

CHAPITRE 4

CONCENTRATION GAUSSIENNE. EXEMPLES D'APPLICATION

4.1. Vecteurs gaussiens

On rappelle brièvement (et en vrac) le vocabulaire des vecteurs gaussiens.

Dans toute la suite, on suppose qu'on a un espace de probabilité abstrait (Ω, \mathbb{P}) . Si E est un espace vectoriel de dimension finie (qu'on identifie en général à un \mathbb{R}^n) un vecteur aléatoire X de E est donc une application mesurable

$$\begin{aligned} X : \Omega &\longrightarrow E \\ \omega &\longrightarrow X_\omega. \end{aligned}$$

Si E est de dimension 1 ($E = \mathbb{R}$) on parle plutôt de *variable* aléatoire. En pratique on oublie Ω (et ω) et on s'intéresse à la mesure image $X_\# \mathbb{P}$ sur E , c'est pour cela qu'on écrit, pour toute partie borélienne $A \subset \mathbb{R}^n$,

$$\mathbb{P}(\{X \in A\}) = \mathbb{P}(\{\omega \in \Omega ; X_\omega \in A\}) \quad (= G_\# \mathbb{P}(A)).$$

Pour parler de vecteur gaussien, il faut se donner une structure euclidienne sur E . Sur \mathbb{R}^n muni de la structure euclidienne standard $|\cdot|$ (c'est-à-dire sur ℓ_2^n), on dit que le vecteur aléatoire X est *gaussien standard* si pour toute partie borélienne $A \subset E$

$$\mathbb{P}(\{X \in A\}) = \gamma_n(A) = \int_A e^{-|x|^2/2} \frac{dx}{(2\pi)^{n/2}}.$$

De manière équivalente, on peut dire que pour toute fonction borélienne (positive ou bornée) $f : E \longrightarrow \mathbb{R}^+$, la *variable* aléatoire $f(X)$ vérifie :

$$\mathbb{E}f(X) = \int_E f(x) d\gamma_n(x).$$

Ainsi, une variable aléatoire g est gaussienne standard si

$$\forall t \in \mathbb{R}, \quad \mathbb{P}(\{g \geq t\}) = \int_t^{+\infty} e^{-s^2} \frac{ds}{\sqrt{2\pi}}.$$

Notez que pour un vecteur gaussien standard X sur \mathbb{R}^n on a

$$\mathbb{E}X = 0 \quad \text{et} \quad \mathbb{E}|X|^2 = n$$

et plus généralement que $\mathbb{E}[(X \cdot e_i)(X \cdot e_j)] = \delta_{i,j}$, soit encore

$$\forall v, w \in \mathbb{R}^n, \quad \mathbb{E}[(X \cdot v)(X \cdot w)] = v \cdot w.$$

On remarque aussi que si $|u| = 1$, alors on a que $X \cdot u$ est un v.a. gaussienne standard. Pour le voir, prenons $u = e_1$ et remarquons que pour $g : \mathbb{R} \rightarrow \mathbb{R}$ on a

$$\mathbb{E}g(X \cdot e_1) = \int_{\mathbb{R}^n} g(x_1) \frac{e^{-|x_1|^2/2}}{\sqrt{2\pi}} \frac{e^{-\sum_{i=2}^n |x_i|^2/2}}{(\sqrt{2\pi})^{n-1}} dx_1 dx_2 \dots dx_n = 1 \times \int_{\mathbb{R}} g(t) d\gamma_1(t).$$

On va voir ci-dessous des extensions utiles de cette observation. Il est bon de retenir que les propriétés caractéristiques des vecteurs gaussiens (standards) dont nous aurons besoin découlent des deux propriétés suivantes caractéristiques de γ_n :

1. La mesure γ_n est invariante par isométries : $\gamma_n(RA) = \gamma_n(A)$ pour $A \subset \mathbb{R}^n$ et $R \in \mathcal{O}_n$.
2. La mesure γ_n est une mesure produit : pour $k, n \in \mathbb{N}^*$, $\gamma_k \otimes \gamma_n = \gamma_{n+k}$ et en particulier $\gamma_n = \gamma_1^{\otimes n}$.

Notez qu'on travaille par commodité de notation sur \mathbb{R}^n mais que les définitions et les propriétés sont les mêmes si l'on est sur un espace euclidien E de dimension finie. Ainsi, si $E \subset \mathbb{R}^n$ est un sous-espace vectoriel d \mathbb{R}^n (avec le produit scalaire usuel induit sur E^*) et si on note γ_E la mesure gaussienne standard sur E , on a que $\gamma_n = \gamma_E \otimes \gamma_{E^\perp}$.

On tire immédiatement des propriétés précédentes les résultats suivants :

Proposition. — Soit X un vecteur gaussien standard de \mathbb{R}^n et $E \subset \mathbb{R}^n$ un sous-espace vectoriel. On note P_E la projection orthogonale sur E dans \mathbb{R}^n . Alors,

$$P_E X$$

est un vecteur gaussien standard de E et les vecteurs

$$P_E X \quad \text{et} \quad P_{E^\perp} X$$

sont indépendants. En particulier, si $E \perp F$, alors $P_E X$ et $P_F X$ sont indépendants.

Démonstration. — Par invariance par rotation de γ_n , on peut supposer que $E = \text{vect}\{e_1, \dots, e_k\}$ (on remplace X par RX pour une certaine rotation R). Alors $P_E(x) = (x_1, \dots, x_k, 0, \dots, 0)$ et comme $\gamma_n = \gamma_E \otimes \gamma_{E^\perp}$ on a bien que $P_E \gamma_n = \gamma_k$ et que $P_E X$ et $P_{E^\perp} X$ sont indépendants. Enfin si $E \perp F$ on a $F \subset E^\perp$ et le résultat s'ensuit. \square

On peut aussi formuler le résultat avec une réciproque.

Proposition. — On considère une décomposition orthogonale de \mathbb{R}^n :

$$\mathbb{R}^n = E_1 \oplus^\perp \dots \oplus^\perp E_k.$$

Alors pour tout vecteur aléatoire X de \mathbb{R}^n on a l'équivalence entre

1. Le vecteur aléatoire X est gaussien standard
2. Chaque $P_{E_i} X$ est un vecteur gaussien standard de E_i , et les vecteurs $\{P_{E_1} X, \dots, P_{E_k} X\}$ sont indépendants.

En particulier, on a

$$X = (X_1, \dots, X_n) \text{ gaussien standard dans } \mathbb{R}^n$$

$$\iff \text{les } X_i, i \leq n, \text{ sont } n \text{ variables aléatoires gaussiennes standard indépendantes.}$$

Démonstration. — Le sens 1. \Rightarrow 2. découle de la proposition précédente, en remarquant que $P_1 X$ indépendant de $P_{E_1^\perp} X$ entraîne que $P_1 X$ indépendant de $\{P_{E_2} X, \dots, P_{E_k} X\}$. Le sens 2. \Rightarrow 1. est tout aussi évident. En effet, \mathbb{R}^n est isométrique à $E_1 \times \dots \times E_n$ et l'indépendance

entraîne que la loi de $(P_{E_1}X, \dots, P_{E_k}X)$ est la loi produit, qui est la loi gaussienne (standard). \square

On retient que se donner un vecteur gaussien standard sur \mathbb{R}^n , c'est équivalent à se donner n variables gaussiennes standard indépendantes (et une base orthonormée, comme par exemple la base canonique).

On tire immédiatement des propriétés précédentes l'observation classique mais utile (qu'on utilise surtout avec $E = \mathbb{R}$ et des variables aléatoires).

Proposition. — Soit X_1, \dots, X_k des vecteurs gaussiens standards indépendants sur \mathbb{R}^n . Alors, pour $\alpha_1, \dots, \alpha_k \in \mathbb{R}$ on a que :

$$\alpha_1 X_1 + \dots + \alpha_k X_k \text{ a la même loi que le vecteur } \left(\sum_{i=1}^k \alpha_i^2 \right)^{1/2} X_1.$$

Démonstration. — Il est suffisant de traiter le cas $n = 1$: le cas général s'obtient en travaillant avec les coordonnées $X_i \cdot e_j$ qui sont toutes indépendantes. On se donne donc des variables aléatoires $X_i = g_i$ gaussiennes standards indépendantes. Remarquez qu'en introduisant le vecteur $u = (\alpha_1, \dots, \alpha_k) \in \mathbb{R}^k$, on a

$$\alpha_1 g_1 + \dots + \alpha_k g_k = X \cdot u$$

où $X := (g_1, \dots, g_k)$ est un vecteur gaussien standard de \mathbb{R}^k d'après la proposition précédente. Mais alors on sait que $X \cdot u$ est une variable gaussienne standard, qui a donc la même loi que g_1 . \square

Remarque. — On peut rappeler que, plus généralement, un vecteur aléatoire G sur \mathbb{R}^n est dit gaussien s'il existe une application affine A sur \mathbb{R}^n ($A = x_0 + B$ avec $B \in \mathcal{L}(\mathbb{R}^n)$ et $x_0 \in \mathbb{R}^n$) et un vecteur gaussien standard G_0 tel que

$$G = AG_0 = \mathbb{E}G + BG_0 \tag{1}$$

Comme l'image d'un vecteur gaussien standard par une isométrie est encore un vecteur gaussien standard, si on écrit $B = |B|U$ avec $|B|$ symétrique positive et $U \in \mathcal{O}(n)$, peut écrire (1) sous la forme

$$G = \mathbb{E}G + |B|\tilde{G}_0$$

où $\tilde{G}_0 = UG_0$ est un vecteur gaussien standard. Soit e_1, \dots, e_n une base orthonormée de \mathbb{R}^n ; on identifie les applications linéaires et leur matrices dans cette base. Pour un vecteur gaussien donné par (1), on a, pour tout $i, j \leq n$,

$$\text{Cov}_{i,j}(G) := \mathbb{E}[(G - \mathbb{E}G) \cdot e_i][(G - \mathbb{E}G) \cdot e_j] = \mathbb{E}[(G_0 \cdot B^* e_i)(G_0 \cdot B^* e_j)] = B^* e_i \cdot B^* e_j.$$

Soit encore, pour la matrice dite de covariance $\text{Cov}(G) = (\text{Cov}_{i,j}(G))_{i,j}$:

$$\text{Cov}(G) = BB^* = |B|^2.$$

Ainsi $|B|$ est déterminé par la matrice de covariance ; on résume cela en disant : la loi d'un vecteur gaussien est déterminée par son espérance et sa matrice de covariance.

4.2. Inégalité de concentration gaussienne

Le théorème suivant est très important : il indique que si f est une fonction 1-lipschitzienne sur \mathbb{R}^n , et si G est un vecteur Gaussien standard, alors la variable $X = f(G) - \mathbb{E}f(G)$ est "presque" constante. Ce qui est surprenant est que l'estimée est indépendante de la dimension : les fonctions lipschitziennes "écrasent" les gaussiennes vers une dimension 1. Le résultat peut s'obtenir comme conséquence directe de l'inégalité de Prékopa-Leindler.

Théorème (Concentration gaussienne). — *Si f est une fonction lipschitzienne sur l'espace euclidien $(\mathbb{R}^n, |\cdot|)$, alors*

$$\forall r \geq 0, \quad \gamma_n \left(\left\{ x \in \mathbb{R}^n ; f(x) \geq \int f d\gamma_n + r \right\} \right) \leq e^{-r^2/(4\|f\|_{\text{lip}}^2)},$$

$$\text{et } \gamma_n \left(\left\{ x \in \mathbb{R}^n ; \left| f(x) - \int f d\gamma_n \right| \geq r \right\} \right) \leq 2e^{-r^2/(4\|f\|_{\text{lip}}^2)}.$$

En d'autres termes, on a

$$\|f - \int f d\gamma_n\|_{\psi_2(\gamma_n)} \leq c\|f\|_{\text{lip}},$$

pour une certaine constante numérique $c > 0$.

Démonstration. — On rappelle qu'une variante de l'inégalité de Prékopa-Leindler, énoncée pour $t = 1/2$ et $H = 1$, est : pour $F, G : \mathbb{R}^n \rightarrow \mathbb{R}^+$ on a tel que, pour tout $x, y \in \mathbb{R}^n$,

$$\left(\forall x, y \in \mathbb{R}^n, \quad 1 \geq e^{-|x-y|^2/4} F(x)G(y) \right) \implies \int F d\gamma_n \int G d\gamma_n \leq 1.$$

On choisit $F(x) = e^{\lambda f(x)}$ avec $\lambda > 0$ (on va donc suivre une méthode de Laplace). On remarque que la fonction $G(y) = e^{-\lambda f(y) - L^2 \lambda^2}$ va satisfaire la condition requise ci-dessus. En effet, pour $x, y \in \mathbb{R}^n$,

$$\lambda f(x) - \lambda f(y) \leq L\lambda|x-y| \leq L^2\lambda^2 + |x-y|^2/4.$$

On a donc

$$\int e^{\lambda f} d\gamma_n \int e^{-\lambda f} d\gamma_n \leq e^{L^2\lambda^2}.$$

Comme, par Jensen, on a $\int e^{-\lambda f} d\gamma_n \geq e^{-\lambda \int f d\gamma_n}$, on en tire

$$\int e^{\lambda[f - \int f d\gamma_n]} d\gamma_n \leq e^{L^2\lambda^2}.$$

D'où, pour $r \geq 0$,

$$\gamma_n \left(\left\{ x \in \mathbb{R}^n ; f(x) \geq \int f d\gamma_n + r \right\} \right) \leq e^{-\lambda r + L^2\lambda^2},$$

et le résultat s'ensuit en optimisant en $\lambda > 0$ (i.e. en prenant $\lambda = r/2L^2$). \square

Remarque. — *On peut montrer que l'inégalité de concentration pour les fonctions lipschitziennes est équivalente à la concentration suivante pour la fonction distance à un ensemble : pour tout ensemble $A \subset \mathbb{R}^n$, on a*

$$\int e^{d(x,A)^2/4} d\gamma_n(x) \leq \frac{1}{\gamma_n(A)}.$$

En particulier, pour tout $A \subset \mathbb{R}^n$ tel que $\gamma_n(A) \geq \frac{1}{2}$, on a

$$\gamma_n(\{x \in \mathbb{R}^n ; d(x, A) \geq r\}) \leq 2e^{-r^2/4}, \quad \forall r \geq 0.$$

Signalons au passage que les constantes numériques 2, 4, etc. ne sont pas les meilleures possibles (mais on s'en fiche pas mal).

Il est important de trouver la bonne manière de formuler le résultat en fonction du contexte. Par exemple, on a la forme suivante due à G. Pisier :

Corollaire. — Soit $(E, \|\cdot\|)$ un Banach et $a_1, \dots, a_m \in E$. On considère le vecteur aléatoire

$$X = \sum_{i=1}^m g_i a_i$$

où les g_i sont des variables aléatoires gaussiennes standard indépendantes. Alors on a

$$\forall t \geq 0, \quad \mathbb{P}(\{ \|\|X\| - \mathbb{E}\|X\| \mid \geq t \}) \leq 2e^{-t^2/(4\sigma_X^2)}$$

où

$$\sigma_X = \sup_{\xi \in E^*, \|\xi\|_{E^*} \leq 1} \left(\sum_{i=1}^m |\xi(a_i)|^2 \right)^{1/2}.$$

Démonstration. — Par hypothèse,

$$G := (g_1, \dots, g_m)$$

un vecteur gaussien standard de $\mathbb{R}^m = \ell_2^m$ (distribué donc suivant γ_m), et si e_1, \dots, e_m désigne la base canonique de \mathbb{R}^m , alors on peut écrire $X = \sum_{i=1}^m (G \cdot e_i) a_i$. Introduisons la fonction $f : \mathbb{R}^m \rightarrow \mathbb{R}$ définie par

$$f(y) = \left\| \sum_{i=1}^m (y \cdot e_i) a_i \right\|.$$

Alors, en utilisant la concentration gaussienne on a, pour tout $t > 0$,

$$\mathbb{P}(\{ \|\|X\| - \mathbb{E}\|X\| \mid \geq t \}) = \gamma_m \left(\left\{ y \in \mathbb{R}^m ; \left| f(y) - \int f d\gamma_m \right| \geq t \right\} \right) \leq 2e^{-t^2/(4\|f\|_{\text{lip}}^2)}.$$

Reste donc à identifier $\|f\|_{\text{lip}}$. Introduisons l'opérateur

$$\begin{aligned} u : \ell_2^m &\longrightarrow E \\ e_i &\longrightarrow a_i \end{aligned}$$

de sorte que $f(y) = \|u(y)\|$. Ainsi,

$$\|f\|_{\text{lip}} = \sup_{y \neq y' \in \ell_2^m} \frac{|\|u(y)\| - \|u(y')\||}{|y - y'|} = \sup_{y \in \ell_2^m} \frac{\|u(y)\|}{|y|} =: \|u\|_{\ell_2^m \rightarrow E}.$$

En utilisant que $\|x\| = \sup_{\|\xi\|_{E^*} \leq 1} \xi(x)$ pour $x \in E$ et en permutant les sup (qui sont en fait des max) on a :

$$\|u\|_{\ell_2^m \rightarrow E} = \sup_{|y|=1} \sup_{\|\xi\|_{E^*} \leq 1} \xi(u(y)) = \sup_{\|\xi\|_{E^*} \leq 1} \sup_{|y|=1} \sum_{i=1}^m (y \cdot e_i) \xi(a_i) = \sigma_X.$$

□

Remarque. — Notez qu'on a, dans le résultat précédent,

$$\sigma_X \leq \left(\sum_{i=1}^m \|a_i\|^2 \right)^{1/2}. \quad (2)$$

De fait, dans la preuve précédente, une borne moins bonne et triviale sur la constante de lipschitz de f (i.e. de la norme d'opérateur de u) aurait été

$$f(y) \leq \sum_{i=1}^m |y \cdot e_i| \|a_i\| \leq \left(\sum_{i=1}^m \|a_i\|^2 \right)^{1/2} |y|.$$

Comparer, par exemple, les termes de (2) dans le cas où $E = \ell_2^m$ et $a_i = e_i$ pour $i = 1, \dots, m$ (où plus généralement dans le cas où u est un opérateur diagonal : $a_i = \alpha_i e_i$, $\alpha_i \in \mathbb{R}$).

4.3. Opérateurs et matrices aléatoires gaussiennes

Soit G une matrice aléatoire gaussienne $N \times n$. Plus précisément, on se donne $n \times N$ variables aléatoires indépendantes $\{g_{i,j}\}_{i \leq N, j \leq n}$ gaussiennes standard (définies sur un espace de probabilité (Ω, \mathbb{P})), et on considère la matrice aléatoire

$$G = (g_{i,j})_{i \leq N, j \leq n}.$$

Si besoin, on notera $\omega \rightarrow g_{i,j}(\omega) \in \mathbb{R}$ les variables aléatoires $g_{i,j}$ et $G_\omega = (g_{i,j}(\omega))_{i \leq N, j \leq n}$ la matrice correspondante. On peut dire que G est un vecteur aléatoire de \mathbb{R}^{nN} mais on préfère voir algébriquement G comme une application linéaire aléatoire de \mathbb{R}^n dans \mathbb{R}^N (en prenant les bases canoniques de ces espaces). En s'intéressant aux lignes et aux colonnes de G ,

$$G = [G_1, \dots, G_n] = \begin{bmatrix} \tilde{G}_1 \\ \vdots \\ \tilde{G}_N \end{bmatrix},$$

on a que G_1, \dots, G_n sont des vecteurs gaussiens standards indépendants de \mathbb{R}^N , et que $\tilde{G}_1, \dots, \tilde{G}_N$ sont des vecteurs gaussiens standards indépendants de \mathbb{R}^n . On peut écrire, pour $x \in \mathbb{R}^n$,

$$Gx = \sum_{i=1}^n x_i G_i \in \mathbb{R}^N.$$

Si on met sur \mathbb{R}^n la structure euclidienne usuelle \cdot , alors on voit alors que, pour $x \in \mathbb{R}^n$ non nul,

$$\frac{1}{|x|} Gx$$

est encore un vecteur gaussien standard de \mathbb{R}^N . De plus, pour $x, y \in \mathbb{R}^n$, on a

$$\mathbb{E}[Gx \cdot Gy] = \sum_{i,k \leq n} x_i y_k \mathbb{E}[G_i \cdot G_k] = N x \cdot y,$$

et en particulier

$$\mathbb{E}|Gx|^2 = N|x|^2.$$

On peut dire que $x \rightarrow \frac{1}{\sqrt{N}}Gx$ est un plongement isométrique de ℓ_2^n dans $L_2(\Omega, \mathbb{P}; \ell_2^N)$. Prenons sur \mathbb{R}^N une certaine norme $\|\cdot\|$ (par nécessairement $|\cdot|$) et posons $E = (\mathbb{R}^n, \|\cdot\|)$. Alors, on peut voir G comme un opérateur (aléatoire) de ℓ_2^n dans E ,

$$G : \ell_2^n \longrightarrow E.$$

Remarque. — La norme de cet opérateur est la variable aléatoire

$$\|G\|_{\ell_2^n \rightarrow E} = \sup_{|x| \leq 1} \|Gx\| = \sup_{|x| \leq 1} \sup_{\|y\|_{E^*} \leq 1} Gx \cdot y.$$

Lorsque $E = \ell_2^N$, la norme $\|G\|_{\ell_2^n \rightarrow \ell_2^N}$ est aussi la plus grande valeur propre de la matrice $({}^tGG)^{1/2}$.

On peut aussi remarquer qu'une matrice gaussienne est toujours de rang maximal, à savoir $\min\{n, N\}$, au sens où

$$\mathbb{P}(\{G \text{ de rang maximal}\}) = 1.$$

Dans la suite, on va faire en sorte que G agisse presque comme une isométrie.

Le lemme suivant, qui est une application directe de la concentration gaussienne, va se révéler être très utile.

Lemme. — On se donne une matrice G aléatoire gaussienne $N \times n$ que l'on voit comme une application aléatoire de \mathbb{R}^n dans \mathbb{R}^N . Étant donné $\|\cdot\|$ une norme sur \mathbb{R}^n et $E := (\mathbb{R}^n, \|\cdot\|)$, on introduit les deux paramètres suivants :

$$m_E := \int_{\mathbb{R}^N} \|x\| d\gamma_N(x) = \mathbb{E}\|Gv\|, \quad \forall v \in \mathbb{R}^k, \quad \text{avec } |v| = 1,$$

et $b_E > 0$ tel que

$$\|\cdot\| \leq b_E |\cdot| \quad \text{sur } \mathbb{R}^n.$$

Alors, pour tout $y \in \mathbb{R}^n$ fixé, on a

$$(1 - \varepsilon)m |y| \leq \|Gy\| \leq (1 + \varepsilon)m |y|$$

avec une probabilité au moins $1 - 2e^{-m^2\varepsilon^2/(4b^2)}$.

Plus généralement, si $\mathcal{S} \subset \mathbb{R}^n$ est un ensemble fini de vecteurs, on a

$$\mathbb{P}(\{\forall y \in \mathcal{S}; (1 - \varepsilon)m |y| \leq \|Gy\| \leq (1 + \varepsilon)m |y|\}) \geq 1 - 2|\mathcal{S}|e^{-m^2\varepsilon^2/(4b^2)}.$$

Démonstration. — On note $m = m_E$ et $b = b_E$. Soit $y \in \mathbb{R}^n$ fixé, $y \neq 0$. Alors $\frac{1}{|y|}Gy$ est un vecteur de loi γ_N , et donc

$$\begin{aligned} \mathbb{P}(\{(1 - \varepsilon)m \leq \frac{1}{|y|}\|Gy\| \leq (1 + \varepsilon)m\}) &= \mathbb{P}(\{|\frac{1}{|y|}\|Gy\| - m| \leq \varepsilon m\}) \\ &= \gamma_N(\{x \in \mathbb{R}^N; \||x| - m| \leq \varepsilon m\}) \\ &\geq 1 - 2e^{-m^2\varepsilon^2/(4b^2)} \end{aligned}$$

puisque la fonction $x \rightarrow \|x\|$ est lipschitzienne de constante b sur $(\mathbb{R}^N, |\cdot|)$.

Si on note, pour chaque $y \in \mathbb{R}^n$, A_y l'événement

$$A_y = \{(1 - \varepsilon)m |y| \leq \|Gy\| \leq (1 + \varepsilon)m |y|\} = \{\omega \in \Omega; (1 - \varepsilon)m |y| \leq \|G_\omega y\| \leq (1 + \varepsilon)m |y|\}$$

et A_y^c l'événement complémentaire, alors on a

$$\begin{aligned} & \mathbb{P}(\{\forall y \in \mathcal{S}; (1 - \varepsilon)m|y| \leq \|Gy\| \leq (1 + \varepsilon)m|y|\}) \\ &= \mathbb{P}\left(\bigcap_{y \in \mathcal{S}} A_y\right) = 1 - \mathbb{P}\left(\bigcup_{y \in \mathcal{S}} A_y^c\right) \geq 1 - \sum_{y \in \mathcal{S}} \mathbb{P}(A_y^c) \geq 1 - 2|\mathcal{S}|e^{-m^2\varepsilon^2/(4b^2)}. \end{aligned}$$

□

4.4. Le lemme « d'aplatissement » de Johnson-Lindenstrauss

Soit (X, d) et (Y, δ) deux espaces métriques. On cherche à réaliser X dans Y de la manière la plus fidèle possible, métriquement parlant.

Definition (D -plongement). — On dit qu'une application $F : X \rightarrow Y$ est un plongement (bi-lipshitzien) si F est lipschitzienne injective et $F_{|F(X)}^{-1}$ est lipschitzienne sur $F(X)$. On dit que F est un D -plongement (ou un plongement de distorsion plus petit que D) si F est un plongement tel que $\|F\|_{\text{lip}} \times \|F_{|F(X)}^{-1}\|_{\text{lip}} \leq D$, c'est-à-dire si

$$\sup_{x \neq y \in X} \frac{\delta(F(x), F(y))}{d(x, y)} \times \sup_{x \neq y \in X} \frac{d(x, y)}{\delta(F(x), F(y))} \leq D$$

En pratique, on utilise plutôt la forme trivialement équivalente suivante :

Fait. — Une application $F : (X, d) \rightarrow (Y, \delta)$ est un D -plongement si, et seulement si, il existe un $r > 0$ tel que

$$\forall x, y \in X, \quad r d(x, y) \leq \delta(F(x), F(y)) \leq rD d(x, y).$$

On cherche donc un D -plongement avec D le plus petit possible (notez que D est toujours plus grand que 1). On dit qu'une application F entre les espaces métriques X et Y est une λ -isométrie ($\lambda > 0$) si $\forall x, y \in X, \delta(F(x), F(y)) = \lambda d(x, y)$. Dans ce cas, $F(X)$ est une « copie » (métriquement parlant) de X dans Y . Notez que $F : X \rightarrow Y$ est un 1-plongement si et seulement si il existe $r > 0$ tel que F est une r -isométrie.

Dans la suite on s'intéresse au cas où Y est un espace vectoriel normé $(E, \|\cdot\|)$ (et donc $\delta(z, z') = \|z - z'\|$). On se donne un espace métrique (X, d) et on cherche à mettre X dans E en déformant X le moins possible. On sait qu'une application $F : X \rightarrow E$ est un 1-plongement exactement si F est une r -isométrie pour un certain $r > 0$, et dans le cas présent cela veut exactement dire que F est un multiple d'une isométrie. En général, on ne peut pas réaliser isométriquement X dans E , mais on cherche à le réaliser avec une distorsion la plus petite possible (c'est-à-dire la plus proche de 1 possible). Une question naturelle est : peut-on trouver un $(1 + \varepsilon)$ -plongement de X dans E pour tout $\varepsilon > 0$?

Retenons, pour la suite, que dire qu'il existe $A, B > 0$ tel que

$$\forall x, y \in X, \quad A d(x, y) \leq \|F(x) - F(y)\| \leq B d(x, y).$$

c'est exactement dire que F est un $\frac{B}{A}$ -plongement de X dans E .

Pour ce genre de problèmes purement métriques, on s'intéresse souvent au cas où X est un espace métrique fini, de cardinal $|X| = n \geq 1$, avec, en général, n grand (tendant vers $+\infty$).

Supposons que l'on se donne n points dans un espace euclidien (muni bien-sûr de la métrique euclidienne). Evidemment, on peut supposer que cet espace euclidien est de dimension $\leq n$, et donc qu'on s'est donné n points dans ℓ_2^n , c'est-à-dire un espace métrique $X \subset \ell_2^n$ à n points

avec la métrique induite par la structure euclidienne. On aimerait pouvoir mettre ces points dans un espace euclidien de dimension N avec

$$N \ll n$$

sans perdre trop. Cela a son importance pour des questions de codage. Le résultat facile suivant – qui est important en théorie de l'apprentissage et de la compression de données – donne une réponse (asymptotiquement surprenante) à cette question :

Théorème (Lemme « d'aplatissement » de Johnson-Lindenstrauss)

Il existe une constante numérique $c > 0$ tel que : Pour tout $\varepsilon \in]0, 1[$, tout $n \geq 2$ et tout ensemble X de n points d'un espace euclidien ($X \subset \ell_2^n$ avec la métrique euclidienne induite), il existe un $(1 + \varepsilon)$ -plongement de X dans ℓ_2^N avec

$$N = c \frac{\log(n)}{\varepsilon^2}$$

(ce plongement étant de plus la restriction à X d'une application linéaire de ℓ_2^n dans ℓ_2^N).

Démonstration. — Soit G une matrice aléatoire gaussienne $N \times n$ comme décrite précédemment, avec $N \leq n$ qui sera fixé plus loin. Dans l'étude qui va suivre, on verra G comme un application (linéaire) de ℓ_2^n dans ℓ_2^N . On va donc appliquer le Lemme 4.3 avec la norme euclidienne $\|\cdot\| = |\cdot|$ sur \mathbb{R}^N (et donc avec $b = 1$). On note donc

$$m := \int_{\mathbb{R}^N} |x| d\gamma_N(x) = \mathbb{E}|Gv| \quad \text{si } |v| = 1,$$

et on rappelle que

$$\frac{\sqrt{N}}{4} \leq m \leq \sqrt{N}.$$

On choisit comme ensemble fini $\mathcal{S} \subset \mathbb{R}^n$ l'ensemble

$$\mathcal{S} := \{x - y ; (x, y) \in X \times X\},$$

dont le cardinal vérifie

$$|\mathcal{S}| \leq n^2.$$

Le Lemme 4.3 appliqué avec $\frac{\varepsilon}{3}$ nous donne donc que

$$\mathbb{P}(\{\forall x, y \in X ; (1 - \frac{\varepsilon}{3})m|x - y| \leq |Gx - Gy| \leq (1 + \frac{\varepsilon}{3})m|x - y|\}) \geq 1 - 2n^2 e^{-m^2 \varepsilon^2 / 36}.$$

En utilisant que $m \geq \frac{\sqrt{N}}{4}$ on a donc

$$\begin{aligned} \mathbb{P}(\{\forall x, y \in X ; (1 - \frac{\varepsilon}{3})m|x - y| \leq |Gx - Gy| \leq (1 + \frac{\varepsilon}{3})m|x - y|\}) &\geq 1 - 2n^2 e^{-N\varepsilon^2 / 576} \\ &\geq 1 - n^3 e^{-N\varepsilon^2 / 576} \\ &\geq 1 - \frac{1}{2} > 0 \end{aligned}$$

en choisissant une fois pour toute le N suivant

$$N = \frac{(4 \cdot 576) \log(n)}{\varepsilon^2}.$$

Mais alors, on a en particulier, pour ce choix de N , que

$$\mathbb{P}(\{\forall x, y \in X ; (1 - \frac{\varepsilon}{3})m|x - y| \leq |Gx - Gy| \leq (1 + \frac{\varepsilon}{3})m|x - y|\}) \neq 0.$$

Par conséquent, l'événement considéré, qui est de probabilité non-nulle, doit être non vide. Il existe donc $\omega_0 \in \Omega$ tel que

$$\forall x, y \in X ; \left(1 - \frac{\varepsilon}{3}\right)m |x - y| \leq |G_{\omega_0}x - G_{\omega_0}y| \leq \left(1 + \frac{\varepsilon}{3}\right)m |x - y|.$$

Soit encore, en posant $F := G_{\omega_0} : \ell_2^n \rightarrow \ell_2^N$,

$$\forall x, y \in X, \quad m\left(1 - \frac{\varepsilon}{3}\right)|x - y| \leq |F(x) - F(y)| \leq m\left(1 + \frac{\varepsilon}{3}\right)|x - y|,$$

ce qui montre que F réalise un plongement de X dans ℓ_2^N de distorsion majorée par

$$\frac{1 + \varepsilon/3}{1 - \varepsilon/3} \leq 1 + \varepsilon.$$

□

4.5. Théorème de Dvoretzky : énoncés

On peut montrer que les espaces ℓ_∞^n et ℓ_2^n sont « loin » l'un de l'autre, à une distance de l'ordre de \sqrt{n} (avec le vocabulaire de la section précédente, toute application linéaire de $u : \ell_2^n \rightarrow \ell_\infty^n$ aura une distorsion au moins égale à \sqrt{n}). De manière étonnante, on peut plonger presque isométriquement ℓ_2^k dans ℓ_∞^n pour des k assez grands (de l'ordre de $\log(n)$).

Ce phénomène a plus généralement lieu pour tout espace vectoriel normé de dimension n (que l'on va identifier à \mathbb{R}^n muni d'une certaine norme, ou à un corps convexe symétrique).

Definition (Constante de Dvoretzky-Milman). — Soit $E = (\mathbb{R}^n, \|\cdot\|)$ un evn de dimension n . On pose alors

$$k(E) := \left(\frac{M}{b}\right)^2$$

où b est la (meilleure) constante telle que

$$\|\cdot\| \leq b|\cdot|$$

et

$$M = M(E) := \int_{\mathbb{R}^n} \|x\| d\gamma_n(x) = \mathbb{E}\|X\| \quad \text{pour } X \text{ gaussien standard de } \mathbb{R}^n.$$

Si K est un corps convexe symétrique de \mathbb{R}^n on pose

$$k(K) := k((\mathbb{R}^n, \|\cdot\|_K)).$$

Notez que $k(E) \leq k(\ell_2^n) \leq n$. Dans la suite, on voudra que $k(E)$ soit grand. On verra qu'on peut s'arranger pour que $k(E)$ soit au moins de l'ordre $\log(n)$. Dans certains cas, on aura même beaucoup mieux.

Remarque (Moyenne sphérique versus moyenne gaussienne)

Dans certains textes, on définit la moyenne de la norme $\|\cdot\|$ sur la sphère S^{n-1} (muni de la probabilité usuelle σ) plutôt que par rapport à la mesure gaussienne :

$$\tilde{M}(E) := \int_{S^{n-1}} \|x\| d\sigma(x).$$

Comme la norme est 1-homogène, on vérifie facilement que

$$\tilde{M}(E) = \int_{S^{n-1}} \|x\| d\sigma(x) = \frac{1}{c_1} \int_{\mathbb{R}^n} \|x\| d\gamma_n(x) = \frac{1}{c_1} M(E)$$

avec $c_1 := \int_{\mathbb{R}^n} |x| d\gamma_n(x) \simeq \sqrt{n}$. Plus précisément, rappelons que $\frac{1}{3}\sqrt{n} \leq c_1 \leq \sqrt{n}$. C'est pourquoi la constante de Dvoretzky-Milman est parfois définie par

$$\tilde{k}(E) := n \left(\frac{\tilde{M}}{b} \right)^2.$$

Cette constante ne diffère de $k(E)$ qu'à une constante numérique près (ce qui pour nous n'a pas d'importance).

Théorème (Théorème de Dvoretzky : version plongement)

Pour tout $\varepsilon \in]0, 1[$, en notant $\eta(\varepsilon) = \frac{\varepsilon^2}{\log(1+\frac{1}{\varepsilon})} > 0$, pour tout $n \geq 2$ et pour tout evn $E = (\mathbb{R}^n, \|\cdot\|)$, il existe une application linéaire $A : \mathbb{R}^k \rightarrow \mathbb{R}^n$ avec

$$k \geq c \eta(\varepsilon) k(E)$$

tel que, pour un certain $m > 0$,

$$\forall y \in \mathbb{R}^k, \quad (1 - \varepsilon)m |y| \leq \|Ay\| \leq (1 + \varepsilon)m |y|. \quad (3)$$

En d'autres termes, pour tout $\varepsilon > 0$ on peut trouver un $(1 + \varepsilon)$ -plongement linéaire de ℓ_2^k dans E pour $k = c\eta(\varepsilon/3)k(E)$.

Rappel (on a maintenant l'habitude) : dans l'énoncé ci-dessus, $c > 0$ est une constante numérique – que l'on pourrait calculer –, c'est-à-dire que l'on devrait écrire : « Il existe une constante numérique $c > 0$ tel que $\forall \varepsilon \in]0, 1[$, $\forall n \geq 2$, $\forall E$, ... ».

Remarque (Important : l'hypothèse cachée). — Dans de certains textes, on trouve une hypothèse sur E , à savoir que $c\eta(\varepsilon)k(E) \geq 1$ (ou ≥ 4 ou n'importe quelle constante). En fait, c'est implicite dans notre énoncé. En effet, si jamais $c\eta(\varepsilon)k(E) \leq 1$, alors l'énoncé est vide, car en dimension 1, toutes les normes sont euclidiennes!!! (On pourrait prendre, pour $x_0 \in \mathbb{R}^n$ fixé, $A(a) = ax_0$, $\forall a \in \mathbb{R}$). Donc, il y a toujours une hypothèse cachée sur la dimension ou sur $k(E)$ pour que $k \geq 2$. On est pas obligé de l'écrire, car le résultat reste vrai sinon, mais il est vide.

On verra plus tard qu'on peut s'arranger pour que $k(E) \geq C \log(n)$. Par conséquent, l'hypothèse $c\eta(\varepsilon)k(E) \geq 1$ pourra être remplacée par une hypothèse portant seulement sur la dimension n (l'énoncé sera alors valable pour tout E de dim n) : $n \geq n(\varepsilon) = \frac{c\eta(\varepsilon)}{\log(n)}$. Si on choisit de prendre $\varepsilon = \frac{1}{2}$, cette hypothèse revient à dire un truc du genre $n \geq 18$.

Remarque ($\exists k \geq c\eta(\varepsilon)k(E)$... c'est dire qu'on peut choisir $k = c\eta(\varepsilon)k(E)$)

Comme $\ell_2^{k'} \subset \ell_2^k$ isométriquement pour $k' \leq k$, chaque fois qu'on aura un énoncé avec un k , il sera vrai pour les dimensions k inférieures. Évidemment, on veut k le plus grand possible. L'expression « $\exists k \geq c\eta(\varepsilon)k(E)$... » ne nous donne pas plus d'information que « pour $k = c\eta(\varepsilon)k(E)$ ». En fait l'énoncé équivalent est « $\forall k \leq c\eta(\varepsilon)k(E)$ on a... ».

Le théorème précédent nous dit que sur le sous-espace $F_k = \{Ay ; y \in \ell_2^k\} \subset \mathbb{R}^n$ qui est k -dimensionnel (pourquoi?) on a, en notant $|x|_A := |A^{-1}x|$ la norme euclidienne (mais pas standard) induite par A sur F_k ,

$$\forall x \in F_k, \quad (1 - \varepsilon)m |x|_A \leq \|x\| \leq (1 + \varepsilon)m |x|_A.$$

En fait, en travaillant un tout petit peu plus, on peut montrer qu'on peut en fait avoir un encadrement du même type avec la norme euclidienne standard.

Théorème (Théorème de Dvoretzky : version section). — Pour tout $\varepsilon \in]0, 1[$, tout $n \geq 2$ et pour tout evn $E = (\mathbb{R}^n, \|\cdot\|)$, il existe un sous-espace vectoriel $F \subset \mathbb{R}^n$ de dimension

$$k = c\eta(\varepsilon)k(E)$$

tel que, pour un certain $m > 0$

$$\forall x \in F, \quad (1 - \varepsilon)m|x| \leq \|x\| \leq (1 + \varepsilon)m|x|. \quad (4)$$

De manière équivalente, pour tout $\varepsilon \in [0, 1]$ et tout $n \geq 2$, et K tout corps convexe symétrique de \mathbb{R}^n (avec $c\eta(\varepsilon)k(E) \geq 4$), on peut trouver un sous-espace $F \subset \mathbb{R}^n$ de dimension $k = c\eta(\varepsilon)k(K)$ tel que

$$(1 - \varepsilon)mB_2^n \cap F \subset K \cap F \subset (1 + \varepsilon)mB_2^n \cap F. \quad (5)$$

Pour que ces résultats soient intéressants, il faut que $k(E)$ soit grand (on veut qu'il tende vers $+\infty$ lorsque $n \rightarrow +\infty$). Regardons le cas des boules ℓ_∞ et ℓ_1 . D'après les calculs fait précédemment, on a :

$$k(\ell_\infty^n) = k(B_\infty^n) \geq C \log(n).$$

Par conséquent, on en déduit le

Théorème (Théorème de Dvoretzky pour le cube). — Pour tout $\varepsilon \in]0, 1[$, tout $n \geq n(\varepsilon)$ (constante dépendant seulement de ε pour que le résultat ne soit pas vide), il existe un sous espace vectoriel $F \subset \mathbb{R}^n$ de dimension

$$k = c\eta(\varepsilon) \log(n)$$

tel que, pour un certain $m > 0$

$$(1 - \varepsilon)mB_2^n \cap F \subset B_\infty^k \cap F \subset (1 + \varepsilon)mB_2^n \cap F. \quad (6)$$

Le résultat est surprenant même en prenant un ε fixé, par exemple : $\varepsilon = \frac{1}{2}$.

Le résultat est encore plus surprenant pour B_1^n . En effet, on voit facilement qu'il existe une constante numérique $C > 0$ tel que

$$k(\ell_1^n) = k(B_1^n) \geq Cn.$$

Par conséquent, on peut trouver une section F de dimension k proportionnelle à n (!!!)

$$k = c\eta(\varepsilon)n$$

tel que

$$(1 - \varepsilon)mB_2^n \cap F \subset B_1^k \cap F \subset (1 + \varepsilon)mB_2^n \cap F. \quad (7)$$

Étudions le cas d'un corps convexe K général. Une position de K est une image TK de K par une application linéaire inversible $T \in GL(\mathbb{R}^n)$. On dit que K est en position de John si $B_2^n \subset K \subset$ et si B_2^n est de volume maximal parmi tous les ellipsoïdes inclus dans K . On montre (voir exercice ci-dessous) que pour tout corps convexe K , il existe une position TK tel que TK soit en position de John.

Théorème (Conséquence de Dvoretzky-Rogers). — Il existe une constante numérique $c > 0$ tel que pour tout $n \geq 2$, si K est un corps convexe symétrique de \mathbb{R}^n en position de John, alors

$$k(K) \geq c \log(n).$$

On peut donc reformuler tous les énoncés précédent en supposant que K (ou E) est en position de John, auquel cas on a pas d'hypothèse à faire sur $k(K)$ et on trouve une section F de dimension k (ou un plongement de ℓ_2^k) avec

$$k \geq c\eta(\varepsilon) \log(n)$$

tel que $(1 - \varepsilon)mB_2^n \cap F \subset K \cap F \subset (1 + \varepsilon)mB_2^n \cap F$. En d'autre termes, lorsque K est en position de John, on a la même résultat que pour le cube B_∞^n .

Si on préfère ne pas mettre le corps en position de John, alors on peut énoncer le résultat comme suit :

Théorème (Théorème de Dvoretzky : version section/position)

Pour tout $\varepsilon \in [0, 1]$ et tout $n \geq n(\varepsilon)$, et tout corps convexe symétrique K de \mathbb{R}^n , on peut trouver un sous-espace $F \subset \mathbb{R}^n$ de dimension $k = c\eta(\varepsilon) \log(n)$ et une application linéaire $T \in GL(\mathbb{R}^n)$ tel que

$$(1 - \varepsilon)mB_2^n \cap F \subset TK \cap F \subset (1 + \varepsilon)mB_2^n \cap F. \quad (8)$$

Cette dernière version (ou son pendant plongement) peut s'exprimer en terme de distance de Banach-Mazur. La distance de Banach-Masure entre E et F , deux espaces de Banach (isomorphes, c'est-à-dire, pour nous, de même dimension) :

$$d(E, F) = \inf\{\|T\|_{E \rightarrow F} \|T^{-1}\|_{F \rightarrow E}; T : E \rightarrow F \text{ isomorphisme}\}.$$

On définit de même la distance de Banach-Mazur entre deux corps convexes symétriques K et L de \mathbb{R}^n comme la distance des evn $(\mathbb{R}^n, \|\cdot\|_K)$ et $(\mathbb{R}^n, \|\cdot\|_L)$, soit encore (exo) :

$$d(K, L) = \inf\left\{\frac{b}{a}; aK \subset TL \subset bK, T \in GL(\mathbb{R}^n)\right\}.$$

On a donc aussi

$$d(K, L) = \inf_{T \in GL(\mathbb{R}^n)} d_g(K, TL)$$

où d_g désigne la distance géométrique définie au chapitre précédent (exercice 1.1). On a donc,

Théorème (Théorème de Dvoretzky : version distance de Banach-Mazur)

Pour tout $\varepsilon \in [0, 1]$, tout $n \geq n(\varepsilon)$ et tout evn $(E, \|\cdot\|_E)$ de dimension n , on peut trouver un sous-espace vectoriel $F \subset E$ de dimension $k = c\eta(\varepsilon) \log(n)$ tel que l'evn $F_E = (F, \|\cdot\|_E)$ vérifie

$$d(\ell_2^k, F_E) \leq 1 + \varepsilon \quad (9)$$

Remarque. — En général, lorsqu'on restreint une structure métrique (une norme) à un sous-espace, ou que l'on regarde $K \cap F$ pour un corps convexe symétrique K , on obtient quelque chose de différent de ce que l'on avait au départ. Si on prend un sous-espace F de dimension $k \leq n$, alors l'espace vectoriel normé obtenu en prenant sur F la (restriction de la) norme $\|\cdot\|_\infty$ de \mathbb{R}^n , ne ressemble pas, en général, à ℓ_∞^k . On peut aussi dire que le corps convexe symétrique de F

$$B_\infty^n \cap F$$

ne ressemble pas en général à un cube dans F . Le théorème de Dvoretzky va beaucoup plus loin. Il nous dit que qu'on peut trouver un sous-espace $F \subset \ell_\infty^n$ de dimension k assez grande (de l'ordre de $\log(n)$) tel que cet espace, muni de la norme $\|\cdot\|_\infty$ vérifie

$$d(\ell_2^k, F) \simeq 1 + \varepsilon$$

De point de vue géométrique, le théorème de Dvoretzky nous dit que $B_\infty^n \cap F$ est très proche d'une boule euclidienne, alors que B_∞^n ne l'était pas du tout.

4.6. Théorème de Dvoretzky : preuves

On va démontrer les deux premières versions. Le cas de la position de John sera faite plus loin en exercice.

Tout repose sur le lemme suivant :

Lemme (Lemme de Milman). — Pour tout $\varepsilon \in]0, 1[$, en notant $\eta(\varepsilon) = \frac{\varepsilon^2}{\log \frac{1}{\varepsilon}} > 0$, pour tout $n \geq 2$ et pour tout evn $E = (\mathbb{R}^n, \|\cdot\|)$, on a que, pour G matrice aléatoire gaussienne $n \times k$ avec $k \leq c\eta(\varepsilon)k(E)$ (et donc pour $k = c\eta(\varepsilon)k(E)$), il existe $m > 0$ tel que

$$\mathbb{P}\left(\left\{\forall y \in \mathbb{R}^k ; (1 - \varepsilon)m|y| \leq \|Gy\| \leq (1 + \varepsilon)m|y|\right\}\right) \geq 1 - e^{-2k} \quad (10)$$

Ce Lemme va nous permettre de construire des plongements et des section aléatoires car il nous garanti une probabilité non nulle pour qu'une matrice aléatoire gaussienne réalise ce que l'on veut. Mais en fait, ce résultat nous donne plus : en général presque toutes les matrices aléatoires vont être des $(1 + \varepsilon)$ -plongements de ℓ_2^k dans E . En effet, puisqu'on a

$$\mathbb{P}\left(\left\{\forall y \in \mathbb{R}^k ; (1 - \varepsilon)m|y| \leq \|Gy\| \leq (1 + \varepsilon)m|y|\right\}\right) \geq 1 - e^{-2c\eta(\varepsilon)k(E)},$$

dès que $k(E)$ est grand, cette proba est proche de 1. Si on est en position de John, on pourra garantir que $k(E) \geq c \log(n)$ et donc que la proba tend vers 1 quand la dimension n tend vers l'infini. Les cas les plus spectaculaires sont les cas $E = \ell_2^n$ ou $E = \ell_1^n$ pour lesquels on a $k(E) \geq cn$ et une convergence exponentielle de la proba. Notons en effet que $k(\ell_2^n) \geq \frac{1}{9}n$ et que si la question des plongements de ℓ_2^k ($k \leq n$) dans ℓ_2^n ou des sections euclidiennes de B_2^n n'a aucun intérêt (!!!), le résultat précédent, lui, donne une information intéressante sur le comportement $\ell_2 - \ell_2$ des matrices gaussiennes. Le fait que G soit en grand probabilité un $(1 + \varepsilon)$ -plongement de $\ell_2^{c(\varepsilon)n}$ dans ℓ_2^n est utilisé dans l'étude de l'invertibilité des matrices aléatoires gaussiennes.

Démonstration du Lemme. — On se donne un $k \leq n$ qui sera fixé plus loin et une matrice aléatoire $n \times k$ gaussienne $G = (g_{i,j})_{i \leq n, j \leq k}$, les $g_{i,j}$ étant donc des variables aléatoire gaussiennes standard indépendantes (définies sur un (Ω, \mathbb{P})). On rappelle que $S^{k-1} = \{y \in \mathbb{R}^k ; |y| = 1\}$. Par linéarité et homogénéité, il suffit de (et il faut) montrer que

$$\mathbb{P}\left(\left\{\forall y \in S^{k-1} ; (1 - \varepsilon)m \leq \|Gy\| \leq (1 + \varepsilon)m\right\}\right) \geq 1 - e^{-2k}$$

Attention, par rapport aux conventions du Lemme fondamental, on travaille ici avec $\ll n = k, N = n \gg$. On voit donc G comme une application aléatoire

$$G : \mathbb{R}^k \longrightarrow \mathbb{R}^n,$$

Le Lemme fondamental 4.3 (appliqué avec $\frac{\varepsilon}{3}$) nous dit que si $\mathcal{S} \subset S^{k-1}$ est un ensemble fini de vecteurs de norme 1, alors on a

$$\mathbb{P}\left(\left\{\forall y \in \mathcal{S} ; (1 - \frac{\varepsilon}{3})m \leq \|Gy\| \leq (1 + \frac{\varepsilon}{3})m\right\}\right) \geq 1 - 2|\mathcal{S}|e^{-m^2\varepsilon^2/(3^2 \cdot 4)}. \quad (11)$$

L'objectif est maintenant de trouver un ensemble $\mathcal{S} \subset S^{k-1}$ de cardinal le moins grand possible permettant de passer d'une inégalité valable pour tout les vecteurs de \mathcal{S} à une inégalité valable pour tous les vecteurs de S^{k-1} . Cela se fait avec un argument de réseau.

Definition. — Pour $\delta > 0$, on dit qu'une partie finie $\mathcal{S} \subset S^{k-1}$ est un δ -réseau de S^{k-1} si pour tout $v \in S^{k-1}$ il existe un $y \in \mathcal{S}$ tel que $|y - v| \leq \delta$.

L'utilité des ε -réseau repose sur le fait très facile suivant :

Fait. — On se donne $\varepsilon \in]0, 1[$, $m > 0$, $A : \mathbb{R}^k \rightarrow \mathbb{R}^n$ une application linéaire et $\|\cdot\|$ une norme sur \mathbb{R}^n . On suppose que \mathcal{S} est un $\frac{\varepsilon}{3}$ -réseau de S^{k-1} pour lequel on a

$$\forall y \in \mathcal{S}, \quad \left(1 - \frac{\varepsilon}{3}\right)m \leq \|Ay\| \leq \left(1 + \frac{\varepsilon}{3}\right)m \quad (12)$$

Alors, on a :

$$\forall v \in S^{k-1}, \quad (1 - \varepsilon)m \leq \|Av\| \leq (1 + \varepsilon)m \quad (13)$$

La démonstration du fait est très facile en introduisant la norme d'opérateur de $A : \ell_2^k \rightarrow (\mathbb{R}^n \|\cdot\|)$,

$$\|A\| = \sup_{x \in \mathbb{R}^k} \frac{\|Ax\|}{|x|}.$$

Soit $v \in S^{k-1}$ quelconque. En prenant un $y \in \mathcal{S}$ tel que $|v - y| \leq \frac{\varepsilon}{3}$ on a

$$\begin{aligned} \left(1 - \frac{\varepsilon}{3}\right)m - \frac{\varepsilon}{3}\|A\| &\leq \|Ay\| - \|Av - Ay\| \leq \|Av\|, \\ \|Av\| &\leq \|Ay\| + \|Av - Ay\| \leq \left(1 + \frac{\varepsilon}{3}\right)m + \frac{\varepsilon}{3}\|A\| \end{aligned} \quad (14)$$

Soit $v_0 \in S^{k-1}$ un vecteur où A atteint sa norme : $\|Av_0\| = \|A\|$. Alors le même raisonnement que ci-dessus donne

$$\|A\| = \|Av_0\| \leq \left(1 + \frac{\varepsilon}{3}\right)m + \frac{\varepsilon}{3}\|A\|$$

et donc

$$\|A\| \leq 2m. \quad (15)$$

En combinant (14) et (15), on obtient bien le résultat voulu (13).

Il reste donc à trouver un ε -réseau qui soit de cardinal pas trop excessif. Cela repose sur la petite estimation volumique suivante :

Lemme. — Pour tout $\delta > 0$ on peut trouver un δ -réseau \mathcal{S} de S^{k-1} tel que

$$|\mathcal{S}| \leq \left(1 + \frac{2}{\delta}\right)^k = e^{k \log(1 + \frac{2}{\delta})}.$$

Démonstration. — (**Joli**). Soit $\mathcal{S} = \{t^1, \dots, t^\ell\} \subset S^{k-1}$ un ensemble (pour l'instant quelconque) de ℓ points de la sphère tels que $|t^i - t^j| > \delta$, $\forall i \neq j$. Alors, dans \mathbb{R}^k , les boules euclidiennes $B(t^i, \frac{\delta}{2}) = \{x \in \mathbb{R}^k ; |x - t^i| \leq \frac{\delta}{2}\}$ sont disjointes, et contenues dans la boule $B(0, 1 + \frac{\delta}{2})$. En comparant les volumes on obtient

$$\ell \times \left(\frac{\delta}{2}\right)^k \leq \left(1 + \frac{\delta}{2}\right)^k,$$

soit encore $|\mathcal{S}| = \ell \leq \left(1 + \frac{2}{\delta}\right)^k$. On choisit maintenant un ensemble $\mathcal{S} \subset S^{k-1}$ de points de la sphère de cardinal maximal vérifiant la condition : $\forall t, s \in \mathcal{S}, t \neq s \Rightarrow |t - s| > \delta$ (ou plus simplement, je commence à construire une suite de points δ -séparé, et je m'arrête quand je

ne peut plus continuer, ce qui doit bien arriver). Alors, \mathcal{S} est automatiquement un δ -réseau de la sphère. \square

Terminons maintenant la démonstration du Lemme de Milman. On se donne donc un $\frac{\varepsilon}{3}$ -réseau \mathcal{S} de S^{k-1} tel que

$$|\mathcal{S}| \leq e^{k \log(1 + \frac{6}{\varepsilon})}.$$

Rappelons que $M = \int \|x\| d\gamma_n(x) = m$ et que $x \rightarrow \|x\|$ est b -lipschitzienne sur (\mathbb{R}^n, \cdot) . On applique avec l'ensemble \mathcal{S} le lemme fondamental et le Fait montré plus haut qui nous dit que

$$\{\forall y \in \mathcal{S}; (1 - \frac{\varepsilon}{3})M |y| \leq \|Gy\| \leq (1 + \frac{\varepsilon}{3})M |y|\} \subset \{\forall y \in S^{k-1}; (1 - \varepsilon)M |y| \leq \|Gy\| \leq (1 + \varepsilon)M |y|\}.$$

On obtient donc,

$$p := \mathbb{P}(\{\forall y \in S^{k-1}; (1 - \varepsilon)M \leq \|Gy\| \leq (1 + \varepsilon)M\}) \geq 1 - 2|\mathcal{S}|e^{-(\varepsilon/3)^2 \cdot M^2 / (4b^2)}.$$

En utilisant $2|\mathcal{S}| \leq |\mathcal{S}|^2$ on a :

$$p \geq 1 - e^{2k \log(1 + \frac{6}{\varepsilon}) - \varepsilon^2 k(E)/36}$$

et donc

$$p \geq 1 - e^{-2k}$$

dès que

$$k \leq \frac{k(E)}{36} \times \frac{\varepsilon^2}{2 + 2 \log(1 + \frac{6}{\varepsilon})}.$$

Soit c_0 une constante telle que $2 + 2 \log(1 + \frac{6}{\varepsilon}) \leq c_0 \log(1 + \frac{1}{\varepsilon})$ pour tout $\varepsilon \in]0, 1[$. Il suffit alors de prendre

$$k \leq \frac{1}{36 c_0} \cdot \frac{\varepsilon^2}{\log(1 + \frac{1}{\varepsilon})} \cdot k(E).$$

\square

Tous les résultats découlent du Lemme que nous venons de démontrer.

Démonstration de la version « plongement ». — Prenons le k du Lemme de Milman, $k = c\eta(\varepsilon)k(E)$. On veut construire un plongement de ℓ_2^k dans E . Rappelons que si $k \leq 1$ il n'y a rien à montrer car l'énoncé est vide! On utilise donc l'hypothèse cachée $k = c\eta(\varepsilon)k(E) \geq 1$. Soit alors G une matrice aléatoire gaussienne $n \times k$. L'événement

$$\{\omega; \forall y \in S^{k-1}; (1 - \varepsilon)m |y| \leq \|G_\omega y\| \leq (1 + \varepsilon)m |y|\}$$

est non vide puisqu'il a une probabilité $\geq 1 - e^{-2} > 0$. On prend alors ω_0 dans cet événement et l'application $A := G_{\omega_0}$ vérifie ce qu'il faut. \square

Démonstration de la version « section ». — On va appliquer encore le lemme de Milman aux espaces vectoriels $E = (\mathbb{R}^n, \|\cdot\|)$ et $\ell_2^n = (\mathbb{R}^n, |\cdot|)$. On rappelle que $k(\ell_2^n) \geq k(E)$, donc en fait si on prend $k := c\eta(\varepsilon)k(E)$, on a aussi $k \leq c\eta(\varepsilon)k(\ell_2^n)$. On a donc, pour G matrice gaussienne $n \times k$,

$$\mathbb{P}\{\omega; \forall y \in S^{k-1}; (1 - \varepsilon)m |y| \leq \|G_\omega y\| \leq (1 + \varepsilon)m |y|\} > \frac{1}{2}$$

et

$$\mathbb{P}\{\omega; \forall y \in S^{k-1}; (1 - \varepsilon)\tilde{m} |y| \leq |G_\omega y| \leq (1 + \varepsilon)\tilde{m} |y|\} > \frac{1}{2}$$

On a alors que l'intersection entre ces deux événements est non-vide, ce qui assure l'existence d'un opérateur $A : \ell_2^k \rightarrow E$ tel que

$$\forall y \in \mathbb{R}^k, \quad M \frac{1-\varepsilon}{1+\varepsilon} |Ay| \leq \|Ay\| \leq M \frac{(1+\varepsilon)}{1-\varepsilon} |Ay|,$$

où l'on a posé $M = m/\tilde{m}$. En posant $F := \{Ay ; y \in \mathbb{R}^k\} \subset \mathbb{R}^n$, on voit que F est un sev de dimension k sur lequel on a :

$$\forall x \in F, \quad M(1-2\varepsilon)|x| \leq \|x\| \leq M(1+2\varepsilon)|x|$$

□

4.7. Lemme de Dvoretzky-Rogers et conséquence

Exercice. — (*Position de John et Dvoretzky-Rogers*)

1. Soit $E = (\mathbb{R}^n, \|\cdot\|)$ un n -evn et notons K sa boule unité (de manière équivalente, on se donne un corps convexe symétrique $K \subset \mathbb{R}^n$). On pose

$$\mathcal{A} := \{u \in GL(\mathbb{R}^n) ; \|u\|_{\ell_2^n \rightarrow E} \leq 1\}.$$

(a) Montrer que la fonction $M \rightarrow |\det M|$ atteint son maximum sur l'ensemble \mathcal{A} .

(b) Montrer que $\mathcal{A} = \{u \in GL(\mathbb{R}^n) ; u(B_2^n) \subset K\}$.

Remarque : si u_0 est une application où le sup du déterminant sur \mathcal{A} est atteint, alors si on considère la norme $\|x\|_{\tilde{E}} := \|u_0(x)\|$, ou de manière équivalente $\tilde{K} = u_0^{-1}K$, alors le sup

$$\sup\{|\det u| ; u : GL(\mathbb{R}^n), \|u\|_{\ell_2^n \rightarrow \tilde{E}} \leq 1\}.$$

est atteint pour $u = \text{Id}_{\mathbb{R}^n}$.

2. On dit qu'un corps convexe symétrique $K \subset \mathbb{R}^n$ – ou que l'evn correspondant $E_K := (\mathbb{R}^n, \|\cdot\|_K)$ – est en *position de John* si la boule euclidienne usuelle B_2^n est incluse dans K et qu'elle a un volume maximal parmi tous les ellipsoïdes contenus dans K .

Montrer qu'on peut toujours trouver une transformation linéaire u tel que uK soit en position de John.

3. Montrer que si K est en position de John, alors pour toute application linéaire v , on a

$$\text{tr}(v) \leq n \|v\|_{\ell_2^n \rightarrow E_K}.$$

Indication : on pourra faire un DL pour $\varepsilon \rightarrow 0$ de $\det(\text{Id}_{\mathbb{R}^n} + \frac{\varepsilon}{\|v\|_{\ell_2^n \rightarrow E_K}} v)$ et utiliser que $|\det(u)| \leq \|u\|_{\ell_2^n \rightarrow E_K}^n$.

En déduire que si F est un sev de dimension $i \in [1, n]$ et P est une projection orthogonale sur F^\perp , on a

$$\|P\|_{\ell_2^n \rightarrow E_K} \geq \frac{n-i}{n}.$$

4. Démontrer le

Théorème (Lemme de Dvoretzky-Rogers). — Soit $K \subset \mathbb{R}^n$ un corps convexe symétrique en position de John. Alors, il existe une base orthonormée w_1, \dots, w_n telle que,

$$\forall i \leq n, \quad \|w_i\|_K \geq \frac{n-i+1}{n}$$

et en particulier que

$$\forall i \leq [n/2], \quad \|w_i\|_K \geq \frac{1}{2}.$$

Montrer également la version evn

Théorème (Lemme de Dvoretzky-Rogers). — Soit $(E, \|\cdot\|)$ un evn de dimension $n \geq 2$ et $m = [n/2]$. Alors il existe m vecteurs w_1, \dots, w_m de E tel que $\|w_j\| \geq \frac{1}{2}$ et

$$\forall a \in \mathbb{R}^m, \quad \left\| \sum_{j=1}^m a_j w_j \right\| \leq \|a\|_{\ell_2^m}.$$

5. Soit $\|\cdot\|$ une norme sur \mathbb{R}^n et e_1, \dots, e_n une base orthonormée de \mathbb{R}^n .

(a) Montrer que

$$\int \|x\| d\gamma_n(x) = \frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \int_{\mathbb{R}^n} \left\| \sum_{i=1}^n \varepsilon_i x_i e_i \right\| d\gamma_n(x).$$

(b) Montrer que pour $x \in \mathbb{R}^n$ fixé, on a :

$$\frac{1}{2^n} \sum_{\varepsilon \in \{-1,1\}^n} \left\| \sum_{i=1}^n \varepsilon_i x_i e_i \right\| \geq \max_{i \leq n} \|x_i e_i\|.$$

6. Soit K un corps convexe symétrique de \mathbb{R}^n . Montrer que si $K \subset \mathbb{R}^n$ est en position de John, alors on a

$$\int \|x\|_K d\gamma_n(x) \geq c\sqrt{\log(n)}$$

pour une certaine constante numérique c .

En déduire que si $K \subset \mathbb{R}^n$ est en position de John, alors la constante de Dvoretzky-Milman vérifie

$$k(K) \geq c \log(n).$$