{tA} pour résoudre le problème. Rappelons que

$$e^{tA} = \sum_{n=0}^{+\infty} \frac{t^n}{n!} A^n = \lim_{n \rightarrow +\infty} \sum_{k=0}^n \frac{t^k}{k!} A^k.$$

Notons que $e^{tI_d} = e^t I_d$ pour tout t réel (ou complexe) ; en particulier, si 0_d désigne la matrice nulle de taille $d \times d$, on a $e^{0_d} = I_d$. Donnons une propriété essentielle.

Lemme. *Si les matrices A et B commutent, on a*

$$e^{A+B} = e^A e^B.$$

En particulier, la matrice e^A est inversible et son inverse est e^{-A} .

On en déduit la propriété importante $e^{tA} e^{-tA} = I_d$. On rappelle aussi que

$$(**) \quad \frac{d}{dt} e^{tA} = A e^{tA} = e^{tA} A.$$

Il est alors facile de voir à partir des propriétés précédentes la première partie de la proposition qui suit. La deuxième partie concerne l'unicité des solutions.

Proposition 6.2.1. *Pour tout $y^{(0)} \in \mathbb{R}^d$ donné, la fonction φ à valeurs dans \mathbb{R}^d , définie sur \mathbb{R} par*

$$\forall t \in \mathbb{R}, \quad \varphi(t) = e^{tA} y^{(0)}$$

est une solution de l'équation (L). Si (I, φ) est une solution de l'équation $y' = Ay$, il existe $y^{(0)} \in \mathbb{R}^d$ tel que

$$\forall t \in I, \quad \varphi(t) = e^{tA} y^{(0)}.$$

Toute solution (I, φ) est donc restriction à I d'une solution définie sur \mathbb{R} par $t \rightarrow e^{tA} y^{(0)}$.

Démonstration. La première partie résulte immédiatement du calcul de la dérivée de e^{tA} . Pour la deuxième, on pose comme toujours $\psi(t) = e^{-tA} \varphi(t)$, qui donne par dérivation (d'un produit matriciel)

$$\psi'(t) = e^{-tA} (-A\varphi(t) + \varphi'(t)) = 0,$$

donc ψ est constante sur I, égale disons à $y^{(0)}$. En multipliant $\psi(t)$ par e^{tA} , on obtient le résultat voulu.

Résumé. *Pour tout intervalle non vide $I \subset \mathbb{R}$, pour tout point $t_0 \in I$ et pour tout vecteur $y^{(0)}$ donné dans \mathbb{R}^d il existe une unique solution (I, φ) de l'équation $y' = Ay$ vérifiant la condition initiale $\varphi(t_0) = y^{(0)}$. Cette solution peut s'exprimer par $\varphi(t) = e^{(t-t_0)A} y^{(0)}$.*

Linéarité. Équation avec second membre

On a dit que si φ_1 et φ_2 sont deux solutions de l'équation différentielle $y' = Ay$ définies sur le même intervalle I, les combinaisons linéaires $c_1\varphi_1 + c_2\varphi_2$ sont également solutions, donc l'ensemble des solutions de l'équation $y' = Ay$ est un sous-espace vectoriel \mathcal{E} de l'espace des fonctions de I dans \mathbb{R}^d ; cet espace \mathcal{E} des solutions est de dimension d . En effet pour $t_0 \in I$ fixé, l'application Φ définie par $\varphi \in \mathcal{E} \rightarrow \varphi(t_0) \in \mathbb{R}^d$ est bijective de \mathcal{E} sur \mathbb{R}^d d'après les résultats de la proposition 6.2.1 : étant donné $y^{(0)} \in \mathbb{R}^d$ quelconque, la fonction φ définie par $\varphi(t) = e^{(t-t_0)A} y^{(0)}$ est une solution telle que $\varphi(t_0) = y^{(0)}$, ce qui montre la surjectivité de Φ , et cette solution est unique, ce qui donne l'injectivité.

Si $v^{(1)}, \dots, v^{(d)}$ sont des données initiales indépendantes dans \mathbb{R}^d , les solutions correspondantes restent indépendantes pour toute valeur de t : c'est clair puisque la matrice e^{tA} est inversible.

Exercice. Wronskien. Soient $y^{(1)}, \dots, y^{(d)}$ des solutions de l'équation $y' = Ay$. Si on désigne par $M(t)$ la matrice $d \times d$ dont les colonnes sont les vecteurs $y^{(1)}(t), \dots, y^{(d)}(t)$, montrer que la fonction $f(t) = \det M(t)$ vérifie l'équation différentielle $f' = (\operatorname{tr} A) f$. On appelle ce déterminant le *déterminant Wronskien*.

Une bonne partie des remarques très simples sur la linéarité s'étend encore aux équations linéaires à coefficients variables, $y'(x) = A(x)y(x)$, où A est une fonction continue définie sur un intervalle I et à valeurs matrices $d \times d$. En revanche, il n'est plus possible en général de résoudre cette équation par calcul de primitive.

Exercice. Soit $x \rightarrow B(x)$ une primitive de $x \rightarrow A(x)$ sur I , c'est-à-dire une fonction à valeurs matrices dont les coefficients sont des primitives des coefficients correspondants de $A(x)$. Si pour tout $x \in I$ la matrice $B(x)$ commute avec la matrice $A(x)$, montrer que l'équation différentielle linéaire $y'(x) = A(x)y(x)$ admet $x \rightarrow e^{B(x)} y^{(0)}$ pour solution.

Remarquons que si φ_1 et φ_2 sont deux solutions de l'équation avec second membre $y'(x) = A(x)y(x) + g(x)$ (où g est une fonction vectorielle continue sur I), la différence $\varphi_1 - \varphi_2$ est solution de l'équation « sans second membre » $y'(x) = A(x)y(x)$. On peut donc encore dire que pour résoudre l'équation avec second membre, il suffit de trouver une solution particulière, et de lui ajouter la « solution générale » de l'équation sans second membre.

Dans le cas de l'équation avec second membre à coefficients constants $y' - Ay = g(t)$, où g est une fonction continue définie sur un intervalle I , à valeurs dans \mathbb{R}^d , la même méthode que lorsque $d = 1$ permet de résoudre le problème. Si on se donne $t_0 \in I$, on voit que les solutions s'expriment par la formule

$$\varphi(t) = e^{(t-t_0)A} \varphi(t_0) + e^{tA} \int_{t_0}^t e^{-uA} g(u) du.$$

Considérons le problème du ressort, $y'' + y = 0$ ou bien $y'' + y = f(t)$. On va le transformer en système différentiel 2×2 en considérant la donnée vectorielle $Y = (y, y') = (Y_1, Y_2)$. On a les équations

$$Y_1' = y' = Y_2; \quad Y_2' = y'' = -y = -Y_1$$

ce qui donne notre système différentiel. On va voir sur cet exemple comment fonctionnent les méthodes théoriques introduites ci-dessus. Il s'agit dans ce paragraphe d'une *illustration*, ça n'est pas la méthode pratique que vous appliquerez par la suite pour résoudre une équation différentielle du second ordre à coefficients constants.

Tout d'abord, la matrice du système qui traduit l'équation $y'' + y = 0$ est

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{et} \quad e^{tA} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

(l'exponentielle est une rotation). On trouve donc la solution générale du système homogène,

$$e^{tA} \begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} \cos t \\ -\sin t \end{pmatrix} + b \begin{pmatrix} \sin t \\ \cos t \end{pmatrix}.$$

Passons à l'équation du ressort dont l'oscillation est forcée par un terme $f(t)$, l'équation étant alors $y'' + y = f(t)$. En passant au système on a

$$Y' = AY + \begin{pmatrix} 0 \\ f(t) \end{pmatrix}.$$

On peut théoriquement résoudre par la formule

$$e^{tA} \int_0^t e^{-uA} g(u) du \quad \text{où } g(u) = \begin{pmatrix} 0 \\ f(u) \end{pmatrix};$$

dans le cas particulier où $f(t) = \sin(\omega t)$, les calculs sont pénibles mais faisables. On aperçoit le problème de résonance quand $\omega = 1$.

Calcul pratique d'exponentielles de matrices

On se sert de la théorie de la réduction des matrices complexes, qui utilise la décomposition de \mathbb{C}^d en sous-espaces caractéristiques de la matrice A et la triangulation des blocs caractéristiques. On sait ainsi que la matrice A peut s'écrire $A = VMV^{-1}$, où V est inversible, et où M est une matrice diagonale par blocs, où chaque bloc diagonal est une matrice triangulaire T_α , avec une valeur constante $\lambda_\alpha \in \mathbb{C}$ sur la diagonale (λ_α est l'une des valeurs propres (complexes) de la matrice A). On obtient alors $e^{tA} = Ve^{tM}V^{-1}$, et e^{tM} est diagonale par blocs, chaque bloc étant égal à e^{tT_α} . Il reste à expliquer l'allure de e^{tT} , où T est l'un quelconque de ces blocs diagonaux de M , de taille $p \times p$. On peut écrire $T = \lambda I_p + N$, où N est nilpotente, ce qui donnera puisque λI_p commute avec N et que $N^p = 0$

$$e^{tT} = e^{t\lambda I_p} e^{tN} = e^{t\lambda} \left(I_p + tN + \frac{t^2}{2!} N^2 + \cdots + \frac{t^{p-1}}{(p-1)!} N^{p-1} \right).$$

Si T a la forme de Jordan on a une information un peu plus précise.

Conséquence. *Si toutes les valeurs propres (réelles ou complexes) de la matrice A sont de partie réelle < 0 , la matrice e^{tA} tend vers 0 lorsque $t \rightarrow +\infty$ (on verra plus bas que c'est si et seulement si). Il en résulte que dans ce cas toutes les solutions de $y' = Ay$ tendent vers 0 lorsque $t \rightarrow +\infty$, quelle que soit la donnée initiale $y^{(0)}$.*

Pour montrer que e^{tA} tend vers 0 quand $t \rightarrow +\infty$, il suffit de montrer que e^{tT} tend vers 0 pour chacun des blocs triangulaires $T = T_\alpha$ de la matrice M introduite ci-dessus ; on a supposé que chacune des valeurs propres λ de A a une partie réelle < 0 , donc $\lambda = a + ib$ avec $a < 0$. On a donc $|e^{t\lambda}| = e^{at}$ qui tend vers 0 rapidement, ce qui entraîne que $e^{t\lambda} t^j$ tend vers 0 quand $t \rightarrow +\infty$ pour tout $j = 0, \dots, p-1$, donc chacun des morceaux de la décomposition de e^{tT} donnée ci-dessus tend vers 0. Inversement, si e^{tA} tend vers 0 quand $t \rightarrow +\infty$, toutes les valeurs propres de la matrice A ont une partie réelle < 0 : si $\lambda = a + ib$ est valeur propre, il existe un vecteur Z complexe non nul tel que $AZ = \lambda Z$, et alors $e^{tA}Z = e^{t\lambda}Z$ tend vers 0, donc $e^{t\lambda}$ tend vers 0 quand $t \rightarrow +\infty$, ce qui entraîne $a < 0$.

Équations différentielles linéaires d'ordre plus élevé, à coefficients constants

Considérons d'abord l'équation du second ordre

$$y'' + py' + qy = 0$$

avec $p, q \in \mathbb{R}$. On peut transformer cette équation en système, en posant $y_1 = y$ et $y_2 = y'$,

$$y'_1 = y_2, \quad y'_2 = -qy_1 - py_2.$$

On obtient que la matrice du système est la matrice transposée de la matrice compagnon du polynôme $P = q + pX + X^2$. Les valeurs propres de la matrice sont donc les racines du trinôme P . Si les valeurs propres λ, μ sont distinctes, les solutions complexes sont de

la forme $c e^{\lambda t} + d e^{\mu t}$. Si ces valeurs propres ne sont pas réelles (quand $p^2 - 4q < 0$), on a $\lambda = \alpha + i\beta$ et $\mu = \alpha - i\beta$, avec $\alpha, \beta \in \mathbb{R}$ et les solutions réelles sont de la forme

$$e^{\alpha t}(c \cos(\beta t) + d \sin(\beta t)).$$

Si α est racine double du trinôme, les solutions sont de la forme $e^{\alpha t}(c + dt)$.

Exemple : oscillation d'un ressort avec amortissement. Résoudre l'équation $y'' + fy' + y = 0$, où $f \geq 0$, selon la valeur de $f \geq 0$ (qui représente un terme de frottement ; oscillations amorties, cas apériodique selon la valeur de f).

Considérons maintenant l'équation différentielle linéaire à coefficients constants d'ordre quelconque. L'équation différentielle d'ordre n

$$(*) \quad y^{(n)} + a_{n-1}y^{(n-1)} + \dots + a_1y' + a_0y = 0$$

se ramène à un système $n \times n$, dont la matrice A est la transposée de la matrice compagnon $M(P)$ du polynôme

$$P = X^n + a_{n-1}X^{n-1} + \dots + a_1X + a_0.$$

Les valeurs propres sont donc les racines du polynôme P .

On compare la résolution de l'équation (*) à celle du système

$$(**) \quad Y' = AY.$$

On établit facilement les principes suivants :

1- Si (I, φ) est solution de l'équation (*), et si on définit la fonction vectorielle Φ sur I par $\Phi(t) = (\varphi(t), \varphi'(t), \dots, \varphi^{(n-1)}(t)) \in \mathbb{R}^n$ (ou \mathbb{C}^n), le couple (I, Φ) est solution du système (**).

2- Si (I, Φ) est solution du système (**), et si $\Phi(t) = (\varphi_1(t), \dots, \varphi_n(t))$, alors (I, φ_1) est solution de l'équation (*) (et en fait chacune des fonctions coordonnées φ_j donne une solution).

La correspondance $\Phi \rightarrow \varphi_1$ est donc une bijection linéaire de l'espace vectoriel des solutions de (**) sur l'espace des solutions de l'équation (*). Cet espace est donc de dimension n (réel ou complexe, selon le cadre fixé). De plus, si t_0 est un point fixé de I , l'application $\varphi \rightarrow (\varphi(t_0), \varphi'(t_0), \dots, \varphi^{(n-1)}(t_0))$ est une bijection linéaire de l'espace des solutions de (*) sur \mathbb{R}^n (ou \mathbb{C}^n).

On rappelle que les valeurs propres de A sont les racines de P . Si $\lambda \in \mathbb{C}$ est une telle racine, on voit facilement que $X = (1, \lambda, \lambda^2, \dots, \lambda^{n-1})$ vérifie $AX = \lambda X$, ce qui donne une solution $\Phi(t) = e^{\lambda t} X$ pour le système, donc $\varphi(t) = e^{\lambda t}$ pour l'équation (*).

Si les racines $(\lambda_1, \dots, \lambda_n)$ de P sont deux à deux distinctes, on obtient ainsi une base de l'espace des solutions, formée des n fonctions $t \rightarrow e^{\lambda_j t}$, pour $j = 1, \dots, n$.

Exercice. Retrouver l'indépendance des solutions dans ce cas en utilisant l'isomorphisme ci-dessus pour $t_0 = 0$ (on rencontrera le déterminant classique de Vandermonde).

Cas des racines multiples pour P

Si

$$P = (X - \lambda_1)^{r_1} \dots (X - \lambda_p)^{r_p}$$

est la décomposition de P avec $(\lambda_1, \dots, \lambda_p)$ racines distinctes, on obtient une base de l'espace des solutions de (*) en faisant la liste des fonctions

$$t \rightarrow e^{\lambda_j t}, t e^{\lambda_j t}, \dots, t^{r_j-1} e^{\lambda_j t},$$

pour $j = 1, \dots, p$.

Index

Abel (transformation d')	31
Absolument convergente (série)	25
Alternée (forme linéaire)	59
Alternée (série)	30
Application identique	50
Application linéaire	49
Application linéaire injective	51
Application linéaire transposée	69
Approché (calcul d'une intégrale)	45
Approximations successives (procédé des)	96
Base canonique de \mathbb{K}^n	49
Base d'un espace vectoriel	53
Base d'une somme directe	55
Base duale	53, 67
Base incomplète (théorème de la)	53
Bertrand (série de)	23
Bézout (théorème de)	90
Bidual d'un espace vectoriel	70
Blocs (calcul matriciel par)	56
Blocs (calcul de déterminants par)	63
Bolzano-Weierstrass (théorème de)	11
Borne inférieure, borne supérieure	3
Borné (ensemble)	2
Calcul approché d'une intégrale	45
Calcul de déterminants par blocs	63
Calcul matriciel par blocs	56
Calcul pratique d'exponentielles de matrices	102
Canonique (base de \mathbb{K}^n)	49
Cantor (ensemble de)	15
Caractéristique (polynôme)	74
Caractéristique (sous-espace)	81, 93, 94
Cauchy (critère de Cauchy pour la convergence des séries)	25
Cayley-Hamilton (théorème de)	88
Changement de base	56
Changement de l'ordre des termes d'une série numérique	27
Chasles (relation de)	42
Codimension	67
Coefficients constants (équation différentielle linéaire à)	96
Combinaison linéaire	52
Commutables (endomorphismes)	64, 75, 78, 84, 85, 86
Compagnon (matrice compagnon d'un polynôme)	75, 87, 87, 91
Comparaison pour les séries à termes positifs (théorème de)	21
Comparaison séries-intégrales	23
Complet (espace)	10
Constante d'Euler	24
Construction des nombres réels	4
Continue (fonction sur un fermé borné de \mathbb{R})	12
Continuité uniforme	14
Convergente (suite de nombres réels)	5
Convergence de suites complexes	8
Convergence des suites de Cauchy	10
Convergente (série)	18

Correspondance séries-suites	19
Coordonnées dans une base	53
Coupures (dans \mathbb{Q} ou \mathbb{R})	1
Critère de Cauchy pour la convergence des séries	25
Critères de convergence des séries numériques	22, 24
Critères de diagonalisabilité	79, 92
d'Alembert (théorème de)	76
Décimal (développement)	21
Déterminant	59
Déterminant dans une base	61
Déterminant des matrices	62
Déterminant d'un endomorphisme	71, 72
Déterminant d'une matrice triangulaire	63
Déterminant wronskien	101
Développement décimal	21
Développement du déterminant d'une matrice suivant une ligne	62
Diagonale par blocs (matrice)	58
Diagonalisabilité et polynôme minimal	92
Diagonalisable (endomorphisme)	79
Dichotomie (démonstration par)	11
Différentiel (système différentiel linéaire)	97
Différentielle (équation)	95
Dimension d'un espace vectoriel	54
Dimension d'une somme directe	55
Divergente (série)	18
Division euclidienne	89
Donnée initiale	95
Droite achevée $\overline{\mathbb{R}}$	5
Dual d'un espace vectoriel	50, 66
Endomorphisme	50, 71
Endomorphismes commutables	64, 75, 78, 84, 85, 86
Endomorphisme diagonalisable	79
Ensemble borné	2
Ensemble de Cantor	15
Ensemble majoré	2
Ensemble ouvert	14
Équation différentielle	95
Équation différentielle linéaire à coefficients constants	96
Équation différentielle linéaire avec second membre	96
Équation différentielle linéaire d'ordre ≥ 2	102
Équivalentes (suites)	6, 21
Escalier (fonction en)	33
Espace complet	10
Espace des matrices à coefficients dans \mathbb{K}	55
Espace dual	66
Espace vectoriel sur un corps \mathbb{K}	49
Euclidienne (division)	89
Exponentielle complexe	28
Exponentielle de matrice	98
Exponentielle de matrice (calcul pratique)	102
Euler (constante d')	24
Fermé (sous-ensemble de \mathbb{R})	12
Fonction continue sur un fermé borné de \mathbb{R}	12
Fonction en escalier	33
Fonction exponentielle complexe	28
Fonction indicatrice	39

Fonction intégrable Riemann	36
Fonction lipschitzienne	38
Fonction uniformément continue	14
Forme linéaire	50
Forme linéaire alternée	59
Formes linéaires coordonnées dans une base	53
Formule de Stirling	24
Générateur (système de vecteurs)	52
Générateur unitaire d'un idéal de polynômes	89
Géométrie (série)	20
Harmonique (série harmonique alternée)	25
Hyperplan	67
Idéal de polynômes	88
Identique (application)	50
Identité (matrice)	56
Image d'une application linéaire	51
Indicatrice (fonction)	39
Indépendance linéaire et changement de corps de base	64
Injectivité (application linéaire)	51
Intégrable Riemann (fonction)	36
Intégrale	36
Intégrale (calcul approché d'une)	45
Intégrale d'une fonction en escalier	35
Intégrale de Lebesgue	36
Intégrale de Riemann	33
Intégrale et primitives	42
Isomorphisme canonique avec le bidual	70
Isomorphisme d'espaces vectoriels	51
Jordan (matrice de)	77, 91
Kronecker (symbole de)	67
Lebesgue (intégrale de)	36
Libre (système de vecteurs)	52
Limite d'une fonction en un point	7
Limite des suites numériques	5
Limite infinie	6
Limite supérieure d'une suite bornée de nombres réels	8, 15
Linéaire (forme)	50
Lipschitzienne (fonction)	38
Majorant	2
Majoré (ensemble)	2
Matrice compagnon d'un polynôme	75, 87, 87, 91
Matrice d'une application linéaire	55
Matrice de Jordan	77, 91
Matrice de passage	56
Matrice diagonale par blocs	58
Matrice identité	56
Matrice triangulaire par blocs	58
Méthode de Simpson	46
Méthode des trapèzes	45
Minimal (polynôme)	88, 91, 93
Minorant	2
Module d'une subdivision	38
Module de continuité	14
Noyau d'une application linéaire	51
Opérations sur les nombres réels	3
Ouvert de \mathbb{R}	14

Paquets (sommmation par paquets des séries)	25
Partie entière (d'un nombre réel)	6, 15
Pas d'une subdivision	38
Passage à la limite des inégalités	7
Passage (matrice de)	56
Polynôme caractéristique	74
Polynôme d'endomorphisme	84
Polynôme minimal d'un endomorphisme, d'une matrice	88, 93, 91, 93
Polynôme unitaire	88
Primitives et intégrale	42
Procédé des approximations successives	96
Produit d'applications linéaires et produit de matrices	56
Produit d'espaces vectoriels	50
Produit de séries numériques	27
Projections d'un produit sur les facteurs	50
Propriétés des fonctions continues sur un fermé borné de \mathbb{R}	12
Réduction des endomorphismes	71
Réduction réelle d'une matrice diagonalisable sur \mathbb{C}	80
Règle $n^\alpha u_n$	24
Relation de Chasles	42
Reste d'une série numérique	23
Restriction d'un endomorphisme	76
Riemann (intégrale de)	33
Riemann (série de)	24
Riemann (théorème de Riemann pour l'intégrale simple)	40
Second membre (équation différentielle linéaire avec)	96
Série à termes complexes	18
Série à termes positifs	20
Série à termes positifs (théorème de comparaison)	21
Série absolument convergente	25
Série alternée	30
Série convergente	18
Série de Bertrand	23
Série de Riemann	24
Série de Taylor	20
Série divergente	18
Série (comparaison avec une intégrale)	23
Série géométrique	20
Série harmonique alternée	25
Série numérique	17
Série numérique double	27
Série semi-convergente	29
Séries numériques (produit de)	27
Signature d'une permutation	60
Simpson (méthode de)	46
Solution d'une équation différentielle	95
Sommation par paquets des séries	25
Somme d'une série numérique	18
Somme de Riemann	39
Somme de sous-espaces vectoriels	50
Somme directe de sous-espaces vectoriels	54, 78
Sous-ensemble fermé de \mathbb{R}	12
Sous-espace caractéristique	81, 93, 94
Sous-espace propre d'un endomorphisme	78
Sous-espace stable	58, 58, 75, 76, 77, 78, 82, 93
Sous-espace stable engendré par un vecteur	86

Sous-espace vectoriel engendré	52
Sous-espaces caractéristiques et triangulation	83
Sous-suite	8, 11
Stable (sous-espace)	58, 58, 75, 76, 77, 78, 82, 93
Stirling (formule de)	24
Subdivision de $[a, b]$	33
Subdivision pointée	38
Suite convergente de nombres réels	5
Suite croissante et majorée	7
Suite de Cauchy	9
Suites équivalentes	6, 21
Symbole de Kronecker	67
Système différentiel linéaire	97
Système générateur	52
Système libre de vecteurs	52
Taylor (série de)	20
Terme général d'une série numérique	18
Termes positifs (série à)	20
Théorème de Bézout	90
Théorème de Bolzano-Weierstrass	11
Théorème de Cayley-Hamilton	88
Théorème de comparaison pour les séries à termes positifs	21
Théorème de d'Alembert	76
Théorème de la base incomplète	53
Théorème de Riemann pour l'intégrale simple	40
Topologie de \mathbb{R}	11
Trace d'un endomorphisme, d'une matrice	74
Transformation d'Abel	31
Transposée (application linéaire)	69
Trapèzes (méthode des)	45
Triangulaire par blocs (matrice)	58
Triangulation d'une matrice réelle	83
Triangulation et sous-espaces caractéristiques	83
Unicité des solutions d'une équation différentielle	96, 100
Uniformément continue (fonction)	14
Unitaire (polynôme)	88
Valeur d'adhérence	9, 15
Valeurs propres, vecteurs propres	73
Valeurs propres (complexes) d'une matrice	77
Vecteur propre	73
Vectoriel (espace)	49
Voisinage d'un point de \mathbb{R}	14
Wronskien (déterminant)	101

Index des notations

0_E : vecteur nul de l'espace vectoriel E	49
$[x]$: partie entière d'un nombre réel	6
$[x_1, \dots, x_n]$: sous-espace vectoriel engendré	52
(x_n) : suite numérique	5
$ a, b $: intervalle ouvert, fermé ou semi-ouvert	39
$\int_a^b f(t) dt$ ou $\int_a^b f$: intégrale de a à b de la fonction f	36

$A \setminus B$: différence ensembliste	1
$A^{\mathbb{C}}$: endomorphisme complexe associé à la matrice A	77
$\operatorname{ch}(z), \operatorname{sh}(z), z \in \mathbb{C}$: cosinus, sinus hyperboliques	29
$\delta_{i,j}$: symbole de Kronecker	67
$\delta(\pi)$: pas ou module d'une subdivision π	38
$\det_{\mathbf{e}}$: déterminant dans la base \mathbf{e}	61
$\det u$: déterminant de l'endomorphisme u	72
$\dim E$: dimension de l'espace vectoriel E	54
$e^z, z \in \mathbb{C}$: exponentielle complexe	29
E^* : dual de l'espace vectoriel E	50
$\mathbf{e}_1, \dots, \mathbf{e}_n$: base canonique de \mathbb{K}^n	49
e_j^* : forme linéaire coordonnée dans la base \mathbf{e}	53
E^{**} : bidual d'un espace vectoriel E	70
E_μ : sous-espace propre d'un endomorphisme	78
$\varepsilon(\sigma)$: signature de la permutation σ	60
$F_1 + F_2, F_1 + \dots + F_n$: somme de sous-espaces vectoriels	50
$F_1 \oplus F_2, F_1 \oplus \dots \oplus F_n$: sommes directes de sous-espaces vectoriels	54
Id_E : application identique de l'espace vectoriel E	50
I_n : matrice identité de taille $n \times n$	56
j_E : injection canonique de E dans E^{**}	70
\mathbb{K} : corps de base	49
$\ker u$: noyau de l'application linéaire u	51
$\mathcal{L}(E, F)$: espace des applications linéaires de E dans F	50
$\mathcal{L}(E)$: algèbre des endomorphismes de E	71
$\lim_n x_n$: limite d'une suite numérique	6
$\limsup_n x_n, \liminf_n x_n$: limite supérieure, inférieure, d'une suite numérique	9
$\lim_{x \rightarrow a, x \in I} f(x)$: limite d'une fonction en un point a	7
$M_{n,m}(\mathbb{K}), M_n(\mathbb{K})$: espace des matrices à coefficients dans \mathbb{K}	55
$\operatorname{mat}(u, \mathbf{e}, \mathbf{f})$: matrice de u par rapport aux bases \mathbf{e} et \mathbf{f}	55
$\operatorname{mat}_{\mathbf{e}}(u)$: matrice de u dans la base \mathbf{e}	72
$M(P)$: matrice compagnon du polynôme P	75
$\max f(X), \min f(X)$: maximum, minimum, de f sur X	13
m_u, m_A : polynôme minimal de l'endomorphisme u , de la matrice A	90
$P(u)$: polynôme en l'endomorphisme u	84
$\overline{\mathbb{R}}$: droite achevée	5
$\operatorname{Re} z, \operatorname{Im} z$: partie réelle, imaginaire, d'un nombre complexe	8
$R_n = \sum_{m=0}^{+\infty} u_m - U_n = u_{n+1} + u_{n+2} + \dots$: reste d'une série numérique	23
S_n : ensemble des permutations de $\{1, \dots, n\}$	60
$\Sigma_{\pi, \xi}(f)$: somme de Riemann de f	39
$\sum_{n=0}^{+\infty} u_n$, ou $u_0 + u_1 + \dots + u_n + \dots$: somme d'une série numérique	18
$\sum u_n$, ou $\sum (u_n)_{n \geq 0}$: série numérique	18
$\sup A, \inf A$: borne supérieure, inférieure, de l'ensemble $A \subset \mathbb{R}$	3
$x_n \sim y_n$: suites équivalentes	21
$\operatorname{tr} u, \operatorname{tr} A$: trace d'un endomorphisme u , d'une matrice A	74
${}^t a$: application linéaire transposée de a	69
$U_n = u_0 + \dots + u_n$: somme partielle d'une série numérique	18
χ_A : fonction indicatrice de l'ensemble A	39
χ_u, χ_A : polynôme caractéristique de l'endomorphisme u , de la matrice A	74, 75