

Mémoire de première année :
Reconstruction d'un vecteur parcimonieux à partir de mesures
linéaires

T. BÉNARD et T. MASSONI
encadré par I. WALDSPURGER

Table des matières

Introduction	3
1 Condition RIP, théorèmes de reconstruction	5
1.1 Constantes d'isométrie restreinte et propriété RIP	5
1.2 Théorème de reconstruction dans le cas d'un vecteur parcimonieux	6
1.3 Théorème de reconstruction dans le cas général	9
1.4 Sur l'optimalité de l'erreur de reconstruction	11
2 Matrices aléatoires gaussiennes	14
2.1 Probabilité de vérifier RIP	14
2.2 Raffinement du résultat	17
2.3 Démonstration des inégalités de concentrations	20
3 Bornes pour les espérances des valeurs singulières	22
3.1 Inégalités de Slepian-Gordon	22
3.2 Application à des processus gaussiens	29
3.3 Démonstration du résultat	32
4 Inégalités de concentration pour des fonctions lipschitziennes	35
4.1 Concentration autour de la moyenne	35
4.2 Concentration autour de la médiane et isopérimétrie gaussienne	37
A Valeurs singulières de matrices	42
B Remarques sur les mesures gaussiennes	43

Introduction

Dans ce texte, on s'intéresse à la question suivante :

Comment identifier un signal inconnu en effectuant le moins de mesures possibles ?

Les applications pratiques sont nombreuses, par exemple dans l'imagerie médicale où limiter le nombre de mesures effectuées par un scanner limite aussi le temps d'exposition du patient aux radiations nocives.

On modélise un signal par un élément $x_0 \in \mathbb{R}^n$. Une mesure est la donnée d'un couple $(u, \langle u, x_0 \rangle)$ où $u \in \mathbb{R}^n$. Se donner m mesures d'un signal $x_0 \in \mathbb{R}^n$ revient donc à se donner un couple (A, Ax_0) où $A \in \mathcal{M}_{m,n}(\mathbb{R})$. Dans ce cas, la matrice A est appelée *matrice de mesure* et Ax_0 la *mesure de x_0 par A* . La question initiale se transpose en le problème suivant :

Soit $x_0 \in \mathbb{R}^n$. Comment trouver une matrice $A \in \mathcal{M}_{m,n}(\mathbb{R})$ avec peu de lignes ($m \ll n$) telle qu'on puisse reconstruire x_0 à partir de la connaissance du couple (A, Ax_0) ?

La difficulté réside dans la condition $m \ll n$ qui implique que A n'est pas injective. Ainsi, quelque soit le choix de A , x_0 ne pourra pas être déterminé par la mesure (A, Ax_0) . La reconstruction n'est donc pas envisageable sans imposer des conditions supplémentaires sur x_0 . La condition privilégiée est la parcimonie du signal, c'est-à-dire le fait qu'il comporte peu de coefficients non nuls ($\|x_0\|_0 \ll n$). Cette condition est naturelle car beaucoup de signaux réels sont quasi-parcimonieux (*i.e.* leurs valeurs sont essentiellement concentrées sur un support petit devant leur dimension).

La démarche proposée dans ce texte comporte quatre étapes :

Dans un premier temps, nous donnons une condition (appelée RIP) pour qu'une matrice de mesure $A \in \mathcal{M}_{m,n}(\mathbb{R})$ permette de reconstruire tout signal $x_0 \in \mathbb{R}^n$ assez parcimonieux. En particulier, nous prouvons le résultat suivant :

Soit $A \in \mathcal{M}_{m,n}(\mathbb{R})$. Pour $T \subset \llbracket 1, n \rrbracket$, on note A_T la matrice de $\mathcal{M}_{m,|T|}(\mathbb{R})$ obtenue à partir de A en conservant les colonnes d'indices dans T et $\lambda_{\min}(A_T)$, $\lambda_{\max}(A_T)$ ses valeurs singulières extrémales.

Si A vérifie

$$\text{RIP}_S \quad : \quad \min_{|T| \leq S} \{1 - \lambda_{\min}^2(A_T), -1 + \lambda_{\max}^2(A_T)\} < \frac{1}{2}$$

alors tout signal $x_0 \in \mathbb{R}^n$ S -parcimonieux (de support de taille au plus S) est l'unique solution du problème :

$$Ax_0 = Ay \quad \text{avec} \quad \|y\|_1 \text{ minimale}$$

Ceci permet de reconstruire x_0 à partir du couple (A, Ax_0) par des méthodes d'optimisation convexe. Nous montrerons plus généralement que cette méthode de reconstruction est efficace dans le cas de mesures imparfaites et de vecteurs quasi-parcimonieux.

Dans la deuxième partie, nous voyons comment obtenir des matrices vérifiant RIP par des considérations probabilistes. Nous montrons essentiellement qu'il existe des constantes $\alpha, K \in \mathbb{R}_+^*$ telles que si $A : \Omega \rightarrow \mathcal{M}_{m,n}(\mathbb{R})$ est une matrice aléatoire gaussienne, alors A vérifie RIP_S avec une probabilité supérieure à $1 - 2e^{-Km}$ pourvu que S vérifie :

$$S \leq \alpha \frac{m}{\ln\left(\frac{n}{m}\right)}$$

Pour cela, il nous faudra établir des inégalités de concentration sur les valeurs singulières des matrices aléatoires gaussiennes.

Les deux dernières parties présentent des résultats généraux (Inégalités de Slepian-Gordon, concentration de la mesure, inégalité isopérimétrique gaussienne) permettant de démontrer ces inégalités de concentration.

On établit dans la troisième partie des bornes pour les espérances des valeurs singulières extrémales d'une matrice aléatoire gaussienne $A : \Omega \longrightarrow \mathcal{M}_{m,n}(\mathbb{R})$:

$$1 - \sqrt{\frac{m}{n}} < \mathbb{E} [\lambda_{\min}(A)] \leq \mathbb{E} [\lambda_{\max}(A)] < 1 + \sqrt{\frac{m}{n}}$$

On démontre dans la dernière partie un théorème de concentration pour les fonctions 1-lipschitziennes $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ autour de leur moyenne pour la mesure gaussienne standard γ_n sur \mathbb{R}^n :

$$\forall r \geq 0, \gamma_n \left(f \geq \int f d\gamma_n + r \right) \leq \exp \left(-\frac{r^2}{2} \right)$$

On s'intéresse également à des inégalités de concentration analogues autour de la médiane qui découlent naturellement de théorèmes sur l'isopérimétrie pour la mesure gaussienne.

1 Condition RIP, théorèmes de reconstruction

Dans cette section, on se donne des entiers naturels m et n vérifiant $m \leq n$, ainsi qu'une matrice de mesure $A \in \mathcal{M}_{m,n}(\mathbb{R})$. On exhibe une condition sur A pour que la donnée de la mesure (A, Ax_0) détermine x_0 pour tout $x_0 \in \mathbb{R}^n$ assez parcimonieux. On constate alors que si la mesure est imparfaite (*i.e.* on connaît seulement (A, ϵ, y) où $\epsilon \geq 0$ est l'erreur de mesure et $y \in B_2(Ax_0, \epsilon)$ est la mesure approchée), on peut quand même déterminer x_0 avec une erreur linéaire en ϵ .

Nous adopterons le principe de reconstruction suivant : Soit $x_0 \in \mathbb{R}^n$, (A, ϵ, y) une mesure approchée de x_0 d'erreur ϵ . Une reconstruction x^\sharp associée à (A, ϵ, y) est un signal de \mathbb{R}^n vérifiant :

$$\|Ax^\sharp - y\|_2 \leq \epsilon \quad \text{avec} \quad \|x^\sharp\|_1 \text{ minimale} \quad (\star)$$

Remarquons que cette reconstruction est définie à partir de A , y et ϵ mais ne dépend pas de x_0 .

Il est commode de poser, pour $x \in \mathbb{R}^n$: $\|x\|_0 := \text{Card}(\{i \in \llbracket 1, n \rrbracket : x_i \neq 0\})$.

La condition $\|\cdot\|_1$ minimale vient du fait que minimiser la norme $\|\cdot\|_1$ permet souvent de minimiser la pseudo-norme $\|\cdot\|_0$ ce qui revient à maximiser la parcimonie du signal. On ne cherche pas à minimiser $\|x^\sharp\|_0$ car $\|\cdot\|_0$ n'est pas une norme. De plus, il existe des méthodes efficaces de minimisation de la norme $\|\cdot\|_1$. Ce problème relève en effet de l'optimisation convexe.

1.1 Constantes d'isométrie restreinte et propriété RIP

Si T est un sous-ensemble de $\llbracket 1, n \rrbracket$, on note A_T la matrice de $\mathcal{M}_{m,|T|}(\mathbb{R})$ obtenue à partir de A en conservant les colonnes d'indices dans T .

Afin de reconstruire un signal x_0 S -parcimonieux de \mathbb{R}^n , il est naturel d'imposer des conditions particulières aux matrices A_T pour des sous-ensembles T de cardinal plus petit que S . Cela revient à imposer des conditions sur les ensembles formés d'au plus S colonnes de A . On introduit ici les *constantes d'isométrie restreinte* qui "mesurent" l'orthogonalité de sous-matrices de A :

Définition 1. Soit $S \in \mathbb{N}^*$, on note Δ l'ensemble des réels positifs δ vérifiant : pour tout $T \subset \llbracket 1, n \rrbracket$ avec $|T| \leq S$ et pour tout $c \in \mathbb{R}^{|T|}$,

$$(1 - \delta)\|c\|_2^2 \leq \|A_T c\|_2^2 \leq (1 + \delta)\|c\|_2^2$$

La constante d'isométrie restreinte à S de A est définie par :

$$\delta_S := \min \Delta$$

Remarque (1). Si $S \leq m$, on a de manière équivalente que δ_S est le plus petit réel positif δ vérifiant : pour tout $T \subset \llbracket 1, n \rrbracket$ avec $|T| \leq S$,

$$1 - \delta \leq \lambda_{\min}^2(A_T) \quad \text{et} \quad \lambda_{\max}^2(A_T) \leq 1 + \delta$$

où $\lambda_{\min}(A_T)$ et $\lambda_{\max}(A_T)$ désignent les valeurs singulières minimale et maximale de A_T (cf. Annexe pour plus de précisions).

Remarque (2). Si $S, S' \in \mathbb{N}^*$ vérifient $S \leq S'$, alors $\delta_S \leq \delta_{S'}$.

Remarque (3). On peut ne considérer que les parties T de $\llbracket 1, n \rrbracket$ de cardinal exactement S dans la définition de Δ ce qui ne modifie pas la valeur de δ_S .

Dans la suite, on s'intéresse à la propriété "Restricted Isometry Property" (RIP) qui s'énonce :

$$\delta_{3S} + 3\delta_{4S} < 2$$

Nous démontrons dans la section suivante que si A vérifie RIP pour l'entier S , alors la reconstruction de vecteurs S -parcimonieux à partir de A peut se faire avec une erreur en norme $\|\cdot\|_2$ linéaire en l'erreur de mesure ϵ .

Nous verrons dans la Section 2 que les matrices aléatoires gaussiennes vérifient cette propriété avec une grande probabilité, ce qui permet une bonne reconstruction de signaux par des mesures aléatoires.

1.2 Théorème de reconstruction dans le cas d'un vecteur parcimonieux

Si $S \in \mathbb{N}$, on dit qu'un vecteur $x \in \mathbb{R}^n$ est S -parcimonieux s'il vérifie : $\|x\|_0 \leq S$.

Soit $x \in \mathbb{R}^n$ et T une partie non vide de $\llbracket 1, n \rrbracket$, la restriction de x à T est le vecteur $x_T \in \mathbb{R}^{|T|}$ obtenue en conservant seulement les composantes de x dont les indices sont dans T .

Théorème 1. Soit $A \in \mathcal{M}_{m,n}(\mathbb{R})$ et $S \in \mathbb{N}$ tels que : $\delta_{3S} + 3\delta_{4S} < 2$.

Il existe une constante $C_S > 0$ telle que pour tout $x_0 \in \mathbb{R}^n$ S -parcimonieux et pour toute mesure approchée (A, ϵ, y) de x_0 , on a :

$$\|x_0 - x^\sharp\|_2 \leq C_S \epsilon$$

où x^\sharp est une reconstruction de x_0 obtenue en résolvant le problème (\star) .

Ce théorème affirme que la condition $\delta_{3S} + 3\delta_{4S} < 2$ est suffisante pour retrouver de façon stable les signaux S -parcimonieux grâce à A : étant donné un signal $x_0 \in \mathbb{R}^n$ S -parcimonieux, la connaissance de sa mesure par A à ϵ près permet de reconstruire x_0 avec une erreur proportionnelle à ϵ . Cette reconstruction est obtenue en minimisant la norme $\|\cdot\|_1$ parmi les signaux de mesures ϵ -proches de la mesure imparfaite.

La condition de validité est vérifiée si $\delta_{4S} < \frac{1}{2}$. On verra dans la démonstration que dans ce cas, la constante C_S ne dépend que de δ_{4S} et prend des valeurs raisonnables si δ_{4S} n'est pas trop près de la borne.

La preuve montre aussi que le résultat reste vrai si l'on suppose seulement que pour un certain α tel que αS est entier,

$$\delta_{\alpha S} + \alpha \delta_{(\alpha+1)S} < \alpha - 1$$

Démonstration. Fixons x_0 , ϵ et y comme ci-dessus. On pose $h := x_0 - x^\sharp$. La définition de x^\sharp réside en deux conditions :

1. Une contrainte cylindrique

L'inégalité $\|Ax^\sharp - y\|_2 \leq \epsilon$ implique par l'inégalité triangulaire que :

$$\|Ah\|_2 \leq 2\epsilon$$

h appartient donc au cylindre de direction $\text{Ker } A$ et de rayon $2r\epsilon$, où r est déterminé par A .

2. Une contrainte conique

On introduit T_0 le support de x_0 et T_0^c son complémentaire dans $\llbracket 1, n \rrbracket$.

En décomposant $x^\sharp = x_0 + h$, l'inégalité $\|x^\sharp\|_1 \leq \|x_0\|_1$ ainsi que l'inégalité triangulaire impliquent que :

$$\|x_0\|_1 - \|h_{T_0}\|_1 + \|h_{T_0^c}\|_1 \leq \|x_0 + h_{T_0}\|_1 + \|h_{T_0^c}\|_1 \leq \|x_0\|_1$$

d'où

$$\|h_{T_0^c}\|_1 \leq \|h_{T_0}\|_1$$

Géométriquement h appartient donc à un cône de sommet 0.

- Pour donner une idée de la suite, supposons provisoirement que x^\sharp et x_0 ont le même support. Dans ce cas, $\|h_{T_0^c}\|_1 = 0$ ce qu'on peut voir comme une contrainte conique renforcée. Alors, $\|h\|_2 = \|h_{T_0}\|_2$. D'autre part, on a :

$$\|Ah\|_2 = \|A_{T_0} h_{T_0}\|_2 \geq \sqrt{1 - \delta_{|T_0|}} \|h_{T_0}\|_2 \geq \sqrt{1 - \delta_S} \|h_{T_0}\|_2$$

La racine est bien définie si $\delta_S \leq 1$ ce qui est vrai quand $\delta_{3S} + 3\delta_{4S} < 2$. En réunissant ces deux observations, il vient :

$$\|Ah\|_2 \geq \sqrt{1 - \delta_S} \|h\|_2$$

Finalement, la contrainte cylindrique permet de majorer :

$$\|h\|_2 \leq \frac{2}{\sqrt{1 - \delta_S}} \epsilon$$

et on a le résultat voulu pour $C_S = \frac{2}{\sqrt{1 - \delta_S}}$.

La démonstration générale suit le même cheminement :

- On montre d'abord que la norme $\|\cdot\|_2$ de h est concentrée sur un support assez proche de T_0 . Pour cela, on introduit $M \in \mathbb{N}^*$ un paramètre (que l'on choisira plus tard) et on décompose T_0^c en $T_1 \sqcup T_2 \sqcup \dots \sqcup T_d$, où T_1 contient les indices des M plus grand coefficients de T_0^c (en valeur absolue), T_2 les indices des M suivants, et ainsi de suite jusqu'à exhaustion de T_0^c (il est bien sûr possible que T_d contienne moins de M éléments).

Avec cette décomposition, la norme $\|\cdot\|_2$ de h est concentrée sur $T_{01} := T_0 \sqcup T_1$. En effet, le k -ième plus grand élément de $h_{T_0^c}$ vérifie :

$$|h_{T_0^c}|_{(k)} \leq \frac{\|h_{T_0^c}\|_1}{k}$$

et ainsi

$$\|h_{T_{01}^c}\|_2^2 = \sum_{k=M+1}^n |h_{T_0^c}|_{(k)}^2 \leq \|h_{T_0^c}\|_1^2 \sum_{k=M+1}^n \frac{1}{k^2} \leq \frac{\|h_{T_0}\|_1^2}{M}$$

où la dernière inégalité est obtenue en utilisant la contrainte du cône et une majoration intégrale de la somme. Or, l'inégalité de Cauchy-Schwartz affirme que :

$$\|h_{T_0}\|_1^2 \leq |T_0| \|h_{T_0}\|_2^2$$

d'où, en posant $\rho := \frac{|T_0|}{M}$, on obtient "l'inégalité de concentration" :

$$\boxed{\|h\|_2^2 \leq (1 + \rho) \|h_{T_{01}}\|_2^2}$$

Il reste donc à majorer $\|h_{T_{01}}\|_2$ pour conclure, ce que l'on va faire en invoquant la contrainte cylindrique.

- Tout d'abord, on a :

$$\|Ah\|_2 = \|A_{T_{01}} h_{T_{01}} + \sum_{j \geq 2} A_{T_j} h_{T_j}\|_2 \geq \|A_{T_{01}} h_{T_{01}}\|_2 - \left\| \sum_{j \geq 2} A_{T_j} h_{T_j} \right\|_2 \geq \|A_{T_{01}} h_{T_{01}}\|_2 - \sum_{j \geq 2} \|A_{T_j} h_{T_j}\|_2$$

d'où

$$\|Ah\|_2 \geq \sqrt{1 - \delta_{M+|T_0|}} \|h_{T_0}\|_2 - \sqrt{1 + \delta_M} \sum_{j \geq 2} \|h_{T_j}\|_2$$

en supposant que $\delta_{M+|T_0|} \leq 1$, ce qui sera vérifié lorsqu'on aura choisi M .

On admet provisoirement l'inégalité

$$\sum_{j \geq 2} \|h_{T_j}\|_2 \leq \sqrt{\rho} \|h_{T_{01}}\|_2 \tag{1}$$

pour obtenir :

$$\|Ah\|_2 \geq C_{|T_0|,M} \|h_{T_0}\|_2$$

où

$$C_{|T_0|,M} := \sqrt{1 - \delta_{M+|T_0|}} - \sqrt{\rho} \sqrt{1 + \delta_M}$$

Supposons avoir $C_{|T_0|,M} > 0$, ce qui sera vrai quand on aura choisi M . La contrainte cylindrique implique alors :

$$\|h_{T_0}\|_2 \leq \frac{2}{C_{|T_0|,M}} \epsilon$$

- Enfin, l'inégalité précédente et l'inégalité de concentration permettent d'obtenir :

$$\|h\|_2 \leq \frac{2\sqrt{1+\rho}}{C_{|T_0|,M}} \epsilon$$

En posant

$$C := \frac{2\sqrt{1+\rho}}{C_{|T_0|,M}}$$

on a alors le résultat :

$$\|x_0 - x^\# \|_2 \leq C \epsilon$$

On se penche maintenant sur le choix de M . Les deux hypothèses nécessaires ont été $\delta_{M+|T_0|} \leq 1$ et $C_{|T_0|,M} > 0$. Un rapide calcul montre qu'en notant $\alpha := \frac{1}{\rho}$, la dernière inégalité équivaut à

$$\delta_{\alpha|T_0|} + \alpha \delta_{(\alpha+1)|T_0|} < \alpha - 1$$

En spécifiant $M = 3|T_0|$ (*i.e.* $\alpha = 3$), cette condition devient $\delta_{3|T_0|} + 3\delta_{4|T_0|} < 2$; comme $|T_0| \leq S$, elle est vérifiée d'après la condition RIP. Ainsi, le choix $M = 3|T_0|$ convient. Il en découle également que $\delta_{M+|T_0|} = \delta_{4|T_0|} \leq \delta_{4S} \leq 1$.

On peut alors majorer la constante C par la constante C_S :

$$C_S := \frac{4}{\sqrt{3} \sqrt{1 - \delta_{4S}} - \sqrt{1 + \delta_{3S}}}$$

qui ne dépend que de δ_{3S} et de δ_{4S} .

- Il ne nous reste plus qu'à démontrer l'inégalité (1).

Pour tout indice j , on a :

$$|h_{T_{j+1}}(t)| \leq \frac{\|h_{T_j}\|_1}{M}$$

d'où

$$\|h_{T_{j+1}}\|_2^2 \leq \frac{\|h_{T_j}\|_1^2}{M}$$

Ainsi :

$$\sum_{j \geq 2} \|h_{T_j}\|_2 \leq \frac{1}{\sqrt{M}} \sum_{j \geq 1} \|h_{T_j}\|_1 = \frac{\|h_{T_0^c}\|_1}{\sqrt{M}} \leq \frac{\|h_{T_0}\|_1}{\sqrt{M}}$$

où la dernière inégalité découle de la contrainte conique, et en utilisant l'inégalité de Cauchy-Schwartz, on obtient bien :

$$\sum_{j \geq 2} \|h_{T_j}\|_2 \leq \frac{\|h_{T_0}\|_1}{\sqrt{M}} \leq \sqrt{\frac{|T_0|}{M}} \|h_{T_0}\|_2 = \sqrt{\rho} \|h_{T_0}\|_2 \leq \sqrt{\rho} \|h_{T_{01}}\|_2$$

ce qui termine la démonstration du théorème. □

Remarque. La condition RIP impose que le nombre de mesures m soit supérieur à la parcimonie S du signal à reconstruire. En effet, si la matrice A vérifie RIP pour S et si x_0 et x'_0 sont deux signaux S -parcimonieux vérifiant $Ax_0 = Ax'_0$, alors le théorème précédent implique que $x_0 = x'_0$. Par conséquent, A est de rang au moins S et comme $m \leq n$, A est de rang au plus m , d'où $S \leq m$.

1.3 Théorème de reconstruction dans le cas général

Dans cette section, on ne suppose plus le signal x_0 parcimonieux et on généralise le Théorème 1 :

Théorème 2. Soit $A \in \mathcal{M}_{m,n}(\mathbb{R})$ et $S \in \mathbb{N}$ tels que : $\delta_{3S} + 3\delta_{4S} < 2$.

Il existe des constantes $C_{1,S}, C_{2,S} > 0$ telles que pour tout $x_0 \in \mathbb{R}^n$ et pour toute mesure approchée (A, ϵ, y) de x_0 , on a :

$$\|x_0 - x^\sharp\|_2 \leq C_{1,S} \epsilon + C_{2,S} \frac{\|x_0 - x_{0S}\|_2}{\sqrt{S}}$$

où x^\sharp est une reconstruction de x_0 obtenue en résolvant le problème (\star) , et où x_{0S} est obtenu à partir de x_0 en conservant ses S plus grands coefficients et en remplaçant les autres par 0.

On retrouve bien le résultat du Théorème 1 pour un signal x_0 S -parcimonieux, puisque dans ce cas : $\|x_0 - x_{0S}\|_2 = 0$.

Ce théorème nous renseigne donc sur l'erreur de reconstruction d'un signal x_0 dans le cas où celui-ci n'est pas parcimonieux ; si ce signal est "quasi- S -parcimonieux", c'est à dire si ses valeurs sont essentiellement concentrées sur un ensemble de cardinal S , alors la quantité $\|x_0 - x_{0S}\|_2$ sera faible ce qui garantira une bonne reconstruction pour la norme $\|\cdot\|_2$.

Comme dans le cas du Théorème 1, le résultat reste valide s'il existe un α tel que αS est entier et vérifie :

$$\delta_{\alpha S} + \alpha \delta_{(\alpha+1)S} < \alpha - 1$$

La démonstration du Théorème 2 est similaire à celle du Théorème 1 et s'obtient en modifiant les inégalités obtenues lors de sa preuve, puisque la "contrainte conique" n'est plus valide mais peut être remplacée par une "contrainte conique modifiée".

Démonstration. On utilise les mêmes notations que dans la démonstration du Théorème 1. On pose T_0 l'ensemble des indices des S plus grands coefficients de x_0 .

Encore une fois, la définition de x^\sharp impose deux conditions :

1. Une contrainte cylindrique

C'est la même que pour le Théorème 1 :

$$\|Ah\|_2 \leq 2\epsilon$$

2. Une contrainte conique modifiée

À partir de l'inégalité triangulaire et de la minimalité de x^\sharp pour la norme $\|\cdot\|_1$, on a :

$$\|x_{0,T_0}\|_1 - \|h_{T_0}\|_1 - \|x_{0,T_0^c}\|_1 + \|h_{T_0^c}\|_1 \leq \|x_{0,T_0} + h_{T_0}\|_1 + \|x_{0,T_0^c} + h_{T_0^c}\|_1 = \|x^\sharp\|_1 \leq \|x_0\|_1$$

On en déduit une contrainte conique "modifiée" :

$$\|h_{T_0^c}\|_1 \leq \|h_{T_0}\|_1 + 2\|x_{0,T_0^c}\|_1$$

- On se donne à nouveau un $M \in \mathbb{N}^*$ que l'on précisera plus tard et on décompose T_0^c en $T_1 \sqcup T_2 \sqcup \dots \sqcup T_d$ de la même manière que dans la démonstration du Théorème 1. Avec un même type de raisonnement, on va en déduire une inégalité de concentration pour h .

En effet, on a toujours

$$\|h_{T_{01}^c}\|_2^2 = \sum_{k=M+1}^n |h_{T_0^c}|_{(k)}^2 \leq \|h_{T_0^c}\|_1^2 \sum_{k=M+1}^n \frac{1}{k^2} \leq \frac{\|h_{T_0^c}\|_1^2}{M}$$

et en posant $\rho = \frac{S}{M}$ et $\eta := \frac{\|x_{0,T_0^c}\|_1}{\sqrt{S}}$, la contrainte conique modifiée permet d'obtenir :

$$\|h\|_2 \leq \|h_{T_{01}}\|_2 + \|h_{T_{01}^c}\|_2 \leq (1 + \sqrt{\rho}) \|h_{T_{01}}\|_2 + 2\sqrt{\rho}\eta$$

On peut remarquer que l'utilisation de l'inégalité triangulaire donne une majoration plus grossière que l'égalité de Pythagore (utilisée pour démontrer l'inégalité de concentration dans le cas parcimonieux), mais elle permet d'avoir une inégalité moins longue et plus explicite :

$$\boxed{\|h\|_2 \leq (1 + \sqrt{\rho}) \|h_{T_{01}}\|_2 + 2\sqrt{\rho}\eta}$$

On majore à nouveau $\|h_{T_{01}}\|_2$ afin de conclure en invoquant la contrainte cylindrique.

- En supposant $\delta_{M+S} \leq 1$, d'après la démonstration du Théorème 1, on a toujours :

$$\|Ah\|_2 \geq \sqrt{1 - \delta_{M+S}} \|h_{T_0}\|_2 - \sqrt{1 + \delta_M} \sum_{j \geq 2} \|h_{T_j}\|_2$$

En admettant provisoirement l'inégalité

$$\sum_{j \geq 2} \|h_{T_j}\|_2 \leq \sqrt{\rho} (\|h_{T_{01}}\|_2 + 2\eta) \tag{2}$$

on obtient :

$$\|Ah\|_2 \geq C_{S,M} \|h_{T_{01}}\|_2 - 2\sqrt{\rho}\eta\sqrt{1 + \delta_M}$$

où $C_{S,M}$ désigne la même constante que dans le Théorème 1. Alors, en supposant $C_{S,M} > 0$ et en utilisant la contrainte cylindrique, on a :

$$\boxed{\|h_{T_{01}}\|_2 \leq \frac{2\epsilon}{C_{S,M}} + \frac{2\sqrt{\rho}\eta\sqrt{1 + \delta_M}}{C_{S,M}}}$$

- Enfin, en combinant l'inégalité précédente avec l'inégalité de concentration modifiée, on obtient :

$$\|h\|_2 \leq \frac{2(1+\sqrt{\rho})}{C_{S,M}} \epsilon + 2\sqrt{\rho} \left(1 + \frac{(1+\sqrt{\rho})\sqrt{1+\delta_M}}{C_{S,M}} \right) \eta$$

d'où en posant :

$$C_1 := \frac{2(1+\sqrt{\rho})}{C_{S,M}} \quad \text{et} \quad C_2 := 2\sqrt{\rho} \left(1 + \frac{(1+\sqrt{\rho})\sqrt{1+\delta_M}}{C_{S,M}} \right)$$

on obtient l'inégalité souhaitée :

$$\boxed{\|x_0 - x^\# \|_2 \leq C_1 \epsilon + C_2 \frac{\|x_{0,T_0^c}\|_1}{\sqrt{S}}}$$

Pour le choix de M , les hypothèses faites sur δ_S et δ_{M+S} étant les mêmes que pour le Théorème 1, on en déduit que pour $M = 3S$, l'hypothèse RIP de l'énoncé justifie $\delta_{M+S} \leq 1$ et $C_{S,M} > 0$. Alors, les constantes C_1 et C_2 deviennent :

$$C_1 = \frac{2(1+\sqrt{3})}{\sqrt{3}\sqrt{1-\delta_{4S}} - \sqrt{1+\delta_{3S}}} \quad \text{et} \quad C_2 = \frac{2}{\sqrt{3}} \left(1 + \frac{1}{2}\sqrt{1+\delta_{3S}} C_{1,S} \right)$$

et ne dépendent que de δ_{3S} et de δ_{4S} .

- Il ne nous reste plus qu'à démontrer l'inégalité (2).

De même dans la démonstration de l'inégalité (1) du Théorème 1, on a :

$$\sum_{j \geq 2} \|h_{T_j}\|_2 \leq \frac{1}{\sqrt{M}} \sum_{j \geq 1} \|h_{T_j}\|_1 = \frac{\|h_{T_0^c}\|_1}{\sqrt{M}}$$

Alors, la contrainte conique modifiée donne :

$$\sum_{j \geq 2} \|h_{T_j}\|_2 \leq \frac{\|h_{T_0}\|_1}{\sqrt{M}} + \frac{2\|x_{0,T_0^c}\|_1}{\sqrt{M}}$$

et par l'inégalité de Cauchy-Schwartz,

$$\sum_{j \geq 2} \|h_{T_j}\|_2 \leq \sqrt{\frac{S}{M}} \|h_{T_0}\|_1 + 2\sqrt{\frac{S}{M}} \frac{\|x_{0,T_0^c}\|_1}{\sqrt{S}} = \sqrt{\rho} (\|h_{T_0}\|_1 + 2\eta) \leq \sqrt{\rho} (\|h_{T_{01}}\|_1 + 2\eta)$$

ce qui termine la démonstration du théorème. □

1.4 Sur l'optimalité de l'erreur de reconstruction

Dans le premier théorème de reconstruction, nous avons vu qu'en connaissant seulement la taille du support d'un signal à reconstruire, si la matrice de mesure considérée vérifie l'hypothèse RIP, alors la reconstruction de ce signal est possible, avec une erreur linéaire en l'erreur de mesure. On peut alors se demander s'il existe une façon de reconstruire le signal inconnu avec une erreur d'un ordre plus faible. La réponse est en général négative. Pour comprendre pourquoi, on peut s'intéresser à une autre

méthode de reconstruction, la méthode "Least-Squares". Celle-ci est très efficace en pratique, mais elle requiert la connaissance du support du signal à reconstruire (et pas seulement sa taille, comme précédemment!).

Cette méthode repose sur le principe suivant : soit $A \in \mathcal{M}_{m,n}(\mathbb{R})$ et $T_0 \subset \llbracket 1, n \rrbracket$ non vide telle la matrice carrée $A_{T_0}^T A_{T_0}$ est inversible. Cette condition est vérifiée si par exemple $\delta_{|T_0|} < 1$, puisque pour tout $x \in \mathbb{R}^{|T_0|}$ non nul, on a :

$$0 < (1 - \delta_{|T_0|}) \|x\|_2^2 \leq \|A_{T_0} x\|_2^2 = x^T A_{T_0}^T A_{T_0} x$$

Ensuite, si $x_0 \in \mathbb{R}^n$ est un signal à reconstruire dont le support est inclus dans T_0 , en notant $y = A x_0 + e$, où $y \in \mathbb{R}^m$ est le signal mesuré, et $e \in \mathbb{R}^m$ s'interprète comme un bruit, on définit :

$$\hat{x} := \begin{cases} (A_{T_0}^T A_{T_0})^{-1} A_{T_0}^T y & \text{sur } T_0, \\ 0 & \text{sur } T_0^c \end{cases}$$

Alors en notant $B_{T_0} := (A_{T_0}^T A_{T_0})^{-1} A_{T_0}^T \in \mathcal{M}_{|T_0|,m}(\mathbb{R})$ et $C_{T_0} := \|B_{T_0}\|$, où $\|\cdot\|$ désigne la norme d'opérateur sur $\mathcal{M}_{|T_0|,m}(\mathbb{R})$, on a :

$$\|\hat{x} - x_0\|_2 \leq C_{T_0} \|e\|_2$$

En effet, comme le support de x_0 est inclus dans T_0 ,

$$B_{T_0} y = (A_{T_0}^T A_{T_0})^{-1} A_{T_0}^T A x_0 + B_{T_0} e = (A_{T_0}^T A_{T_0})^{-1} A_{T_0}^T A_{T_0} x_{0,T_0} + B_{T_0} e = x_{0,T_0} + B_{T_0} e$$

Ainsi, comme \hat{x} et x_0 sont tous les deux nuls sur T_0^c , on a :

$$\|\hat{x} - x_0\|_2 = \|B_{T_0} y - x_{0,T_0}\|_2 = \|B_{T_0} e\|_2 \leq C_{T_0} \|e\|_2$$

On constate alors que l'erreur de reconstruction est également linéaire en l'erreur de mesure, et qu'il n'est pas possible de faire mieux, car il existe toujours $e \in \mathbb{R}^m$ tel que $\|B_{T_0} e\|_2 = C_{T_0} \|e\|_2$.

De façon plus générale, on considère une matrice $A \in \mathcal{M}_{m,n}(\mathbb{R})$, un ensemble $T \subset \llbracket 1, n \rrbracket$ non vide quelconque ainsi qu'un signal $x_0 \in \mathbb{R}^n$ inconnu, dont le support est inclus dans T . On souhaite reconstruire x_0 à partir de $y = A x_0 + e$, où $e \in \mathbb{R}^m$ est l'erreur de mesure. Pour cela, on considère un "algorithme de reconstruction", qui est une fonction $\mathcal{F} : \mathbb{R}^m \rightarrow \mathbb{R}^n$. Alors l'erreur de reconstruction par un tel algorithme est au mieux linéaire en $\|e\|_2$, comme le montre la proposition suivante :

Proposition 1. *Soit $A \in \mathcal{M}_{m,n}(\mathbb{R})$, il existe une constante $C > 0$ (dépendant seulement de A) telle que pour $\epsilon > 0$, pour tout $T \subset \llbracket 1, n \rrbracket$ non vide et pour tout algorithme de construction $\mathcal{F} : \mathbb{R}^m \rightarrow \mathbb{R}^n$, il existe $x_0 \in \mathbb{R}^n$ à support inclus dans T et $e \in \mathbb{R}^m$ vérifiant :*

$$\|e\|_2 \leq \epsilon \quad \text{et} \quad \|\mathcal{F}(A x_0 + e) - x_0\|_2 \geq C \epsilon$$

Démonstration. Posons $C := \frac{1}{2} \|A\|^{-1}$, où $\|\cdot\|$ désigne la norme d'opérateur sur $\mathcal{M}_{m,n}(\mathbb{R})$, et fixons un $\epsilon > 0$ et un algorithme de reconstruction $\mathcal{F} : \mathbb{R}^m \rightarrow \mathbb{R}^n$.

Soit $x_1 \in \mathbb{R}^n$ de support inclus dans T , ainsi que $h \in \mathbb{R}^n$ de support inclus dans T et vérifiant $\|h\|_2 = \|A\|^{-1} \epsilon$. On pose alors $e_1 = A h$ et $x_2 = x_1 + h$, de façon à avoir :

$$\begin{aligned} 2 C \epsilon &= \|h\|_2 = \|x_1 - x_2\|_2 \\ &\leq \|x_1 - \mathcal{F}(A x_2)\|_2 + \|\mathcal{F}(A x_2) - x_2\|_2 \\ &= \|x_1 - \mathcal{F}(A x_1 + e_1)\|_2 + \|\mathcal{F}(A x_2) - x_2\|_2 \end{aligned}$$

On a donc nécessairement :

$$\|x_1 - \mathcal{F}(A x_1 + e_1)\|_2 \geq C \epsilon \quad \text{ou} \quad \|\mathcal{F}(A x_2) - x_2\|_2 \geq C \epsilon$$

Dans le premier cas, on prend $x_0 = x_1$ et $e = e_1$ afin d'avoir :

$$\|\mathcal{F}(Ax_0 + e) - x_0\|_2 \geq C \epsilon$$

$$\|e\|_2 = \|Ah\|_2 \leq \|A\| \|h\|_2 = \|A\| \|A\|^{-1} \epsilon = \epsilon$$

Dans le deuxième cas, on prend $x_0 = x_2$ et $e = 0$. Les inégalités recherchées sont également vérifiées. \square

2 Matrices aléatoires gaussiennes

Maintenant que nous savons quelle condition imposer sur une matrice de mesure afin que celle-ci permette une bonne reconstruction de signaux parcimonieux, nous souhaitons trouver des matrices qui vérifient en effet cette condition. Cependant, il est en pratique très compliqué de déterminer si une matrice vérifie la condition RIP puisque ses constantes d'isométrie restreintes sont difficiles à calculer. Compte tenu de la taille des matrices que nous étudions, il n'est pas question de tester des matrices pour vérifier qu'elles satisfont bien la condition RIP. Une stratégie alternative consiste donc à considérer des matrices aléatoires dont les coefficients suivent des lois connues, et d'imposer des conditions sur la parcimonie des vecteurs à reconstruire pour que la condition RIP soit vérifiée avec une grande probabilité.

Dans la suite, nous nous intéressons à des matrices aléatoires dont les coefficients suivent des lois gaussiennes. Ceci est justifié par le fait que la loi gaussienne est une loi naturelle dont on connaît de nombreuses propriétés. De plus, de telles matrices aléatoires sont faciles à implémenter.

Nous montrons que si la parcimonie des vecteurs à reconstruire n'est pas trop grande alors la probabilité qu'une matrice aléatoire gaussienne vérifie la condition RIP est très proche de 1. Notons que des résultats similaires existent pour des matrices aléatoires suivant d'autres lois.

On fixe dans la suite un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ ainsi que deux entiers naturels m et n vérifiant $m \leq n$.

2.1 Probabilité de vérifier RIP

Définition 2. Soit $A : \Omega \rightarrow \mathcal{M}_{p,q}(\mathbb{R})$ une matrice aléatoire, A est une matrice aléatoire gaussienne si ses coefficients $a_{i,j} : \Omega \rightarrow \mathbb{R}$ pour $(i, j) \in \llbracket 1, p \rrbracket \times \llbracket 1, q \rrbracket$ sont des variables aléatoires indépendantes et de même loi $\mathcal{N}\left(0, \frac{1}{p}\right)$.

Théorème 3 (Inégalités de concentration des valeurs singulières).

Soient p, q des entiers naturels non nuls vérifiant $p \geq q$.

Soit $A : \Omega \rightarrow \mathcal{M}_{p,q}(\mathbb{R})$ une matrice aléatoire gaussienne. En posant $\lambda_{\min}(A) : \Omega \rightarrow \mathbb{R}$ (resp. $\lambda_{\max}(A) : \Omega \rightarrow \mathbb{R}$) sa plus petite (resp. plus grande) valeur singulière, $\lambda_{\min}(A)$ et $\lambda_{\max}(A)$ sont des variables aléatoires qui vérifient les inégalités de concentration pour tout réel positif r :

$$\mathbb{P}\left(\lambda_{\min}(A) \leq 1 - \sqrt{\frac{q}{p}} - r\right) \leq \exp\left(-\frac{pr^2}{2}\right) \quad (\text{C1})$$

$$\mathbb{P}\left(\lambda_{\max}(A) \geq 1 + \sqrt{\frac{q}{p}} + r\right) \leq \exp\left(-\frac{pr^2}{2}\right) \quad (\text{C2})$$

Ce théorème affirme que pour des matrices gaussiennes avec beaucoup de lignes (p grand), les valeurs singulières sont concentrées dans un petit voisinage de l'intervalle $\left[1 - \sqrt{\frac{q}{p}}, 1 + \sqrt{\frac{q}{p}}\right]$ avec une forte probabilité. Or, d'après la remarque (1) de la section 1.2., plus les valeurs singulières d'une matrice sont concentrées et plus ses constantes d'isométrie sont petites. On comprend donc que ce théorème va nous aider à minorer la probabilité de vérifier la condition RIP.

Ce théorème est démontré dans la sous-section 2.3.

Lemme 1. Soient a, b des entiers naturels avec $a \leq b$, on a :

$$\binom{b}{a} \leq \exp\left(b H\left(\frac{a}{b}\right)\right)$$

où H est la "fonction d'entropie binaire" :

$$H : x \in [0, 1] \mapsto -x \ln x - (1-x) \ln(1-x), \quad H(0) = H(1) = 0$$

Démonstration. Remarquons d'abord que l'inégalité à démontrer peut se réécrire :

$$a^a (b-a)^{(b-a)} \binom{b}{a} \leq b^b$$

Nous allons faire appel à un argument combinatoire pour l'obtenir. En notant $\mathcal{P}_a(\llbracket 1, b \rrbracket)$ l'ensemble des parties à a éléments de $\llbracket 1, b \rrbracket$, on définit une application

$$\Phi : \llbracket 1, a \rrbracket^{\llbracket 1, a \rrbracket} \times \llbracket 1, b-a \rrbracket^{\llbracket 1, b-a \rrbracket} \times \mathcal{P}_a(\llbracket 1, b \rrbracket) \longrightarrow \llbracket 1, b \rrbracket^{\llbracket 1, b \rrbracket}$$

par : si $f \in \llbracket 1, a \rrbracket^{\llbracket 1, a \rrbracket}$, $g \in \llbracket 1, b-a \rrbracket^{\llbracket 1, b-a \rrbracket}$ et $X \in \mathcal{P}_a(\llbracket 1, b \rrbracket)$, en notant $X = \{x_1, \dots, x_a\}$ avec $x_1 < \dots < x_a$ et $X^c = \{y_1, \dots, y_{b-a}\}$ avec $y_1 < \dots < y_{b-a}$,

$$\begin{cases} \forall i \in \llbracket 1, a \rrbracket, & \Phi(f, g, X)[x_i] = f(i) \\ \forall j \in \llbracket 1, b-a \rrbracket, & \Phi(f, g, X)[y_j] = g(j) + a \end{cases}$$

Alors si $h = \Phi(f, g, X)$, on remarque que $X = h^{-1}(\llbracket 1, a \rrbracket)$ puis que f et g peuvent être reconstruites en connaissant h , en ordonnant les éléments de X et X^c . On en déduit donc que Φ est injective, d'où l'inégalité souhaitée. \square

Théorème 4. *Sous l'hypothèse $n \geq 3m$, il existe des réels $K >$ et $\alpha > 0$ tels que pour toute matrice aléatoire gaussienne $A : \Omega \longrightarrow \mathcal{M}_{m,n}(\mathbb{R})$ et pour tout $S \in \mathbb{N}$ vérifiant*

$$S \leq \alpha \frac{m}{\ln(\frac{n}{m})}$$

on a :

$$\mathbb{P} \left(\delta_S(A) < \frac{1}{2} \right) \geq 1 - 2e^{-Km}$$

Les valeurs obtenues dans la démonstration sont : $K \approx 0.00316$ et $\alpha \approx 0.000156$.

Démonstration. Fixons $A : \Omega \longrightarrow \mathcal{M}_{m,n}(\mathbb{R})$ ainsi que $S \in \mathbb{N}$. On cherche à minorer la probabilité de l'événement $\{\delta_S(A) < 1/2\}$, quitte à imposer certaines conditions sur S . Pour alléger les notations, on désignera $\delta_S(A)$ par δ_S

Posons $\mathcal{T}_S := \{T \subset \llbracket 1, n \rrbracket : |T| = S\}$. Pour $T \in \mathcal{T}_S$ on note E_T l'événement

$$E_T := \left\{ \lambda_{\min}(A_T) \leq \sqrt{\frac{1}{2}} \right\} \cup \left\{ \lambda_{\max}(A_T) \geq \sqrt{\frac{3}{2}} \right\}$$

D'après la Remarque 1.2.1, on a alors :

$$\{\delta_S \geq 1/2\} = \bigcup_{T \in \mathcal{T}_S} E_T$$

Ainsi, nous voulons majorer la probabilité $\mathbb{P} \left(\bigcup_{T \in \mathcal{T}_S} E_T \right)$.

Fixons deux constantes $c, c' > 0$ qui vérifient :

$$\sqrt{c'} + \sqrt{2c} = \sqrt{\frac{3}{2}} - 1$$

On peut prendre par exemple :

$$c' = \frac{1}{4} \left(\sqrt{\frac{3}{2}} - 1 \right)^2, \quad c = \frac{1}{8} \left(\sqrt{\frac{3}{2}} - 1 \right)^2$$

On suppose dans la suite que l'on a : $S \leq c'm$, de façon à avoir :

$$\sqrt{\frac{S}{m}} + \sqrt{2c} \leq \sqrt{\frac{3}{2}} - 1 \leq 1 - \sqrt{\frac{1}{2}}$$

En fixant $T \in \mathcal{T}_S$, les inégalités de concentrations du Théorème 3 donnent alors :

$$\begin{aligned} \mathbb{P}\left(\lambda_{\min}(A_T) \leq \sqrt{\frac{1}{2}}\right) &\leq \mathbb{P}\left(\lambda_{\min}(A) \leq 1 - \sqrt{\frac{S}{m}} - \sqrt{2c}\right) \leq \exp(-cm) \\ \mathbb{P}\left(\lambda_{\max}(A_T) \geq \sqrt{\frac{3}{2}}\right) &\leq \mathbb{P}\left(\lambda_{\max}(A) \geq 1 + \sqrt{\frac{S}{m}} + \sqrt{2c}\right) \leq \exp(-cm) \end{aligned}$$

d'où

$$\mathbb{P}(E_T) \leq 2 \exp(-cm)$$

On a alors :

$$\mathbb{P}\left(\bigcup_{T \in \mathcal{T}_S} E_T\right) \leq 2 \exp(-cm) |\mathcal{T}_S|$$

c'est-à-dire

$$\mathbb{P}(\delta_S \geq 1/2) \leq 2 \exp(-cm) \binom{n}{S}$$

En utilisant l'inégalité :

$$\forall x \in [0, 1], \quad -(1-x) \ln(1-x) \leq x$$

et d'après le lemme précédent, on a :

$$\binom{n}{S} \leq \exp\left(-S \ln\left(\frac{S}{n}\right) + S\right)$$

Ainsi

$$\ln[\mathbb{P}(\delta_S \geq 1/2)] \leq \ln(2) - cm + S \left(\ln\left(\frac{n}{S}\right) + 1\right) \quad (3)$$

On cherche une condition sur S pour que le terme $S \left(\ln\left(\frac{n}{S}\right) + 1\right)$ soit plus petit que $\frac{c}{2}m$ ce qui permettra de conclure.

Supposons alors que S vérifie :

$$S \leq \alpha \frac{m}{\ln\left(\frac{n}{m}\right)}$$

où $0 < \alpha < 1$ est un paramètre que l'on spécifiera plus tard.

On a alors les inégalités :

$$0 < S \leq \alpha \frac{m}{\ln\left(\frac{n}{m}\right)} < \frac{n}{3}$$

Ceci permet d'écrire :

$$\begin{aligned} S \left(\ln\left(\frac{n}{S}\right) + 1\right) &\leq 2S \ln\left(\frac{n}{S}\right) \\ &\leq 2\alpha \frac{m}{\ln\left(\frac{n}{m}\right)} \ln\left(\frac{n}{\alpha m} \ln\left(\frac{n}{m}\right)\right) \\ &\leq 2\alpha m \left(1 + \ln\left(\frac{1}{\alpha}\right) + \frac{\ln\left(\ln\left(\frac{n}{m}\right)\right)}{\ln\left(\frac{n}{m}\right)}\right) \\ &\leq 2\alpha \left(2 + \ln\left(\frac{1}{\alpha}\right)\right) m \end{aligned}$$

On choisit maintenant α comme étant l'unique solution dans $]0, 1[$ de l'équation : $4x(2 - \ln(x)) = c$.
On obtient alors :

$$S \left(\ln \left(\frac{n}{S} \right) + 1 \right) \leq \frac{c}{2} m$$

Puis, en reprenant l'inégalité (3),

$$\ln[\mathbb{P}(\delta_S \geq 1/2)] \leq \ln(2) - \frac{c}{2} m$$

c'est-à-dire, en posant $K := \frac{c}{2} \approx 0.00316$,

$$\mathbb{P}(\delta_S < 1/2) \geq 1 - 2e^{-Km}$$

En résumé, nous avons prouvé que sous les conditions :

$$\begin{aligned} S &\leq c' m \\ S &\leq \alpha \frac{m}{\ln \left(\frac{n}{m} \right)} \end{aligned}$$

on a :

$$\mathbb{P}(\delta_S < 1/2) \geq 1 - 2e^{-Km}$$

Or, $a \approx 0.000156$ et $c' \approx 0.0127$ donc $\alpha \leq c'$ et la deuxième condition implique la première, ce qui achève la preuve. \square

2.2 Raffinement du résultat

Dans la pratique, l'entier m correspond au nombre de mesures effectuées et on peut supposer qu'il est assez grand ($m \geq 2000$). Le théorème affirme que si S est assez petit, la probabilité que A vérifie RIP pour S est supérieure à 0,99.

Malheureusement, la valeur de α calculée ici est très petite et restreint S de façon aberrante. Par exemple, pour récupérer un vecteur 1-parcimonieux, il faudrait au moins effectuer 25 641 mesures (et que le nombre de composantes du signal dépasse 76 923). Ce résultat est décevant car il reste très inférieur à ce qu'on observe dans la réalité, où la constante α autorisée empiriquement est beaucoup plus grande.

On peut néanmoins améliorer cette constante en modifiant la condition sur S , et en restreignant le nombre d'inégalités utilisées.

En effet, on peut donner une condition sur S qui soit "optimale" dans le sens où elle ne repose que sur trois inégalités fondamentales :

1. Les deux inégalités de concentrations, que l'on ne peut améliorer sans complexifier notablement la preuve,
2. L'inégalité de majoration du coefficient binomial, qui est satisfaisante pour de grandes valeurs et qui ne peut être améliorée sans rendre les calculs très fastidieux,
3. L'inégalité $\mathbb{P} \left(\bigcup_{T \in \mathcal{T}_S} E_T \right) \leq \sum_{T \in \mathcal{T}_S} \mathbb{P}(E_T)$, que l'on ne peut améliorer sans des considérations très précises (donc compliquées) sur les constantes d'isométrie restreintes des matrices aléatoires gaussiennes.

Avec *seulement* ces inégalités, on obtient le théorème :

Théorème 5. *Il existe un réel $K > 0$ tel que pour toute matrice aléatoire gaussienne $A : \Omega \rightarrow \mathcal{M}_{m,n}(\mathbb{R})$ et pour tout $S \in \mathbb{N}$ vérifiant $S \leq \frac{n}{2}$, on a :*

$$\mathbb{P} \left(\delta_S(A) < \frac{1}{2} \right) \geq 1 - 2e^{-Km}$$

sous la condition

$$H(\xi) \leq K\eta$$

où $\xi := \frac{S}{n}$, $\eta := \frac{m}{n}$ et $H : x \in [0, 1] \mapsto -x \ln(x) - (1-x) \ln(1-x)$.

La valeur de K obtenue dans la démonstration est : $K \approx 0.009$.

Démonstration. D'après la démonstration du théorème précédent, en considérant $c, c' > 0$ vérifiant :

$$\sqrt{c'} + \sqrt{2c} = \sqrt{\frac{3}{2}} - 1$$

et en supposant $S \leq c'm$, i.e. $\xi \leq c'\eta$, on a :

$$\ln [\mathbb{P}(\delta_S \geq 1/2)] \leq \ln 2 - cm + nH(\xi)$$

Alors

$$\mathbb{P} \left(\delta_S < \frac{1}{2} \right) \geq 1 - 2e^{-\frac{c}{2}m} \quad (4)$$

dès que

$$H(\xi) \leq \frac{c}{2}\eta$$

Il ne nous reste plus qu'à fixer correctement les constantes c et c' afin de trouver la constante K de l'énoncé du théorème.

Les hypothèses à vérifier pour obtenir (4) deviennent alors :

$$2H(\xi) \leq c\eta \quad (5)$$

$$\xi \leq c'\eta \quad (6)$$

En choisissant convenablement les constantes c et c' , nous allons montrer que l'inégalité (5) implique l'inégalité (6). Dans ces conditions, en posant $K = \frac{c}{2}$, l'inégalité

$$H(\xi) \leq K\eta$$

permet bien d'obtenir le résultat :

$$\mathbb{P} \left(\delta_S < \frac{1}{2} \right) \geq 1 - 2e^{-Km}$$

Intéressons nous au choix de c et c' . On peut supposer $\xi \leq \frac{1}{2}$, ce qui n'est pas une condition très restrictive en pratique. Alors en imposant

$$2H(c') = c$$

comme la fonction H est concave et croissante sur $[0, \frac{1}{2}]$ et que $c' \leq \frac{1}{2}$, on a :

$$\forall y \in [0, 1], \forall x \in [0, c'], \frac{2H(x)}{c} \leq y \Rightarrow \frac{x}{c'} \leq y$$

donc comme $\eta \leq 1$, on a bien l'implication voulue. Les conditions sur c et c' sont alors :

$$\begin{cases} 2H(c') & = & c \\ \sqrt{c'} + \sqrt{2c} & = & \sqrt{\frac{3}{2}} - 1 \end{cases}$$

Ce système admet bien une solution (unique), qui vaut approximativement :

$$\begin{cases} c & \approx & 0.01815 \\ c' & \approx & 0.00117 \end{cases}$$

d'où $K \approx 0.00907$ convient. □

Compte tenu de la constante K du théorème, cette probabilité est très proche de 1 dès que m vaut quelques centaines. Par ailleurs, la condition à vérifier sur n , m et S montre que seules les fractions $\frac{S}{n}$ et $\frac{m}{n}$ importent. En pratique, la taille du signal de départ n est connue. Si l'on cherche à connaître le nombre de mesures minimales à effectuer connaissant la parcimonie S du signal initial, alors la condition du théorème s'avère utile. Mais dans le cas contraire, si l'on connaît une borne du nombre de mesures possibles et que l'on veut calculer la parcimonie maximale des signaux que l'on peut reconstruire, la formulation du théorème n'est pas la plus parlante. On lui préfère alors la formulation du corollaire suivant, qui est une amélioration du Théorème 4 :

Corollaire. *Avec les mêmes notations que dans le Théorème 4, il existe une constante $\alpha > 0$ telle que si S vérifie*

$$S \leq \alpha \frac{m}{1 + \ln\left(\frac{n}{m}\right)}$$

alors :

$$\mathbb{P}\left(\delta_S(A) < \frac{1}{2}\right) \geq 1 - 2e^{-Km}$$

La valeur de α obtenue dans la démonstration est : $\alpha \approx 0.00117$.

Preuve du Corollaire. Comme la fonction H est croissante sur $[0, \frac{1}{2}]$, si pour une certaine constante $\alpha > 0$ on a $S \leq \alpha \frac{m}{1 + \ln\left(\frac{n}{m}\right)}$, alors :

$$H(\xi) = H\left(\frac{S}{n}\right) \leq H\left(\frac{\alpha m}{n\left(1 + \ln\left(\frac{n}{m}\right)\right)}\right)$$

En posant $x := \frac{n}{m} \geq 1$, la condition

$$H(\xi) \leq K\eta$$

se réécrit :

$$xH\left(\frac{\alpha}{x(1 + \ln(x))}\right) \leq K$$

On cherche alors α de telle sorte que l'inégalité précédente soit vérifiée pour tout $x \geq 1$. Comme pour tout $y \in [0, 1]$, $-(1-y)\ln(1-y) \leq y$, on a :

$$xH\left(\frac{\alpha}{x(1 + \ln(x))}\right) \leq \frac{\alpha}{1 + \ln(x)} \left(\ln\left(\frac{x(1 + \ln(x))}{\alpha}\right) + 1\right) = G_\alpha(x)$$

On peut alors renforcer la condition sur α par :

$$G_\alpha(x) \leq K$$

Un calcul de dérivée montre que G_α est décroissante sur $[\exp(e\alpha - 1), +\infty[$. Ainsi, si $\alpha \leq \frac{1}{e} \approx 0.3679$, G_α est décroissante sur $[1, +\infty[$. Par ailleurs, $G_\alpha(1) = \alpha - \alpha \ln(\alpha)$ ce qui nous incite à prendre α la solution sur $[0, 1]$ de l'équation $\alpha - \alpha \ln(\alpha) = K$. Dans ce cas, $\alpha \approx 0.00117$ et on a donc :

$$\forall x \geq 1, G_\alpha(x) \leq K$$

ce qui achève la preuve. □

2.3 Démonstration des inégalités de concentrations

On consacre cette sous-section à la démonstration du Théorème 3 sur la concentration des valeurs singulières énoncé dans la partie 2.1.

Pour cela, on fixe des entiers naturels p, q vérifiant $p \geq q$ ainsi qu'une matrice aléatoire gaussienne $A : \Omega \rightarrow \mathcal{M}_{p,q}(\mathbb{R})$.

Dans la section 3, nous démontrons les inégalités sur les valeurs singulières de la matrice A :

Théorème.

$$1 - \sqrt{\frac{q}{p}} < \mathbb{E}[\lambda_{\min}(A)] \leq \mathbb{E}[\lambda_{\max}(A)] < 1 + \sqrt{\frac{q}{p}}$$

Dans la première partie de la section 4, nous démontrons le résultat suivant sur la concentration de fonction lipschitziennes autour de leur moyenne :

Théorème. *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application 1-lipschitzienne pour la norme euclidienne, en notant $\mathbb{E}[f] = \int f d\gamma_n$ sa moyenne pour la mesure gaussienne standard γ_n sur \mathbb{R}^n , on a pour tout réel $r \geq 0$:*

$$\begin{aligned} \gamma_n(f \geq \mathbb{E}[f] + r) &\leq \exp\left(-\frac{r^2}{2}\right) \\ \gamma_n(f \leq \mathbb{E}[f] - r) &\leq \exp\left(-\frac{r^2}{2}\right) \end{aligned}$$

Nous démontrons en Annexe que les applications λ_{\min} et λ_{\max} qui à une matrice de $\mathcal{M}_{p,q}(\mathbb{R})$ associent sa plus petite et sa plus grande valeur singulière sont 1-lipschitziennes pour la norme euclidienne sur $\mathcal{M}_{p,q}(\mathbb{R})$. Nous pouvons donc leur appliquer le théorème précédent pour obtenir pour tout $r \geq 0$:

$$\begin{aligned} \gamma_n(\lambda_{\max} \geq \mathbb{E}[\lambda_{\max}] + r) &\leq \exp\left(-\frac{r^2}{2}\right) \\ \gamma_n(\lambda_{\min} \leq \mathbb{E}[\lambda_{\min}] - r) &\leq \exp\left(-\frac{r^2}{2}\right) \end{aligned}$$

En notant μ_A la mesure image de A , compte tenu des résultats présentés en Annexe, on a :

$$\begin{aligned} \gamma_n(\lambda_{\max} \geq \mathbb{E}[\lambda_{\max}] + r) &= \mu_A\left(\frac{1}{\sqrt{p}} \cdot (\lambda_{\max} \geq \mathbb{E}[\lambda_{\max}] + r)\right) \\ &= \mu_A(\{M : \lambda_{\max}(\sqrt{p}M) \geq \sqrt{p}\mathbb{E}_{\mathbb{P}}[\lambda_{\max}(A)] + r\}) \\ &= \mu_A\left(\lambda_{\max}(M) \geq \mathbb{E}_{\mathbb{P}}[\lambda_{\max}(A)] + \frac{r}{\sqrt{p}}\right) \\ &= \mathbb{P}\left(\lambda_{\max}(A) \geq \mathbb{E}_{\mathbb{P}}[\lambda_{\max}(A)] + \frac{r}{\sqrt{p}}\right) \end{aligned}$$

On obtient de la même façon :

$$\gamma_n(\lambda_{\min} \leq \mathbb{E}[\lambda_{\min}] - r) = \mathbb{P}\left(\lambda_{\min}(A) \leq \mathbb{E}_{\mathbb{P}}[\lambda_{\min}(A)] - \frac{r}{\sqrt{p}}\right)$$

On en déduit alors les deux inégalités de concentration suivantes :

$$\begin{aligned}\mathbb{P}(\lambda_{\max}(A) \geq \mathbb{E}[\lambda_{\max}(A)] + r) &\leq \exp\left(-\frac{pr^2}{2}\right) \\ \mathbb{P}(\lambda_{\min}(A) \geq \mathbb{E}[\lambda_{\max}(A)] - r) &\leq \exp\left(-\frac{pr^2}{2}\right)\end{aligned}$$

Enfin, d'après les bornes sur les espérances des valeurs singulières de A , on obtient les inclusions :

$$\begin{aligned}\left\{\lambda_{\max}(A) \geq 1 + \sqrt{\frac{q}{p}} + r\right\} &\subset \{\lambda_{\max}(A) \geq \mathbb{E}[\lambda_{\max}(A)] + r\} \\ \left\{\lambda_{\min}(A) \geq 1 - \sqrt{\frac{q}{p}} - r\right\} &\subset \{\lambda_{\min}(A) \geq \mathbb{E}[\lambda_{\max}(A)] - r\}\end{aligned}$$

Ce qui permet d'aboutir aux inégalités recherchées :

$$\mathbb{P}\left(\lambda_{\min}(A) \leq 1 - \sqrt{\frac{q}{p}} - r\right) \leq \exp\left(-\frac{pr^2}{2}\right) \quad (\text{C1})$$

$$\mathbb{P}\left(\lambda_{\max}(A) \geq 1 + \sqrt{\frac{q}{p}} + r\right) \leq \exp\left(-\frac{pr^2}{2}\right) \quad (\text{C2})$$

Les deux parties suivantes du mémoire sont donc consacrées à la démonstration des deux théorèmes énoncés précédemment. Ces parties sont plus techniques et présentent quelques résultats généraux et importants en théorie des probabilités.

3 Bornes pour les espérances des valeurs singulières

Soient p, q des entiers naturels non nuls vérifiant $p \geq q$ et $A : \Omega \rightarrow \mathcal{M}_{p,q}(\mathbb{R})$ une matrice aléatoire gaussienne. Comme les applications λ_{\min} et λ_{\max} sont lipschitziennes (cf. les propositions en Annexe), les variables aléatoires $\lambda_{\min}(A)$ et $\lambda_{\max}(A) : \Omega \rightarrow \mathbb{R}$ sont mesurables. L'objectif de cette section est de montrer les inégalités :

$$1 - \sqrt{\frac{q}{p}} < \mathbb{E}[\lambda_{\min}(A)] \leq \mathbb{E}[\lambda_{\max}(A)] < 1 + \sqrt{\frac{q}{p}}$$

En posant pour tout $(u, v) \in S^{p-1} \times S^{q-1}$:

$$\begin{aligned} X_{u,v} : \mathcal{M}_{p,q}(\mathbb{R}) &\longrightarrow \mathbb{R} \\ M &\longmapsto \langle Mv, u \rangle \end{aligned}$$

on peut écrire (cf. Annexe) :

$$\begin{aligned} \lambda_{\min} &= - \max_{v \in S^{q-1}} \min_{u \in S^{p-1}} X_{u,v} \\ \lambda_{\max} &= \max_{(u,v) \in S^{p-1} \times S^{q-1}} X_{u,v} \end{aligned}$$

En notant μ_A la mesure image de A , on se ramène à prouver :

$$\begin{aligned} \mathbb{E}_{\mu_A} \left[\max_{v \in S^{q-1}} \min_{u \in S^{p-1}} X_{u,v} \right] &< -1 + \sqrt{\frac{q}{p}} \\ \mathbb{E}_{\mu_A} \left[\max_{(u,v) \in S^{p-1} \times S^{q-1}} X_{u,v} \right] &< 1 + \sqrt{\frac{q}{p}} \end{aligned}$$

Le raisonnement tient alors en 3 étapes :

1. On donne une façon générale d'obtenir de telles majorations (Inégalités de Slepian-Gordon),
2. On voit pourquoi on peut l'appliquer dans le cas qui nous intéresse,
3. On en déduit le résultat.

3.1 Inégalités de Slepian-Gordon

Définition 3. Un vecteur aléatoire $X : \omega \in \Omega \mapsto (X_1(\omega), \dots, X_n(\omega)) \in \mathbb{R}^n$ est un vecteur aléatoire gaussien si pour tout $(a_1, \dots, a_n) \in \mathbb{R}^n$, $Z = a_1 X_1 + \dots + a_n X_n$ est une variable aléatoire gaussienne. Il est dit de moyenne nulle si pour tout $i \in \llbracket 1, n \rrbracket$, $\mathbb{E}[X_i] = 0$.

Définition 4. Si $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ est un vecteur aléatoire, sa matrice de covariance est la matrice $\Gamma_X := (\text{Cov}(X_i, X_j))_{i,j \in \llbracket 1, n \rrbracket}$, où $\text{Cov}(X_i, X_j) := \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j]$.

On rappelle un résultat important de réduction des vecteurs aléatoires gaussiens :

Proposition 2. Soit $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ un vecteur aléatoire gaussien de moyenne nulle et Γ^X sa matrice de covariance, alors il existe $r \in \llbracket 0, n \rrbracket$ ainsi que des matrices $O \in \text{O}_n(\mathbb{R})$ et $\Delta, \tilde{\Delta} \in \mathcal{M}_n(\mathbb{R})$ diagonales telles que $O\Delta^2 O^T = \Gamma^X$, $\Delta\tilde{\Delta} = I_r$ et $Y := \tilde{\Delta} O^T X$ est de la forme $Y = (Y_1, \dots, Y_r, 0, \dots, 0)$, où les Y_i sont des variables aléatoires indépendantes et de même loi $\mathcal{N}(0, 1)$. De plus, $X = O\Delta Y$.

On en déduit un résultat utile d'intégration par parties sur les vecteurs aléatoires gaussiens :

Lemme (Intégration par parties). Soit $F : \mathbb{R}^n \rightarrow \mathbb{R}$ une application de classe \mathcal{C}^1 à croissance modérée, c'est-à-dire vérifiant :

$$\forall a > 0, \lim_{x \rightarrow \infty} F(x) e^{-a\|x\|^2} = 0$$

On suppose également que les dérivées partielles de F sont à croissance modérée.

Soit $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ un vecteur aléatoire gaussien de moyenne nulle, alors pour tout $i \in \llbracket 1, n \rrbracket$,

$$\mathbb{E} [X_i F(X)] = \sum_{j=1}^n \mathbb{E} [X_i X_j] \mathbb{E} \left[\frac{\partial F}{\partial x_j}(X) \right]$$

De plus, si $g : \mathbb{R}^n \rightarrow \mathbb{R}$ une application de classe \mathcal{C}^2 dont les dérivées partielles sont à croissance modérée, on a :

$$\mathbb{E} [\nabla g(X)^T X] = \text{tr} (\Gamma^X \mathbb{E} [\nabla^2 g(X)])$$

Preuve du lemme. Fixons $X = (X_1, \dots, X_n) \in \mathbb{R}^n$ un vecteur aléatoire gaussien de moyenne nulle. En général, il n'est pas possible d'exprimer la loi de X comme une densité contre la mesure de Lebesgue. On procède donc en réduisant X comme dans la proposition précédente : fixons les matrices $U = O\Delta$ et $V = \tilde{\Delta}O^T$ correspondantes. En posant $\tilde{F} : y \in \mathbb{R} \mapsto F(Uy)$, \tilde{F} est toujours de classe \mathcal{C}^1 à croissance modérée, ce qui justifiera l'existence des intégrales que nous allons écrire.

Alors comme les coefficients non nuls de $Y = VX = (Y_1, \dots, Y_r, 0, \dots, 0)$ sont indépendants et de même loi $\mathcal{N}(0, 1)$, la loi de (Y_1, \dots, Y_r) admet pour densité contre la mesure de Lebesgue :

$$g_r : x \in \mathbb{R}^r \mapsto \frac{1}{\sqrt{2\pi}^r} \exp\left(-\frac{x_1^2 + \dots + x_r^2}{2}\right)$$

En remarquant que \tilde{F} ne dépend que des r premières coordonnées de $y \in \mathbb{R}^n$, puisque les $n - r$ dernières colonnes de U sont nulles, on a pour tout $i \in \llbracket 1, r \rrbracket$,

$$\begin{aligned} \mathbb{E} [Y_i \tilde{F}(Y)] &= \int_{\Omega} Y_i \tilde{F}(Y) d\mathbb{P} = \int_{\mathbb{R}^r} y_i \tilde{F}(y) g_r(y) dy \\ &= \int_{\mathbb{R}^{r-1}} \left(\int_{\mathbb{R}} \tilde{F}(y) y_i \exp\left(-\frac{y_i^2}{2}\right) dy_i \right) g_r(y_1, \dots, y_{i-1}, 0, y_{i+1}, \dots, y_r) \prod_{j \neq i} dy_j \end{aligned}$$

Or, par intégration par parties,

$$\begin{aligned} \int_{\mathbb{R}} \tilde{F}(y) y_i \exp\left(-\frac{y_i^2}{2}\right) dy_i &= \left[-\tilde{F}(y_1, \dots, y_i, \dots, y_r) \exp\left(-\frac{y_i^2}{2}\right) \right]_{y_i=-\infty}^{+\infty} - \int_{\mathbb{R}} -\frac{\partial \tilde{F}}{\partial y_i}(y) \exp\left(-\frac{y_i^2}{2}\right) dy_i \\ &= \int_{\mathbb{R}} \frac{\partial \tilde{F}}{\partial y_i}(y) \exp\left(-\frac{y_i^2}{2}\right) dy_i \end{aligned}$$

Ainsi,

$$\begin{aligned} \mathbb{E} [Y_i \tilde{F}(Y)] &= \int_{\mathbb{R}^{r-1}} \left(\int_{\mathbb{R}} \frac{\partial \tilde{F}}{\partial y_i}(y) \exp\left(-\frac{y_i^2}{2}\right) dy_i \right) g_r(y_1, \dots, y_{i-1}, 0, y_{i+1}, \dots, y_r) \prod_{j \neq i} dy_j \\ &= \int_{\mathbb{R}^r} \frac{\partial \tilde{F}}{\partial y_i}(y) g_r(y) dy = \int_{\Omega} \frac{\partial \tilde{F}}{\partial y_i}(Y) d\mathbb{P} = \mathbb{E} \left[\frac{\partial \tilde{F}}{\partial y_i}(Y) \right] \end{aligned}$$

De plus, les $n - r$ dernières colonnes de U étant nulles, on a : $\forall i \geq r + 1$, $\frac{\partial \tilde{F}}{\partial y_i} = 0$, donc l'égalité précédente reste vérifiée pour $i \geq r + 1$.

Pour $i \in \llbracket 1, n \rrbracket$, en notant $L_{U,i}$ la i -ème ligne de U , on obtient :

$$\begin{aligned} \mathbb{E}[X_i F(X)] &= \mathbb{E}\left[L_{U,i} Y \tilde{F}(Y)\right] = L_{U,i} \mathbb{E}\left[\nabla \tilde{F}(Y)\right] = L_{U,i} \mathbb{E}\left[U^T \nabla F(X)\right] = \mathbb{E}\left[(L_{U,i} U^T) \nabla F(X)\right] \\ &= \sum_{j=1}^n \mathbb{E}[X_i X_j] \mathbb{E}\left[\frac{\partial F}{\partial x_j}(X)\right] \end{aligned}$$

puisque $UU^T = O\Delta O^T = \Gamma^X$, donc la première égalité est démontrée.

La deuxième s'en déduit facilement, car en l'appliquant aux dérivées partielles de g , on obtient :

$$\begin{aligned} \mathbb{E}[\nabla g(X)^T X] &= \sum_{i=1}^n \mathbb{E}\left[X_i \frac{\partial g}{\partial x_i}(X)\right] \\ &= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}[X_i X_j] \mathbb{E}\left[\frac{\partial^2 g}{\partial x_i \partial x_j}(X)\right] \\ &= \text{tr}(\Gamma^X \mathbb{E}[\nabla^2 g(X)]) \end{aligned}$$

□

Théorème 6 (Inégalités de Slepian-Gordon). *Soit (Ω, F, \mathbb{P}) un espace de probabilité et $X, Y : \Omega \rightarrow \mathbb{R}^n$ des vecteurs aléatoires gaussiens de moyenne nulle. On pose pour tout $i, j \in \llbracket 1, n \rrbracket$,*

$$\delta_{i,j}^X = \mathbb{E}\left[(X_i - X_j)^2\right] \quad \delta_{i,j}^Y = \mathbb{E}\left[(Y_i - Y_j)^2\right]$$

1. Si $\forall (i, j) \in \llbracket 1, n \rrbracket^2$, $\delta_{i,j}^X \leq \delta_{i,j}^Y$, alors :

$$\mathbb{E}\left[\max_{i \in \llbracket 1, n \rrbracket} X_i\right] \leq \mathbb{E}\left[\max_{i \in \llbracket 1, n \rrbracket} Y_i\right]$$

2. Si $(T_i)_{i \in \llbracket 1, k \rrbracket}$ est un recouvrement de $\llbracket 1, n \rrbracket$ tel que :

- Pour tout $i, i' \in \llbracket 1, k \rrbracket$, $i \neq i'$, pour tout $(j, j') \in T_i \times T_{i'}$, $\delta_{j,j'}^X \leq \delta_{j,j'}^Y$,
- Pour tout $i \in \llbracket 1, k \rrbracket$, pour tout $(j, j') \in T_i^2$, $\delta_{j,j'}^X \geq \delta_{j,j'}^Y$,

Alors :

$$\mathbb{E}\left[\max_{i \in \llbracket 1, k \rrbracket} \min_{j \in T_i} X_j\right] \leq \mathbb{E}\left[\max_{i \in \llbracket 1, k \rrbracket} \min_{j \in T_i} Y_j\right]$$

Démonstration. On démontre en détail le premier point. On donnera ensuite les idées essentielles pour obtenir le deuxième point.

• Exposons d'abord les idées du raisonnement :

1. On se donne une famille de fonctions de classe \mathcal{C}^∞ : $f_\lambda : \mathbb{R}^n \rightarrow \mathbb{R}$, $\lambda \in \mathbb{R}_+^*$ qui approche uniformément la fonction max quand λ tend vers $+\infty$.
2. On exhibe un chemin de vecteurs aléatoires gaussiens $(Z_t)_{t \in [0,1]}$ qui relie X à Y et tel que, pour tout $\lambda \in \mathbb{R}_+^*$, $t \mapsto \mathbb{E}[f_\lambda(Z_t)]$ est croissante.
3. On a en particulier que pour $\lambda \in \mathbb{R}_+^*$: $\mathbb{E}[f_\lambda(X)] \leq \mathbb{E}[f_\lambda(Y)]$. On obtient alors le résultat par passage à la limite quand $\lambda \rightarrow +\infty$.

• Remarquons que l'on peut supposer X et Y indépendants.

En effet, soit \tilde{X} et \tilde{Y} les deux vecteurs aléatoires sur $\Omega \times \Omega$ (muni de la mesure produit) définis par :

$$\forall (\omega_1, \omega_2) \in \Omega \times \Omega, \tilde{X}(\omega_1, \omega_2) = X(\omega_1), \tilde{Y}(\omega_1, \omega_2) = Y(\omega_2)$$

Les vecteurs \tilde{X} et \tilde{Y} sont indépendants (par définition de la mesure produit). On vérifie aisément que \tilde{X} et \tilde{Y} satisfont les hypothèses du théorème dès que X et Y les satisfont, ainsi que :

$$\mathbb{E} \left[\max_{i \in [1, n]} X_i \right] = \mathbb{E} \left[\max_{i \in [1, n]} \tilde{X}_i \right], \quad \mathbb{E} \left[\max_{i \in [1, n]} Y_i \right] = \mathbb{E} \left[\max_{i \in [1, n]} \tilde{Y}_i \right]$$

Dans la suite on suppose donc que X et Y sont indépendants.

- Pour tout $\lambda \in \mathbb{R}_+^*$, on pose :

$$\begin{aligned} f_\lambda : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \frac{1}{\lambda} \ln \left(\sum_{i=1}^n e^{\lambda x_i} \right) \end{aligned}$$

Alors chaque f_λ est de classe \mathcal{C}^∞ , de dérivées partielles à croissance modérée, car :

$$\nabla f_\lambda(x) = \frac{1}{\sum_{i=1}^n e^{\lambda x_i}} (e^{\lambda x_1}, \dots, e^{\lambda x_n})^T$$

On a de plus, pour tout $x \in \mathbb{R}^n$:

$$\max_{i \in [1, n]} x_i \leq f_\lambda(x) \leq \frac{1}{\lambda} \ln n + \max_{i \in [1, n]} x_i \quad (7)$$

ce qui montre que la famille (f_λ) , $\lambda \in \mathbb{R}_+^*$, approche uniformément la fonction max quand $\lambda \rightarrow +\infty$.

- On pose ensuite pour tout $t \in [0, 1]$:

$$Z_t = \sqrt{1-t} X + \sqrt{t} Y$$

Alors $(Z_t : \Omega \rightarrow \mathbb{R}^n)_{t \in [0, 1]}$ est un chemin de vecteurs aléatoires gaussiens d'extrémités $Z_0 = X$ et $Z_1 = Y$.

- Soit $\lambda \in \mathbb{R}_+^*$. Considérons l'application :

$$\begin{aligned} h : [0, 1] &\longrightarrow \mathbb{R} \\ t &\longmapsto \mathbb{E} [f_\lambda(Z_t)] \end{aligned}$$

on va montrer que :

1. h est bien définie,
2. h est continue sur $[0, 1]$, dérivable sur $]0, 1[$,
3. $\forall t \in]0, 1[, h'(t) \geq 0$.

Admettons ces trois points provisoirement. La croissance de h donne en particulier $h(0) \leq h(1)$ i.e. $\mathbb{E} [f_\lambda(X)] \leq \mathbb{E} [f_\lambda(Y)]$. Ceci est valable pour tout $\lambda > 0$. L'inégalité (7) permet alors de passer à la limite dans l'espérance pour $\lambda \rightarrow +\infty$ et on obtient :

$$\mathbb{E} \left[\max_{i \in [1, n]} X_i \right] \leq \mathbb{E} \left[\max_{i \in [1, n]} Y_i \right]$$

On consacre la fin de la démonstration à la preuve des trois points précédents. Fixons $\lambda \in \mathbb{R}_+^*$. Pour alléger les notations, on note désormais f à la place de f_λ .

1. Fixons $t \in [0, 1]$. Par l'inégalité (7), on a :

$$|f(Z_t)| \leq \frac{1}{\lambda} \ln n + \left| \max_{i \in \llbracket 1, n \rrbracket} Z_{t,i} \right| \leq \frac{1}{\lambda} \ln n + \sum_{i=1}^n |Z_{t,i}| \leq \frac{1}{\lambda} \ln n + \sum_{i=1}^n |X_i| + |Y_i|$$

Ainsi, comme les X_i et les Y_i sont intégrables, h est bien définie.

2. Posons :

$$F : \Omega \times [0, 1] \longrightarrow \mathbb{R} \\ (\omega, t) \longmapsto f(Z_t)(\omega) = f(\sqrt{1-t}X(\omega) + \sqrt{t}Y(\omega))$$

de façon à avoir :

$$\forall t \in [0, 1], h(t) = \int_{\Omega} F(\omega, t) d\mathbb{P}(\omega)$$

D'après la domination uniforme en t du point 1. et comme l'application $t \longrightarrow F(\omega, t)$ est continue sur $[0, 1]$ pour tout $\omega \in \Omega$, h est continue sur $[0, 1]$ d'après le théorème de continuité sous l'intégrale. Pour la dérivabilité :

◇ À ω fixé, $F(\omega, \cdot)$ est dérivable sur $]0, 1[$, de dérivée :

$$\frac{\partial F}{\partial t}(\omega, t) = \nabla f(Z_t(\omega))^T \left(\frac{Y(\omega)}{2\sqrt{t}} - \frac{X(\omega)}{2\sqrt{1-t}} \right)$$

◇ Fixons $\epsilon \in]0, 1[$. En posant $M := \frac{1}{2\sqrt{\epsilon}}$ on a :

$$\forall \omega \in \Omega, \forall t \in]\epsilon, 1 - \epsilon[, \left\| \frac{Y(\omega)}{2\sqrt{t}} - \frac{X(\omega)}{2\sqrt{1-t}} \right\|_{2, \mathbb{R}^n} \leq M(\|Y(\omega)\|_{2, \mathbb{R}^n} + \|X(\omega)\|_{2, \mathbb{R}^n})$$

Par ailleurs,

$$\forall x \in \mathbb{R}^n, \|\nabla f(x)\|_{2, \mathbb{R}^n}^2 = \frac{\sum_{i=1}^n e^{2\lambda x_i}}{(\sum_{i=1}^n e^{\lambda x_i})^2} \leq 1$$

L'inégalité de Cauchy-Schwarz permet alors d'obtenir :

$$\forall \omega \in \Omega, \forall t \in]\epsilon, 1 - \epsilon[, \left| \frac{\partial F}{\partial t}(\omega, t) \right| \leq M(\|Y(\omega)\|_{2, \mathbb{R}^n} + \|X(\omega)\|_{2, \mathbb{R}^n})$$

où le membre de droite ne dépend que de ω et est intégrable.

Les deux points montrent que h est dérivable avec dérivation sous l'intégrale sur $]\epsilon, 1 - \epsilon[$ pour tout $\epsilon > 0$, donc sur $]0, 1[$.

3. Fixons $t \in]0, 1[$. Par 2., on a :

$$h'(t) = \mathbb{E} \left[\nabla f(Z_t)^T \left(\frac{Y}{2\sqrt{t}} - \frac{X}{2\sqrt{1-t}} \right) \right]$$

Calculons d'abord la quantité $\mathbb{E} \left[\nabla f(Z_t)^T \frac{Y}{2\sqrt{t}} \right]$.

En notant μ_X (resp. μ_Y) la loi de X (resp. de Y), par indépendance de X et Y et d'après le théorème de Fubini, on peut écrire :

$$\begin{aligned}\mathbb{E} \left[\nabla f(Z_t)^T \frac{Y}{2\sqrt{t}} \right] &= \int_{\mathbb{R}^n \times \mathbb{R}^n} \nabla f(\sqrt{1-t}x + \sqrt{t}y)^T \frac{y}{2\sqrt{t}} d\mu_X \otimes d\mu_Y(x, y) \\ &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \nabla f(\sqrt{1-t}x + \sqrt{t}y)^T \frac{y}{2\sqrt{t}} d\mu_Y(y) \right) d\mu_X(x) \\ &= \int_{\mathbb{R}^n} \mathbb{E} \left[\nabla f(\sqrt{1-t}x + \sqrt{t}Y)^T \frac{Y}{2\sqrt{t}} \right] d\mu_X(x)\end{aligned}$$

On utilise alors le lemme d'intégration par parties pour la fonction $g : y \mapsto \frac{1}{\sqrt{t}}f(\sqrt{1-t}x + \sqrt{t}y)$:

$$\begin{aligned}\mathbb{E} \left[\nabla f(\sqrt{1-t}x + \sqrt{t}Y)^T \frac{Y}{2\sqrt{t}} \right] &= \frac{1}{2\sqrt{t}} \mathbb{E} [\nabla g(Y)^T Y] \\ &= \frac{1}{2\sqrt{t}} \text{tr}(\Gamma^Y \mathbb{E}[\nabla^2 g(Y)]) \\ &= \frac{1}{2\sqrt{t}} \text{tr}(\Gamma^Y \mathbb{E}[\sqrt{t} \nabla^2 f(\sqrt{1-t}x + \sqrt{t}Y)]) \\ &= \frac{1}{2} \mathbb{E} [\text{tr}(\Gamma^Y \nabla^2 f(\sqrt{1-t}x + \sqrt{t}Y))]\end{aligned}$$

Enfin, en utilisant le théorème de Fubini dans le sens inverse, on trouve :

$$\mathbb{E} \left[\nabla f(Z_t)^T \frac{Y}{2\sqrt{t}} \right] = \frac{1}{2} \mathbb{E} [\text{tr}(\nabla^2 f(Z_t) \Gamma^Y)]$$

En procédant de même pour X , on obtient :

$$\mathbb{E} \left[\nabla f(Z_t)^T \frac{X}{2\sqrt{1-t}} \right] = \frac{1}{2} \mathbb{E} [\text{tr}(\nabla^2 f(Z_t) \Gamma^X)]$$

ce qui implique :

$$h'(t) = \frac{1}{2} \mathbb{E} [\text{tr}(\nabla^2 f(Z_t)(\Gamma^Y - \Gamma^X))]$$

Soit $x \in \mathbb{R}^n$, notons $p(x) = \nabla f(x)$. Un calcul direct montre que la matrice hessienne de f s'écrit :

$$\nabla^2 f(x) = \lambda (\text{Diag}(p(x)) - p(x)p(x)^T)$$

Comme $\sum_{i=1}^n p_i(x) = 1$, on a :

$$\begin{aligned}\text{tr}(\text{Diag}(p(x))(\Gamma^Y - \Gamma^X)) &= \sum_{i=1}^n p_i(x)(\Gamma_{i,i}^Y - \Gamma_{i,i}^X) \\ &= \frac{1}{2} \sum_{1 \leq i, j \leq n} p_i(x) p_j(x) (\Gamma_{i,i}^Y - \Gamma_{i,i}^X + \Gamma_{j,j}^Y - \Gamma_{j,j}^X)\end{aligned}$$

De plus,

$$\text{tr}(p(x)p(x)^T(\Gamma^Y - \Gamma^X)) = \sum_{1 \leq i, j \leq n} p_i(x) p_j(x) (\Gamma_{i,j}^Y - \Gamma_{i,j}^X)$$

On en déduit que :

$$\begin{aligned} \text{tr} (\nabla^2 f(x)(\Gamma^Y - \Gamma^X)) &= \frac{\lambda}{2} \sum_{1 \leq i, j \leq n} p_i(x) p_j(x) (\Gamma_{i,i}^Y - \Gamma_{i,i}^X + \Gamma_{j,j}^Y - \Gamma_{j,j}^X - 2\Gamma_{i,j}^Y + 2\Gamma_{i,j}^X) \\ &= \frac{\lambda}{2} \sum_{1 \leq i, j \leq n} p_i(x) p_j(x) \underbrace{(\delta_{i,j}^Y - \delta_{i,j}^X)}_{\geq 0} \geq 0 \end{aligned}$$

Ainsi, $h'(t) \geq 0$. Ceci conclut la démonstration du premier point.

La démonstration du deuxième point suit exactement le même raisonnement, mais nécessite de remplacer les fonctions f_λ par les fonctions $f_{\lambda,\mu}$ définies par :

$$\forall \lambda, \mu \in \mathbb{R}_+^*, \forall x \in \mathbb{R}^n, f_{\lambda,\mu}(x) = \frac{1}{\lambda} \ln \left(\sum_{i=1}^k \exp \left(-\frac{\lambda}{\mu} \ln \left(\sum_{j \in T_i} \exp(-\mu x_j) \right) \right) \right)$$

Pour tout $\lambda, \mu \in \mathbb{R}_+^*$, ces fonctions vérifient l'encadrement :

$$\forall x \in \mathbb{R}^n, \max_{i \in [1,k]} \min_{j \in T_i} x_j - \frac{1}{\mu} \ln n \leq f_{\lambda,\mu}(x) \leq \max_{i \in [1,k]} \min_{j \in T_i} x_j + \frac{1}{\lambda} \ln n$$

Ainsi, les $f_{\lambda,\mu}$ convergent uniformément vers la fonction "max min" lorsque $\lambda \rightarrow +\infty$, $\mu \rightarrow +\infty$.

On peut toujours supposer que X et Y sont indépendants, quitte à les remplacer par les variables \tilde{X} et \tilde{Y} définies au début de la preuve.

On garde également les notations précédentes, en remplaçant f_λ par $f_{\lambda,\mu}$ dans la définition de h . Alors h est toujours dérivable sur $]0, 1[$ avec :

$$\forall t \in]0, 1[, h'(t) = \frac{1}{2} \mathbb{E} [\text{tr} (\nabla^2 f_{\lambda,\mu}(Z_t)(\Gamma^Y - \Gamma^X))]]$$

En fixant $t \in]0, 1[$, on obtient après quelques calculs :

$$h'(t) = \frac{1}{4} \sum_{i,j} \mathbb{E} \left[\frac{\partial^2 f_{\lambda,\mu}}{\partial x_i \partial x_j} (Z_t) \right] (\delta_{i,j}^X - \delta_{i,j}^Y)$$

On remarque par ailleurs que pour tout $i, j \in [1, n]$, $i \neq j$,

- Si i et j appartiennent à un même T_l , alors :

$$\forall x \in \mathbb{R}^n, \frac{\partial^2 f_{\lambda,\mu}}{\partial x_i \partial x_j} (x) \geq 0$$

- Si i et j appartiennent à deux T_l différents,

$$\forall x \in \mathbb{R}^n, \frac{\partial^2 f_{\lambda,\mu}}{\partial x_i \partial x_j} (x) \leq 0$$

Ainsi, d'après les hypothèses faites sur les $\delta_{i,j}^X$ et les $\delta_{i,j}^Y$, on a bien $h'(t) \geq 0$. On conclut alors la preuve comme dans le premier point. \square

Remarque. Si $X : (\Omega_1, \mathcal{F}_1, \mathbb{P}_1) \rightarrow \mathbb{R}^N$ et $Y : (\Omega_2, \mathcal{F}_2, \mathbb{P}_2) \rightarrow \mathbb{R}^N$ sont deux vecteurs aléatoires gaussiens vérifiant les hypothèses du théorème de Slepian-Gordon, alors la conclusion reste valable, même si X et Y ne sont pas définis sur le même espace.

Pour le voir, il suffit de considérer l'espace de probabilité produit $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mathbb{P}_1 \otimes \mathbb{P}_2)$ ainsi que les projections $p_1 : \Omega_1 \times \Omega_2 \rightarrow \Omega_1$ et $p_2 : \Omega_1 \times \Omega_2 \rightarrow \Omega_2$, de lois respectives \mathbb{P}_1 et \mathbb{P}_2 , puis d'appliquer le théorème aux variables aléatoires $X \circ p_1$ et $Y \circ p_2$. Il est immédiat que celles-ci vérifient aussi les conditions du théorème et que les inégalités obtenues se transportent à X et Y .

On déduit du théorème et de la remarque le corollaire suivant :

Corollaire. Soient $m, n \in \mathbb{N}^*$ et $X, Y \in \mathbb{R}^{mn}$ deux vecteurs aléatoires gaussiens de moyenne nulle, à mn variables indexées par $\llbracket 1, m \rrbracket \times \llbracket 1, n \rrbracket$. On suppose que pour tout $k, k' \in \llbracket 1, m \rrbracket \times \llbracket 1, n \rrbracket$,

$$\delta_{k, k'}^X \leq \delta_{k, k'}^Y$$

avec égalité si $k_2 = k'_2$. Alors :

1. $\mathbb{E} \left[\max_{(i, j) \in \llbracket 1, m \rrbracket \times \llbracket 1, n \rrbracket} X_{(i, j)} \right] \leq \mathbb{E} \left[\max_{(i, j) \in \llbracket 1, m \rrbracket \times \llbracket 1, n \rrbracket} Y_{(i, j)} \right]$
2. $\mathbb{E} \left[\max_{j \in \llbracket 1, n \rrbracket} \min_{i \in \llbracket 1, m \rrbracket} X_{(i, j)} \right] \leq \mathbb{E} \left[\max_{j \in \llbracket 1, n \rrbracket} \min_{i \in \llbracket 1, m \rrbracket} Y_{(i, j)} \right]$

C'est exactement ce résultat que nous utiliserons dans la suite.

3.2 Application à des processus gaussiens

On donne d'abord quelques définitions et propriétés générales.

Définition 5. Soit $(X_i : \Omega \rightarrow \mathbb{R})_{i \in I}$ une famille de variables aléatoires sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On dit que $(X_i)_{i \in I}$ est un processus gaussien si pour tout $J \subset I$ fini, la famille $(X_j)_{j \in J}$ est un vecteur aléatoire gaussien.

Soit $N \in \mathbb{N}$. Plaçons nous dans l'espace de probabilité (\mathbb{R}^N, γ_N) , où γ_N est la mesure de probabilité gaussienne standard avec pour densité contre la mesure de Lebesgue :

$$g_N : x \in \mathbb{R}^N \mapsto \frac{1}{\sqrt{2\pi}^N} \exp\left(-\frac{\|x\|_2^2}{2}\right)$$

Pour $u \in \mathbb{R}^N$, on considère la forme linéaire $Z_u := \langle \cdot, u \rangle$ comme une variable aléatoire réelle sur (\mathbb{R}^N, γ_N) .

On a alors les propriétés suivantes :

- Lemme 2.**
1. La famille $(Z_u)_{u \in \mathbb{R}^n}$ est un processus gaussien de moyenne nulle sur (\mathbb{R}^N, γ_N) ,
 2. L'application linéaire $u \in (\mathbb{R}^N, \|\cdot\|_2) \mapsto Z_u \in L^2(\mathbb{R}^N, \gamma_N)$ est une isométrie.

Démonstration. 1. Puisque l'application $u \in \mathbb{R}^N \mapsto Z_u \in (\mathbb{R}^N)^*$ est linéaire, cela revient à montrer que pour tout $u \in \mathbb{R}^N$, Z_u est une variable aléatoire gaussienne de moyenne nulle sur (\mathbb{R}^N, γ_N) .

Soit $u \in \mathbb{R}^N - \{0\}$, on fixe $O \in O_n(\mathbb{R})$ une matrice orthogonale telle que $Ou = (\|u\|_2, 0, \dots, 0)$. Alors, pour toute application mesurable $f : \mathbb{R} \rightarrow \mathbb{R}_+$, on a :

$$\begin{aligned} \mathbb{E}[f(Z_u)] &= \int_{\mathbb{R}^N} f(Z_u(x)) d\gamma_N(x) = \int_{\mathbb{R}^N} f(\langle x, u \rangle) g_N(x) dx \\ &= \int_{\mathbb{R}^N} f(\langle O^{-1}x, u \rangle) g_N(O^{-1}x) dx = \int_{\mathbb{R}^N} f(\langle x, Ou \rangle) g_N(x) dx \\ &= \int_{\mathbb{R}^N} f(\|u\|_2 x_1) g_N(x) dx = \int_{\mathbb{R}} f(\|u\|_2 x) g_1(x) dx \\ (y = \|u\|_2 x) &= \int_{\mathbb{R}} f(y) \frac{1}{\sqrt{2\pi}\|u\|_2} \exp\left(-\frac{y^2}{2\|u\|_2^2}\right) dy \end{aligned}$$

Ainsi, Z_u est un vecteur aléatoire gaussien de moyenne nulle et de variance $\|u\|_2$.

2. D'après ce qui précède, pour $f : x \mapsto x^2$, on a :

$$\begin{aligned}\mathbb{E}[Z_u^2] &= \int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi}\|u\|_2} \exp\left(-\frac{x^2}{2\|u\|_2^2}\right) dx \\ &= \frac{\|u\|_2^2}{\sqrt{2\pi}} \int_{\mathbb{R}} x^2 \exp\left(-\frac{x^2}{2}\right) dx = \frac{\|u\|_2^2}{\sqrt{2\pi}} \left(\left[-x \exp\left(-\frac{x^2}{2}\right) \right]_{-\infty}^{+\infty} - \int_{\mathbb{R}} -\exp\left(-\frac{x^2}{2}\right) dx \right) \\ &= \|u\|_2^2\end{aligned}$$

d'où le résultat. □

Soient $p, q \in \mathbb{N}^*$ et $S := S^{p-1} \times S^{q-1}$,

1. On définit sur $(\mathbb{R}^{pq}, \gamma_{pq})$ le processus gaussien :

$$X = (Z_{u \otimes v})_{(u,v) \in S}$$

avec $u \otimes v = (u_i v_j)_{i,j} \in \mathcal{M}_{p,q}(\mathbb{R}) \cong \mathbb{R}^{pq}$

2. On définit sur $(\mathbb{R}^{p+q}, \gamma_{p+q})$ le processus gaussien :

$$Y = (Z_{(u,v)})_{(u,v) \in S}$$

Nous allons appliquer les inégalités de Slepian-Gordon à X et Y . On vérifie d'abord les conditions d'application.

Lemme 3. 1. X et Y sont des processus gaussiens de moyenne nulle.

2. pour tout $k, k' \in S$, $\delta_{k,k'}^X \leq \delta_{k,k'}^Y$ avec égalité si $k_2 = k'_2$.

Démonstration. 1. C'est une conséquence immédiate du lemme précédent.

2. Soit $k = (u, v)$, $k' = (u', v') \in S$. D'après le lemme précédent, $\delta_{k,k'}^X = \|u \otimes v - u' \otimes v'\|_2$ et $\delta_{k,k'}^Y = \|(u, v) - (u', v')\|_2$. Il s'agit donc de montrer que $\|u \otimes v - u' \otimes v'\|_2 \leq \|(u, v) - (u', v')\|_2$ avec égalité si $v = v'$.

Par un calcul direct, on a :

$$\begin{aligned}\|(u, v) - (u', v')\|_2^2 &= \sum_{i=1}^p (u_i - u'_i)^2 + \sum_{j=1}^q (v_j - v'_j)^2 \\ \|u \otimes v - u' \otimes v'\|_2^2 &= \sum_{i=1}^p \sum_{j=1}^q (u_i v_j - u'_i v'_j)^2 \\ &= \sum_{i,j} ((u_i - u'_i)v_j + (v_j - v'_j)u'_i)^2 \\ &= \sum_{i,j} (u_i - u'_i)^2 v_j^2 + (v_j - v'_j)^2 u'_i{}^2 + 2(u_i - u'_i)(v_j - v'_j)u'_i v_j \\ &= \sum_i (u_i - u'_i)^2 + \sum_j (v_j - v'_j)^2 + 2 \sum_{i,j} (u_i - u'_i)(v_j - v'_j)u'_i v_j \\ &= \|(u, v) - (u', v')\|_2^2 + 2 \sum_{i,j} (u_i - u'_i)(v_j - v'_j)u'_i v_j\end{aligned}$$

Or :

$$\sum_{i,j} (u_i - u'_i) (v'_j - v_j) u'_i v_j = \left(\sum_i u_i u'_i - u_i'^2 \right) \left(\sum_j v_j v'_j - v_j^2 \right) = (u^T u' - 1) (v^T v' - 1)$$

Ainsi :

$$\|(u, v) - (u', v')\|_2^2 - \|u \otimes v - u' \otimes v'\|_2^2 = (u^T u' - 1) (v^T v' - 1) \geq 0$$

avec égalité si $v = v'$, puisque $v^T v = 1$.

□

On peut ainsi appliquer les inégalités de Slepian-Gordon et obtenir directement :

Corollaire. *Pour toutes parties finies $I \subset S^{p-1}, J \subset S^{q-1}$, on a :*

1. $\mathbb{E} \left[\max_{k \in I \times J} X_k \right] \leq \mathbb{E} \left[\max_{k \in I \times J} X_k \right]$
2. $\mathbb{E} \left[\max_{v \in J} \min_{u \in I} X_{u,v} \right] \leq \mathbb{E} \left[\max_{v \in J} \min_{u \in I} Y_{u,v} \right]$

On montre alors que ces relations sont toujours vraies si $I = S^{p-1}$ et $J = S^{q-1}$. L'argument principal repose sur la compacité de S .

Proposition 3.

$$\mathbb{E} \left[\max_{\substack{u \in S^{p-1} \\ v \in S^{q-1}}} X_{u,v} \right] \leq \mathbb{E} \left[\max_{\substack{u \in S^{p-1} \\ v \in S^{q-1}}} Y_{u,v} \right] \quad (I_1)$$

$$\mathbb{E} \left[\max_{v \in S^{q-1}} \min_{u \in S^{p-1}} X_{u,v} \right] \leq \mathbb{E} \left[\max_{v \in S^{q-1}} \min_{u \in S^{p-1}} Y_{u,v} \right] \quad (I_2)$$

Démonstration. On se contente de démontrer (I_1) , (I_2) s'obtenant de façon très similaire.

Il suffit de remarquer que pour tout compact K de $(\mathbb{R}^N, \|\cdot\|_2)$,

$$\mathbb{E} \left[\max_{u \in K} Z_u \right] = \sup_{\substack{S \subset K \\ S \text{ fini}}} \mathbb{E} \left[\max_{s \in S} Z_s \right]$$

et appliquer le dernier corollaire.

On a directement pour tout $S \subset K$, S fini,

$$\max_{u \in K} Z_u \geq \max_{s \in S} Z_s$$

d'où

$$\mathbb{E} \left[\max_{u \in K} Z_u \right] \geq \sup_{\substack{S \subset K \\ S \text{ fini}}} \mathbb{E} \left[\max_{s \in S} Z_s \right]$$

Pour démontrer l'autre inégalité, fixons $\epsilon > 0$.

Par compacité de K , il existe une partie finie $S = \{s_1, \dots, s_r\} \subset K$ telle que : $K \subset \bigcup_{k=1}^r B(t_k, \epsilon)$.

On utilise alors l'inégalité élémentaire :

$$\forall u, u' \in K, \forall x \in \mathbb{R}^N, |Z_u(x) - Z_{u'}(x)| = |\langle x, u - u' \rangle| \leq \|x\|_2 \|u - u'\|_2$$

de façon à avoir : pour tout $u \in K$, en fixant $k \in \llbracket 1, r \rrbracket$ tel que $\|u - s_k\|_2 \leq \epsilon$,

$$\forall x \in \mathbb{R}^N, Z_u(x) \leq Z_{s_k} + \|x\|_2 \epsilon$$

Ainsi :

$$\forall x \in \mathbb{R}^N, \left(\max_{u \in K} Z_u \right) (x) \leq \left(\max_{s \in S} Z_s \right) (x) + \|x\|_2 \epsilon$$

On obtient alors, en passant à l'espérance :

$$\mathbb{E} \left[\max_{u \in K} Z_u \right] \leq \mathbb{E} \left[\max_{s \in S} Z_s \right] + \underbrace{\mathbb{E} [\|\cdot\|_2]}_{< \infty} \epsilon$$

Alors en notant $M := \mathbb{E} [\|\cdot\|_2] < \infty$, on a pour tout $\epsilon > 0$:

$$\mathbb{E} \left[\max_{u \in K} Z_u \right] \leq \sup_{\substack{S \subset K \\ S \text{ fini}}} \mathbb{E} \left[\max_{s \in S} Z_s \right] + M \epsilon$$

En faisant tendre ϵ vers 0, on obtient l'autre inégalité :

$$\mathbb{E} \left[\max_{u \in K} Z_u \right] \leq \sup_{\substack{S \subset K \\ S \text{ fini}}} \mathbb{E} \left[\max_{s \in S} Z_s \right]$$

ce qui termine la preuve. □

3.3 Démonstration du résultat

Avant de passer à la preuve des inégalités portant sur l'espérance des valeurs singulières, on détaille quelques calculs que l'on peut omettre lors d'une première lecture.

Lemme 4. 1. On a la formule :

$$\int_{\mathbb{R}^N} \|x\|_2 g_N(x) dx = \sqrt{2} \frac{\Gamma\left(\frac{N+1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)}$$

2. On a l'inégalité :

$$\sqrt{2} \frac{\Gamma\left(\frac{N+1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)} < \sqrt{N}$$

3. Si $p, q \in \mathbb{N}^*$, $p \geq q$, on a l'inégalité :

$$\sqrt{2} \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} - \sqrt{2} \frac{\Gamma\left(\frac{q+1}{2}\right)}{\Gamma\left(\frac{q}{2}\right)} > \sqrt{p} - \sqrt{q}$$

Démonstration. 1. Notons :

$$I := \int_{\mathbb{R}^N} \|x\|_2 g_N(x) dx = \int_{\mathbb{R}^N} \|x\|_2 \frac{1}{\sqrt{2\pi}^N} \exp\left(-\frac{\|x\|_2^2}{2}\right) dx$$

Un changement de variable polaire donne :

$$I = \int_{\mathbb{R}_+} r \frac{1}{\sqrt{2\pi}^N} \exp\left(-\frac{r^2}{2}\right) S_N r^{n-1} dr$$

où S_N est la surface de la sphère de rayon 1 en dimension N .

On a alors :

$$\begin{aligned} I &= \frac{S_N}{\sqrt{2\pi}^N} \int_{\mathbb{R}_+} r^N \exp\left(-\frac{r^2}{2}\right) dr \stackrel{(r=\sqrt{2t})}{=} \frac{S_N}{\sqrt{2\pi}^N} \int_{\mathbb{R}_+} \sqrt{2t}^N \exp(-t) \frac{dt}{\sqrt{2t}} \\ &= \frac{S_N \sqrt{2}^{N-1}}{\sqrt{2\pi}^N} \int_{\mathbb{R}_+} t^{\frac{N-1}{2}} \exp(-t) dt = \frac{S_N}{\sqrt{2}\sqrt{\pi}^N} \Gamma\left(\frac{N+1}{2}\right) \end{aligned}$$

Un calcul classique donne :

$$S_N = \frac{2\sqrt{\pi}^N}{\Gamma\left(\frac{N}{2}\right)}$$

d'où le résultat :

$$I = \sqrt{2} \frac{\Gamma\left(\frac{N+1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)}$$

2. On rappelle que la fonction Γ est log-convexe sur \mathbb{R}_+^* , c'est-à-dire :

$$\forall x, y \in \mathbb{R}_+^*, \forall \alpha \in]0, 1[, \Gamma(\alpha x + (1-\alpha)y) < \Gamma(x)^\alpha \Gamma(y)^{1-\alpha}$$

En particulier,

$$\Gamma\left(\frac{1}{2} \frac{N}{2} + \frac{1}{2} \left(\frac{N}{2} + 1\right)\right) < \Gamma\left(\frac{N}{2}\right)^{\frac{1}{2}} \Gamma\left(\frac{N}{2} + 1\right)^{\frac{1}{2}} = \Gamma\left(\frac{N}{2}\right)^{\frac{1}{2}} \left(\frac{N}{2} \Gamma\left(\frac{N}{2}\right)\right)^{\frac{1}{2}}$$

Ce qui donne :

$$\Gamma\left(\frac{N+1}{2}\right) < \sqrt{\frac{N}{2}} \Gamma\left(\frac{N}{2}\right)$$

C'est-à-dire :

$$\sqrt{2} \frac{\Gamma\left(\frac{N+1}{2}\right)}{\Gamma\left(\frac{N}{2}\right)} < \sqrt{N}$$

3. Ce point est nettement plus technique à démontrer. L'idée générale de la preuve est de vérifier que l'inégalité est vraie pour de très grandes valeurs de p et q , en utilisant le développement asymptotique de la fonction Γ , puis de la démontrer pour tout p et q par récurrence décroissante. \square

Remarque. Dans le théorème qui suit, on fera attention au fait que l'on considère deux mesures de probabilité différentes sur \mathbb{R}^N . La première est la mesure gaussienne standard γ_N . La deuxième est la mesure image d'une matrice aléatoire gaussienne sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

En considérant $A : \Omega \rightarrow \mathcal{M}_{p,q}(\mathbb{R})$ une matrice aléatoire gaussienne, sa mesure image μ_A sur $\mathcal{M}_{p,q}(\mathbb{R}) \cong \mathbb{R}^{pq}$ est la mesure de densité contre la mesure de Lebesgue :

$$x \in \mathbb{R}^{pq} \mapsto \frac{1}{\sqrt{2\pi}^{\frac{1}{p}N}} \exp\left(-\frac{\|x\|_2^2}{2\frac{1}{p}}\right)$$

Il s'agit d'une mesure gaussienne centrée de variance différente de 1. On trouvera en Annexe plus de détails sur le passage de ce type de mesure à une mesure gaussienne standard.

Théorème 7. Soit $p, q \in \mathbb{N}^*$, $p \geq q$ et $A : \Omega \rightarrow \mathcal{M}_{p,q}(\mathbb{R})$ une matrice aléatoire gaussienne, alors :

$$1 - \sqrt{\frac{q}{p}} < \mathbb{E}[\lambda_{\min}(A)] \leq \mathbb{E}[\lambda_{\max}(A)] < 1 + \sqrt{\frac{q}{p}}$$

Démonstration. Pour tout $M \in \mathcal{M}_{p,q}(\mathbb{R})$, $x \in \mathbb{R}^p$ et $y \in \mathbb{R}^q$, on a les égalités :

$$\max_{\substack{u \in S^{p-1} \\ v \in S^{q-1}}} Z_{u \otimes v}(M) = \max_{\substack{u \in S^{p-1} \\ v \in S^{q-1}}} \langle Mv, u \rangle = \max_{v \in S^{q-1}} \|Mv\|_2 = \lambda_{\max}(M) \quad (a)$$

$$\max_{\substack{u \in S^{p-1} \\ v \in S^{q-1}}} Z_{(u,v)}((x, y)) = \max_{u \in S^{p-1}} \langle x, u \rangle + \max_{v \in S^{q-1}} \langle y, v \rangle = \|x\|_2 + \|y\|_2 \quad (b)$$

$$\max_{v \in S^{q-1}} \min_{u \in S^{p-1}} Z_{u \otimes v}(M) = \max_{v \in S^{q-1}} \min_{u \in S^{p-1}} \langle Mv, u \rangle = -\lambda_{\min}(M) \quad (c)$$

$$\max_{v \in S^{q-1}} \min_{u \in S^{p-1}} Z_{(u,v)}((x, y)) = \max_{v \in S^{q-1}} \langle y, v \rangle + \min_{u \in S^{p-1}} \langle x, u \rangle = \|y\|_2 - \|x\|_2 \quad (d)$$

Or on a montré qu'on pouvait appliquer les inégalités de Slepian-Gordon pour obtenir :

$$\mathbb{E}_{\gamma_{pq}} \left[\max_{\substack{u \in S^{p-1} \\ v \in S^{q-1}}} Z_{u \otimes v} \right] \leq \mathbb{E}_{\gamma_{pq}} \left[\max_{\substack{u \in S^{p-1} \\ v \in S^{q-1}}} Z_{(u,v)} \right] \quad (I_1)$$

$$\mathbb{E}_{\gamma_{pq}} \left[\max_{v \in S^{q-1}} \min_{u \in S^{p-1}} Z_{u \otimes v} \right] \leq \mathbb{E}_{\gamma_{pq}} \left[\max_{v \in S^{q-1}} \min_{u \in S^{p-1}} Z_{(u,v)} \right] \quad (I_2)$$

- Les égalités (a) et (b), l'inégalité (I₁) et les calculs effectués au Lemme 3 permettent d'obtenir :

$$\begin{aligned} \mathbb{E}_{\gamma_{pq}} [\lambda_{\max}] &\leq \mathbb{E}_{\gamma_{pq}} [\|x\|_2 + \|y\|_2] = \int_{\mathbb{R}^p \times \mathbb{R}^q} (\|x\|_2 + \|y\|_2) g_{p+q}((x, y)) dx dy \\ &= \int_{\mathbb{R}^p} \|x\|_2 g_p(x) dx + \int_{\mathbb{R}^q} \|y\|_2 g_q(y) dy \\ &= \sqrt{2} \frac{\Gamma(\frac{q+1}{2})}{\Gamma(\frac{q}{2})} + \sqrt{2} \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} < \sqrt{q} + \sqrt{p} \end{aligned}$$

- De même, les égalités (b) et (c), l'inégalité (I₂) et le Lemme 2 permettent d'obtenir :

$$\begin{aligned} -\mathbb{E}_{\gamma_{pq}} [\lambda_{\min}] &\leq \mathbb{E}_{\gamma_{pq}} [\|x\|_2 - \|y\|_2] = \int_{\mathbb{R}^p \times \mathbb{R}^q} (-\|x\|_2 + \|y\|_2) g_{p+q}((x, y)) dx dy \\ &= -\int_{\mathbb{R}^p} \|x\|_2 g_p(x) dx + \int_{\mathbb{R}^q} \|y\|_2 g_q(y) dy \\ &= -\sqrt{2} \frac{\Gamma(\frac{p+1}{2})}{\Gamma(\frac{p}{2})} + \sqrt{2} \frac{\Gamma(\frac{q+1}{2})}{\Gamma(\frac{q}{2})} < -\sqrt{p} + \sqrt{q} \end{aligned}$$

- Des deux points précédents, on déduit :

$$\sqrt{p} - \sqrt{q} < \mathbb{E}_{\gamma_{pq}} [\lambda_{\min}] \leq \mathbb{E}_{\gamma_{pq}} [\lambda_{\max}] < \sqrt{p} + \sqrt{q}$$

En passant à la mesure μ_A , on obtient (*cf.* Annexe) :

$$\sqrt{p} - \sqrt{q} < \sqrt{p} \mathbb{E}_{\mathbb{P}} [\lambda_{\min}(A)] \leq \sqrt{p} \mathbb{E}_{\mathbb{P}} [\lambda_{\max}(A)] < \sqrt{p} + \sqrt{q}$$

ce qui donne le resultat en divisant par \sqrt{p} . □

4 Inégalités de concentration pour des fonctions lipschitziennes

Dans cette section, on s'intéresse à la concentration de fonctions lipschitziennes sur \mathbb{R}^n pour la mesure gaussienne. Les inégalités de concentration autour de la moyenne nous servent à la démonstration des inégalités de concentration utilisées dans la section 2. On présente également des inégalités de concentration autour de la médiane, qui n'interviennent pas dans la démonstration des résultats des parties précédentes.

4.1 Concentration autour de la moyenne

Le théorème que nous avons déjà mentionné dans la sous-section 2.3. et que nous allons maintenant démontrer est le suivant :

Théorème 8. *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application 1-lipschitzienne, pour la norme euclidienne, en notant $\mathbb{E}[f] = \int f d\gamma_n$ sa moyenne pour la mesure gaussienne standard γ_n , on a pour tout réel $r \geq 0$:*

$$\begin{aligned}\gamma_n(f \geq \mathbb{E}[f] + r) &\leq \exp\left(-\frac{r^2}{2}\right) \\ \gamma_n(f \leq \mathbb{E}[f] - r) &\leq \exp\left(-\frac{r^2}{2}\right)\end{aligned}$$

La démonstration de ces inégalités de concentration découle directement de la proposition suivante :

Proposition 4. *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application 1-lipschitzienne de moyenne nulle pour γ_n , on a pour tout réel λ :*

$$\int_{\mathbb{R}^n} \exp(\lambda f) d\gamma_n \leq \exp\left(\frac{\lambda^2}{2}\right)$$

Nous commençons d'abord par démontrer le théorème à partir de cette proposition.

Démonstration. D'après la proposition, pour tout réel λ , on a :

$$\int_{\mathbb{R}^n} \exp(\lambda(f - \mathbb{E}[f])) d\gamma_n \leq \exp\left(\frac{\lambda^2}{2}\right)$$

Or, en fixant $r \geq 0$ et en notant $A := \{f \geq \mathbb{E}[f] + r\}$, on a :

$$\int_{\mathbb{R}^n} \exp(\lambda(f - \mathbb{E}[f])) d\gamma_n \geq \int_A \exp(\lambda(f - \mathbb{E}[f])) d\gamma_n \geq \int_A \exp(\lambda r) d\gamma_n = \gamma_n(A) \exp(\lambda r)$$

Ainsi :

$$\forall \lambda \in \mathbb{R}, \gamma_n(f \geq \mathbb{E}[f] + r) \leq \exp\left(-\lambda r + \frac{\lambda^2}{2}\right)$$

Le membre de gauche de l'inégalité ci-dessus atteint son minimum pour $\lambda = r$ et vaut alors $\exp\left(-\frac{r^2}{2}\right)$, ce qui permet d'obtenir la première inégalité du théorème. La deuxième s'obtient alors en appliquant la première inégalité à $-f$. \square

La démonstration de la proposition est un peu technique et nécessite d'introduire des outils très importants dans la théorie des processus stochastiques. Nous ne rentrerons pas dans les détails de certains calculs classiques.

Pour toute application $f : \mathbb{R}^n \rightarrow \mathbb{R}$ mesurable bornée et pour tout réel $t \geq 0$, on définit l'application $P_t f : \mathbb{R}^n \rightarrow \mathbb{R}$ par :

$$\forall x \in \mathbb{R}^n, P_t f(x) = \int_{\mathbb{R}^n} f\left(e^{-t}x + \sqrt{1 - e^{-2t}}y\right) d\gamma_n(y)$$

On notera que $P_0 f = f$, et on pourrait montrer que $P_t \circ P_{t'}(f) = P_{t+t'}(f)$ pour $t, t' \geq 0$. Par ailleurs, par convergence dominée, $\lim_{t \rightarrow +\infty} P_t f = \int_{\mathbb{R}^n} f d\gamma_n$. On peut donc interpréter $P_t f$ comme une déformation de f au cours du temps, convergeant simplement vers sa moyenne.

Si f est de classe \mathcal{C}^1 bornée et de dérivées bornées, une application directe du théorème de dérivation sous l'intégrale permet d'obtenir :

$$\nabla P_t f = e^{-t} P_t (\nabla f)$$

où $P_t (\nabla f)$ désigne $\left(P_t \left(\frac{\partial f}{\partial x_1} \right), \dots, P_t \left(\frac{\partial f}{\partial x_n} \right) \right)$.

Si f est de classe \mathcal{C}^2 bornée et de dérivées premières et secondes bornées, on définit également l'application $L f : \mathbb{R}^n \rightarrow \mathbb{R}$ par :

$$\forall x \in \mathbb{R}^n, L f(x) = \Delta f(x) - \langle \nabla f(x), x \rangle$$

Par un calcul direct, on obtient pour tout $t \geq 0$ et tout $x \in \mathbb{R}^n$:

$$\frac{d}{dt} (P_t f(x)) = L (P_t f)(x)$$

Pour $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 bornées et de dérivées bornées, la formule d'intégration par parties permet d'obtenir :

$$\int_{\mathbb{R}^n} f L g d\gamma_n = - \int_{\mathbb{R}^n} \langle \nabla f, \nabla g \rangle d\gamma_n$$

Nous pouvons désormais établir la preuve de la proposition.

Démonstration. Commençons par considérer $f : \mathbb{R}^n \rightarrow \mathbb{R}$ 1-lipschitzienne, de classe \mathcal{C}^2 , bornée et de dérivées bornées, et de moyenne nulle. Fixons $\lambda \in \mathbb{R}$ et posons l'application $G : \mathbb{R}_+ \rightarrow \mathbb{R}$ définie par :

$$\forall t \geq 0, G(t) = \int_{\mathbb{R}^n} \exp(\lambda P_t f) d\gamma_n$$

Puisque f est 1-lipschitzienne, pour tout $x \in \mathbb{R}^n$, on a $\|\nabla f(x)\|_2 \leq 1$, donc :

$$\forall t \geq 0, \forall x \in \mathbb{R}^n, \|\nabla P_t f(x)\|_2 \leq e^{-t} \int_{\mathbb{R}^n} \left\| \nabla f \left(e^{-t} x + \sqrt{1 - e^{-2t}} y \right) \right\|_2 d\gamma_n(y) \leq e^{-t}$$

De plus, comme f est de moyenne nulle, $P_t f$ converge simplement vers 0 lorsque $t \rightarrow +\infty$, par application du théorème de convergence dominée, donc par le même argument, $\lim_{t \rightarrow +\infty} G(t) = 1$. On constate également que G est dérivable, ce qui permet d'écrire, pour tout $t \geq 0$:

$$G(t) = 1 - \int_t^{+\infty} G'(s) ds = 1 - \lambda \int_t^{+\infty} \left(\int_{\mathbb{R}^n} L(P_s f) \exp(\lambda P_s f) \right) ds$$

En appliquant la propriété d'intégration par parties énoncée précédemment, on obtient :

$$G(t) = 1 + \lambda^2 \int_t^{+\infty} \left(\int_{\mathbb{R}^n} \|\nabla(P_s f)\|_2^2 \exp(\lambda P_s f) \right) ds \leq 1 + \lambda^2 \int_t^{+\infty} e^{-2s} G(s) ds$$

En posant $H : t \in \mathbb{R}_+ \mapsto \ln \left(1 + \lambda^2 \int_t^{+\infty} e^{-2s} G(s) ds \right)$, l'inégalité précédente donne :

$$-H'(t) = \frac{\lambda^2 e^{-2t} G(t)}{1 + \lambda^2 \int_t^{+\infty} e^{-2s} G(s) ds} \leq \frac{\lambda^2 e^{-2t} G(t)}{G(t)} = \lambda^2 e^{-2t}$$

Ainsi :

$$\ln \left(\int_{\mathbb{R}^n} \exp(\lambda f) d\gamma_n \right) = \ln(G(0)) \leq H(0) = \int_0^{+\infty} -H'(t) dt \leq \int_0^{+\infty} \lambda^2 e^{-2t} dt = \frac{\lambda^2}{2}$$

La proposition est donc prouvée pour des fonctions suffisamment régulières.

Dans le cas général, si $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est seulement supposée 1-lipschitzienne et de moyenne nulle, des résultats techniques de théorie de la mesure permettent de construire une suite $(f_k)_{k \in \mathbb{N}}$ de fonctions 1-lipschitziennes, bornées, de classe \mathcal{C}^2 , de dérivées bornées et de moyenne nulle qui converge presque partout vers f . Alors pour tout $\lambda \in \mathbb{R}$ et pour tout $k \in \mathbb{N}$, on a :

$$\int_{\mathbb{R}^n} \exp(\lambda f_k) d\gamma_n \leq \exp\left(\frac{\lambda^2}{2}\right)$$

Le lemme de Fatou permet alors d'obtenir :

$$\int_{\mathbb{R}^n} \exp(\lambda f) d\gamma_n = \int_{\mathbb{R}^n} \liminf_{k \rightarrow +\infty} \exp(\lambda f_k) d\gamma_n \leq \liminf_{k \rightarrow +\infty} \int_{\mathbb{R}^n} \exp(\lambda f_k) d\gamma_n \leq \exp\left(\frac{\lambda^2}{2}\right)$$

La proposition est donc démontrée dans le cas général. □

4.2 Concentration autour de la médiane et isopérimétrie gaussienne

Commençons par quelques résultats généraux sur la concentration de fonctions lipschitziennes autour de la médiane.

Définition 6. Soit μ une mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ et $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application borélienne. On appelle médiane de f pour μ tout réel m vérifiant :

$$\mu(f \leq m) \geq \frac{1}{2} \quad \text{et} \quad \mu(f \geq m) \geq \frac{1}{2}$$

Remarque. Avec cette définition et d'après les propriétés élémentaires des fonctions de répartition de variables aléatoires, on vérifie que f admet au moins une médiane. Cependant, elle peut en admettre plusieurs. Dans le cas où μ est strictement positive sur les ouverts non vides et f est continue, on peut facilement montrer que f admet une unique médiane pour μ , que l'on note $\text{Med}(f)$.

Définition 7. Si μ est une mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, la fonction de concentration de μ est l'application $\alpha_\mu : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ définie par :

$$\forall r \in \mathbb{R}_+, \quad \alpha_\mu(r) = \sup \left\{ \mu(\mathbb{R}^n \setminus A_r) : A \in \mathcal{B}(\mathbb{R}^n), \mu(A) \geq \frac{1}{2} \right\}$$

où $A_r := \{x \in \mathbb{R}^n : d(x, A) < r\}$ pour $A \subset \mathbb{R}^n$, d étant la distance euclidienne sur \mathbb{R}^n .

Il découle de cette définition :

Proposition 5 (Inégalités de Lévy). Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application 1-lipschitzienne pour la norme euclidienne et admettant une médiane $\text{Med}(f) \in \mathbb{R}$ par rapport à μ . Alors, pour tout $r \in \mathbb{R}_+$:

$$\begin{aligned} \mu(f \geq \text{Med}(f) + r) &\leq \alpha_\mu(r) \\ \mu(f \leq \text{Med}(f) - r) &\leq \alpha_\mu(r) \end{aligned}$$

Démonstration. Soit $r \geq 0$, posons $A := \{f \leq \text{Med}(f)\} \in \mathcal{B}(\mathbb{R}^n)$. Par définition de $\text{Med}(f)$, $\mu(A) \geq \frac{1}{2}$ donc $\mu(\mathbb{R}^n \setminus A_r) \leq \alpha_\mu(r)$. Puisque f est 1-lipschitzienne, on a $\{f \geq \text{Med}(f) + r\} \subset \mathbb{R}^n \setminus A_r$, donc :

$$\mu(f \geq \text{Med}(f) + r) \leq \alpha_\mu(r)$$

La deuxième inégalité s'obtient de façon similaire, en considérant $B := \{f \geq \text{Med}(f)\} \in \mathcal{B}(\mathbb{R}^n)$. □

Afin de démontrer des inégalités de concentrations autour de la médiane, nous sommes amenés à déterminer l'expression de α_μ dans le cas où μ est la mesure gaussienne standard sur \mathbb{R}^n . Nous devons alors nous intéresser à l'élégant Théorème Isopérimétrique Gaussien, que nous admettons, qui décrit la géométrie des parties mesurables de \mathbb{R}^n qui minimisent leur "périmètre" à volume fixé. On trouvera une démonstration de ce théorème dans [2].

Définition 8. Si μ est une mesure de probabilité sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, la mesure de bord associée à μ , notée μ^+ , est définie par :

$$\forall A \in \mathcal{B}(\mathbb{R}^n), \mu^+(A) = \liminf_{r \rightarrow 0^+} \frac{\mu(A_r) - \mu(A)}{r}$$

Remarque. Si μ est une mesure borélienne finie sur \mathbb{R}^n et si $A \in \mathcal{B}(\mathbb{R}^n)$, l'application $r \mapsto \mu(A_r)$ est continue sur \mathbb{R}_+ . Cependant, cette application n'est pas nécessairement dérivable. Il suffit de prendre $\mu = \gamma_n$ et $A = \mathbb{Q}$ pour s'en rendre compte.

Les théorèmes isopérimétriques s'énoncent souvent comme une inégalité de la forme :

$$\forall A \in \mathcal{B}(\mathbb{R}^n), \mu^+(A) \geq \mathcal{I}_\mu(\mu(A))$$

où $\mathcal{I}_\mu : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ est "suffisamment régulière" et "optimise" en un certain sens l'inégalité ci-dessus.

Dans la suite, on considère la mesure gaussienne standard sur $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, notée γ_n . On note Φ la fonction de répartition de γ_1 sur \mathbb{R} , c'est-à-dire :

$$\forall r \in \mathbb{R}, \Phi(r) = \int_{-\infty}^r \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

Alors Φ est strictement croissante sur \mathbb{R} et de classe \mathcal{C}^∞ . On définit alors la *fonction isopérimétrique gaussienne* $\mathcal{I} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ par : $\mathcal{I} := \Phi' \circ \Phi^{-1}$.

Commençons par calculer la mesure de bord associée à γ_n sur des sous-ensembles particuliers, les demi-espaces :

Proposition 6. Soit $u \in S^{n-1}$ et $a \in \mathbb{R}$, on considère le demi-espace $H^{u,a} = \{x \in \mathbb{R}^n : \langle x, u \rangle < a\}$. Alors :

$$\begin{aligned} \forall r \geq 0, \gamma_n(H_r^{u,a}) &= \Phi(a+r) = \Phi(\Phi^{-1}(\gamma_n(H^{u,a})) + r) \\ \gamma_n^+(H^{u,a}) &= \Phi'(a) = \mathcal{I}(\gamma_n(H^{u,a})) \end{aligned}$$

Démonstration. Remarquons d'abord que l'on a : $H_r^{u,a} = H^{u,a+r}$. Comme γ_n est invariante par l'action de $O_n(\mathbb{R})$, il nous suffit de calculer $\gamma_n(H_r)$ où $H := \{x \in \mathbb{R}^n : x_1 < 0\}$ pour $r \in \mathbb{R}$. Or :

$$\begin{aligned} \gamma_n(H_r) &= \int_{-\infty}^r \int_{\mathbb{R}^{n-1}} \frac{1}{\sqrt{2\pi}^n} \exp\left(-\frac{x_1^2 + \dots + x_n^2}{2}\right) dx_1 dx_2 \dots dx_n \\ &= \int_{-\infty}^r \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \Phi(r) \end{aligned}$$

La deuxième égalité découle immédiatement de la première par dérivation. □

Les demi-espaces sont exactement les parties qui réalisent l'égalité dans l'Inégalité Isopérimétrique Gaussienne. Celle-ci s'énonce :

Théorème 9 (Théorème Isopérimétrique Gaussien). Pour tout $A \in \mathcal{B}(\mathbb{R}^n)$, on a l'inégalité :

$$\gamma_n^+(A) \geq \mathcal{I}(\gamma_n(A)) \tag{TIG}$$

On en déduit alors le Théorème de Concentration Gaussien (TCG), qui est en fait équivalent au TIG :

Théorème 10. *Soit $A \in \mathcal{B}(\mathbb{R}^n)$, alors pour tout $r \geq 0$, on a l'inégalité :*

$$\gamma_n(A_r) \geq \Phi(\Phi^{-1}(\gamma_n(A)) + r) \quad (\text{TCG})$$

Nous nous contentons de démontrer l'équivalence entre le TIG et le TCG.

Démonstration. • Commençons par démontrer le sens "TIG \Rightarrow TCG".

Pour cela, on pose pour tout $r \geq 0$:

$$R_r : [0, 1] \longrightarrow \mathbb{R} \\ p \longmapsto \Phi(\Phi^{-1}(p) + r)$$

avec la convention $R_r(0) = 1 - R_r(1) = 0$. Les R_r sont continues, strictement croissantes et vérifient la relation : $R_r \circ R_{r'} = R_{r+r'}$ pour tout $r, r' \in \mathbb{R}_+$. De plus, pour tout $p \in [0, 1]$, $r \longmapsto R_r(p)$ est dérivable et sa dérivée en 0 vaut $\mathcal{I}(p)$.

L'inégalité (TCG) se réécrit alors :

$$\forall A \in \mathcal{B}(\mathbb{R}^n), \forall r \in \mathbb{R}_+, \gamma_n(A_r) \geq R_r(\gamma_n(A))$$

Soit $A \in \mathcal{B}(\mathbb{R}^n)$ tel que $\gamma_n(A) \in]0, 1[$ (le cas où $\gamma_n(A) \in \{0, 1\}$ est trivial). Fixons $\sigma > 1$ et notons $R_{r,\sigma} := R_{\frac{r}{\sigma}}$. On pose $\Delta_A := \{r \in \mathbb{R}_+ : \gamma_n(A_r) \geq R_{r,\sigma}(\gamma_n(A))\} \subset \mathbb{R}_+$, et on montre que Δ_A est \mathbb{R}_+ tout entier. Pour cela, il suffit de prouver que Δ_A contient 0, est ouvert à droite et fermé dans \mathbb{R}_+ .

◇ Δ_A contient 0 car $\gamma_n(A) = R_{0,\sigma}(\gamma_n(A))$.

◇ Soit $r \in \Delta_A$. La définition de la mesure de bord et le développement de Taylor en 0^+ à l'ordre 1 de $r' \longmapsto R_{r',\sigma}(\gamma_n(A_r))$ donnent au voisinage de 0^+ :

$$\gamma_n(A_{r+r'}) = \gamma_n((A_r)_{r'}) \geq \gamma_n(A_r) + r' \gamma_n^+(A_r) + o(r') \\ R_{r',\sigma}(\gamma_n(A_r)) = \gamma_n(A_r) + \frac{r'}{\sigma} \mathcal{I}(\gamma_n(A_r)) + o(r')$$

D'après l'inégalité (TIG), on a : $\gamma_n^+(A_r) \geq \mathcal{I}(\gamma_n(A_r)) > \frac{1}{\sigma} \mathcal{I}(\gamma_n(A_r))$. Alors, pour r' assez petit :

$$\gamma_n(A_{r+r'}) \geq R_{r',\sigma}(\gamma_n(A_r))$$

Or, comme $r \in \Delta_A$, $\gamma_n(A_r) \geq R_{r,\sigma}(\gamma_n(A))$. Ainsi, pour r' assez petit,

$$\gamma_n(A_{r+r'}) \geq R_{r',\sigma} \circ R_{r,\sigma}(\gamma_n(A)) = R_{r+r',\sigma}(\gamma_n(A))$$

Ainsi, Δ_A est ouvert à droite dans \mathbb{R}_+ .

◇ Enfin, Δ_A est fermé dans \mathbb{R}_+ puisque les applications $r \longmapsto \gamma_n(A_r)$ et $r \longmapsto R_{r,\sigma}(\gamma_n(A))$ sont continues.

Par un argument de connexité, on a donc $\Delta_A = \mathbb{R}_+$. Ainsi, pour tout $\sigma > 1$ et pour tout $r \geq 0$, on a :

$$\gamma_n(A_r) \geq \Phi\left(\Phi^{-1}(\gamma_n(A)) + \frac{r}{\sigma}\right)$$

On en déduit alors l'inégalité (TCG) pour A en faisant tendre σ vers 1.

- Le sens "TCG \Rightarrow TIG" est beaucoup plus facile. En effet, si $A \in \mathcal{B}(\mathbb{R}^n)$, on a :

$$\forall r \geq 0, \frac{\gamma_n(A_r) - \gamma_n(A)}{r} \geq \frac{\Phi(\Phi^{-1}(\gamma_n(A)) + r) - \Phi(\Phi^{-1}(\gamma_n(A)))}{r}$$

Un passage à la limite inférieure pour $r \rightarrow 0^+$ donne directement le résultat.

□

Le TCG permet facilement d'obtenir :

Proposition 7. *La fonction de concentration de γ_n sur \mathbb{R}^n vaut : $\alpha_{\gamma_n} = 1 - \Phi$.*

Démonstration. Fixons $A \in \mathcal{B}(\mathbb{R}^n)$ avec $\gamma_n(A) \geq \frac{1}{2}$. Alors pour tout $r \geq 0$,

$$1 - \gamma_n(A_r) \leq 1 - \Phi(\Phi^{-1}(\gamma_n(A)) + r) \leq 1 - \Phi\left(\Phi^{-1}\left(\frac{1}{2}\right) + r\right) = 1 - \Phi(r)$$

puisque $\Phi^{-1}\left(\frac{1}{2}\right) = 0$, d'où $\alpha_{\gamma_n}(r) \leq 1 - \Phi(r)$.

De plus, pour $H = \{x \in \mathbb{R}^n : x_1 < 0\}$, $\gamma_n(H) = \Phi(0) = \frac{1}{2}$ donc pour tout $r \geq 0$:

$$\alpha_{\gamma_n}(r) \geq 1 - \gamma_n(H_r) = 1 - \Phi(r)$$

Ainsi, on en déduit : $\alpha_{\gamma_n} = 1 - \Phi$.

□

Dans le cas gaussien, les inégalités de Lévy se réécrivent donc :

Théorème 11 (Concentration autour de la médiane). *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une application 1-lipschitzienne pour la norme euclidienne et de médiane $\text{Med}(f)$ par rapport à γ_n . Alors, pour tout $r \in \mathbb{R}_+$,*

$$\begin{aligned} \gamma_n(f \geq \text{Med}(f) + r) &\leq 1 - \Phi(r) \\ \gamma_n(f \leq \text{Med}(f) - r) &\leq 1 - \Phi(r) \end{aligned}$$

Le calcul classique de l'équivalent de la queue de la loi gaussienne donne lorsque $r \rightarrow +\infty$:

$$1 - \Phi(r) \sim \frac{1}{r\sqrt{2\pi}} \exp\left(-\frac{r^2}{2}\right)$$

On obtient ainsi des inégalités de concentration pour des fonctions lipschitziennes autour de la médiane légèrement plus fortes asymptotiquement que celles obtenues pour des concentrations autour de la moyenne dans la section précédente.

Références

- [1] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities : A Non-asymptotic Theory of Independence*. Oxford University Press, 2013.
- [2] Julien Bureaux. Propriétés d'isopérimétrie et de concentration gaussiennes. 2011.
- [3] Emmanuel Candes, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. 2005.
- [4] Emmanuel Candes and Terence Tao. Near optimal signal recovery from random projections : Universal encoding strategies ? 2006.
- [5] Kenneth R. Davidson and Stanislaw J. Szarek. *Local Operator Theory, Random Matrices and Banach Spaces*.
- [6] Michel Ledoux. Isoperimetry and gaussian analysis. 1994.

A Valeurs singulières de matrices

On fixe deux entiers naturels m, n vérifiant $m \geq n$.

Définition 9. Soit $M \in \mathcal{M}_{m,n}(\mathbb{R})$. La matrice $M^T M \in \mathcal{M}_n(\mathbb{R})$ est symétrique positive. On appelle valeurs singulières les racines carrées des valeurs propres de $M^T M$.

Dans la suite, on notera λ_{\max} (resp. λ_{\min}) l'application qui à une matrice associe sa plus grande (resp. sa plus petite) valeur singulière.

Proposition 8. Soit $M \in \mathcal{M}_{m,n}(\mathbb{R})$. On a alors :

$$\begin{aligned}\lambda_{\max}(M) &= \max_{u \in S^{n-1}} \|Mu\|_2 = \|M\| \\ \lambda_{\min}(M) &= \min_{u \in S^{n-1}} \|Mu\|_2 = - \max_{u \in S^{n-1}} \min_{v \in S^{m-1}} \langle Mu, v \rangle\end{aligned}$$

Démonstration. Par un calcul direct :

$$\left(\max_{u \in S^{n-1}} \|Mu\|_2 \right)^2 = \max_{u \in S^{n-1}} (\|Mu\|_2)^2 = \max_{u \in S^{n-1}} \langle Mu, Mu \rangle = \max_{u \in S^{n-1}} \langle M^T M u, u \rangle = (\lambda_{\max}(M))^2$$

où la dernière égalité s'obtient classiquement en diagonalisant $M^T M$ dans une base orthonormée. Ainsi :

$$\lambda_{\max}(M) = \max_{u \in S^{n-1}} \|Mu\| = \|M\|$$

On a de même : $\lambda_{\min}(M) = \min_{u \in S^{n-1}} \|Mu\|_2$

Il reste à montrer l'égalité : $\min_{u \in S^{n-1}} \max_{v \in S^{m-1}} \langle Mu, v \rangle = - \max_{u \in S^{n-1}} \min_{v \in S^{m-1}} \langle Mu, v \rangle$.

Fixons $u \in S^{n-1}$. L'image de l'application $\langle Mu, \cdot \rangle : S^{m-1} \rightarrow \mathbb{R}$ est symétrique par rapport à 0 donc $\max_{v \in S^{m-1}} \langle Mu, v \rangle = - \min_{v \in S^{m-1}} \langle Mu, v \rangle$. Le résultat en découle immédiatement. □

Proposition 9. Les applications $\lambda_{\max}, \lambda_{\min} : \mathcal{M}_{m,n}(\mathbb{R}) \rightarrow \mathbb{R}_+$ sont 1-lipschitziennes pour la norme d'opérateur $\|\cdot\|$ sur $\mathcal{M}_{m,n}(\mathbb{R})$, et donc pour la norme euclidienne sur \mathbb{R}^{mn} puisque $\|\cdot\| \leq \|\cdot\|_{2, \mathbb{R}^{mn}}$.

Démonstration. $\lambda_{\max} = \|\cdot\|$ est 1-lipschitzienne d'après l'inégalité triangulaire. Montrons le résultat pour λ_{\min} .

Soit $M, M' \in \mathcal{M}_{m,n}(\mathbb{R})$, on a :

$$\begin{aligned}\lambda_{\min}(M) &= \lambda_{\min}(M' + (M - M')) = \min_{u \in S^{n-1}} \|M'u + (M - M')u\|_2 \\ &\leq \min_{u \in S^{n-1}} (\|M'u\|_2 + \|(M - M')u\|_2) \\ &\leq \min_{u \in S^{n-1}} \|M'u\|_2 + \|M - M'\| \\ &= \lambda_{\min}(M') + \|M - M'\|\end{aligned}$$

En échangeant M et M' dans l'inégalité précédente, on conclut :

$$|\lambda_{\min}(M) - \lambda_{\min}(M')| \leq \|M - M'\|$$

□

B Remarques sur les mesures gaussiennes

Soient $p, q \in \mathbb{N}$. On identifie \mathbb{R}^{pq} et $\mathcal{M}_{p,q}(\mathbb{R})$ en voyant les colonnes d'une matrice comme les tranches successives d'un vecteur.

Dans ce mémoire, nous avons muni \mathbb{R}^{pq} de la mesure de probabilité gaussienne standard γ_{pq} , définie par sa densité contre la mesure de Lebesgue :

$$g_{pq} : x \in \mathbb{R}^{pq} \mapsto \frac{1}{\sqrt{2\pi}^{pq}} \exp\left(-\frac{\|x\|_2^2}{2}\right)$$

Par ailleurs, si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace de probabilité et $A : \Omega \rightarrow \mathcal{M}_{p,q}(\mathbb{R})$ est une matrice aléatoire gaussienne, on peut munir \mathbb{R}^{pq} de la mesure image par A que l'on note μ_A . Comme les coefficients de A sont des variables aléatoires indépendantes, c'est une mesure de probabilité qui admet une densité contre la mesure de Lebesgue :

$$x \in \mathbb{R}^{pq} \mapsto \frac{1}{\sqrt{2\pi}^{\frac{1}{p}pq}} \exp\left(-\frac{\|x\|_2^2}{2\frac{1}{p}}\right)$$

Il s'agit en fait d'une mesure gaussienne centrée non réduite, de variance $\frac{1}{p}$.

De manière plus générale on définit pour tout $\sigma > 0$ la mesure gaussienne centrée de variance σ^2 sur \mathbb{R}^n , que l'on note γ_n^σ , par sa densité contre la mesure de Lebesgue :

$$g_n^\sigma : x \in \mathbb{R}^n \mapsto \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left(-\frac{\|x\|_2^2}{2\sigma^2}\right)$$

On a ainsi :

$$\mu_A = \gamma_{pq}^{\frac{1}{\sqrt{p}}}$$

Par ailleurs, un simple changement de variable linéaire permet d'écrire pour tout borélien $B \in \mathcal{B}(\mathbb{R}^n)$:

$$\gamma_n^\sigma(B) = \gamma_n\left(\frac{1}{\sigma} \cdot B\right)$$

Dans le cas qui nous intéresse, cela se réécrit :

$$\mu_A(B) = \gamma_n(\sqrt{p} \cdot B)$$

On s'intéresse maintenant à la compatibilité entre les moyennes pour les deux mesures. Toujours par changement de variable linéaire, on obtient la proposition suivante :

Proposition 10. *Soit $\sigma > 0$ et $f : \mathbb{R}^n \rightarrow \mathbb{R}$ mesurable et positivement homogène (i.e. $f(\lambda x) = \lambda f(x)$ pour tout $\lambda \in \mathbb{R}_+$, $x \in \mathbb{R}^n$), alors f est intégrable par rapport à γ_n^σ si et seulement si f est intégrable par rapport à γ_n et dans ce cas :*

$$\mathbb{E}_{\gamma_n^\sigma}[f] = \sigma \mathbb{E}_{\gamma_n}[f]$$

On en déduit donc, par homogénéité et intégrabilité des applications λ_{\min} et λ_{\max} :

$$\mathbb{E}_{\mathbb{P}}[\lambda_{\min/\max}(A)] = \frac{1}{\sqrt{p}} \mathbb{E}_{\gamma_{pq}}[\lambda_{\min/\max}]$$