

Introduction au Compressed sensing.

Null Space Property et Restricted Isometry property

Guillaume Lécué¹

Résumé

Nous présentons les conditions *Null Space Property* et *Restricted Isometry Property* des matrices de mesures. Puis nous démontrons que les matrices aléatoires Gaussiennes vérifient ces propriétés avec grande probabilité pour le nombre optimal de mesures en $s \log(eN/s)$.

1 Introduction

La procédure du Basis Pursuit a été introduit par relaxation convexe dans le cours précédent. On rappelle sa définition.

Définition 1.1 ([2, 1]). *Etant donné une matrice $A \in \mathbb{R}^{m \times N}$ et $y \in \mathbb{R}^m$ la procédure de **Basis Pursuit** (BP) renvoie*

$$\hat{x} \in \underset{t \in \mathbb{R}^N : At=y}{\operatorname{argmin}} \|t\|_1 \quad (\text{BP})$$

quand il y a une solution à l'équation $At = y$ et \emptyset sinon.

On a vu qu'on pouvait réécrire (BP) comme un problème de programmation linéaire et que donc on pouvait implémenter cette procédure de manière efficace même en grandes dimensions N .

D'un point de vue théorique, on a montré que si une matrice de mesure $A \in \mathbb{R}^{m \times N}$ permettait la reconstruction de n'importe quel signal de Σ_s par Basis Pursuit alors nécessairement $m \gtrsim s \log(eN/s)$, c'est à dire on doit avoir au moins $s \log(eN/s)$ mesures linéaires pour résoudre le problème du Compressed Sensing par BP sur Σ_s .

Dans ce chapitre, on va introduire des conditions sur la matrice A qui assurent la reconstruction exacte par BP de tous les vecteurs de Σ_s . La propriété qu'on cherche à obtenir est rappelée ici.

Définition 1.2. *Soit $A : \mathbb{R}^N \rightarrow \mathbb{R}^m$ une matrice telle que $m \leq N$ et s un entier plus petit que N . On dit que A vérifie la propriété de **reconstruction exacte d'ordre s** quand pour tout vecteur s -sparse x , on a*

$$\underset{t \in \mathbb{R}^N}{\operatorname{argmin}} (\|t\|_1 : At = Ax) = \{x\}.$$

On dit alors que A vérifie $RE(s)$.

On sait déjà que si A vérifie $RE(s)$ alors nécessairement A a au moins $s \log(eN/s)$ lignes (toujours à constante absolue près). L'objectif de ce chapitre est de montrer qu'on peut construire des matrices vérifiant $RE(s)$ et ayant un nombre de lignes *minimal*, c'est à dire de l'ordre de grandeur de $s \log(eN/s)$. On verra que des matrices aléatoires vérifient ces propriétés.

1. CNRS, CREST, ENSAE, Université Paris-Saclay. Bureau E31, 3 avenue Pierre Larousse. 92 245 Malakof. Email : guillaume.lecue@ensae.fr.

2 Null Space Property et Restricted Isometry Property

Dans le chapitre précédent, on a vu que le noyau de A joue un rôle important pour la procédure de minimisation ℓ_0 : cette procédure résoud le problème de CS sur Σ_s si et seulement si $\text{Ker}(A) \cap \Sigma_{2s} = \{0\}$.

Le seul élément autorisé à être parcimonieux dans $\text{Ker}(A)$ est donc 0. Cette condition est aussi nécessaire tout simplement pour avoir $Ax \neq Ay$ pour tout $x \neq y \in \Sigma_s$, càd pour que le problème de CS puisse avoir une chance d'être résolu sur Σ_s .

Dans cette section, on montre que $\text{Ker}(A)$ joue aussi un rôle essentiel pour la propriété $RE(s)$. On va vouloir que $\text{Ker}(A)$ soit en quelque sorte loin de Σ_{2s} et non pas seulement d'intersection nulle avec Σ_{2s} . Cette propriété est connue en CS sous le nom de *Null Space Property*. On rappelle que si $v \in \mathbb{R}^N$ et $J \subset \{1, \dots, N\}$ alors v_J est le vecteur de \mathbb{R}^N coïncidant avec v sur J et de coordonnées nulles en dehors de J .

Définition 2.1. Soit $A \in \mathbb{R}^{m \times N}$ et $1 \leq s \leq N$. On dit que A satisfait la **Null Space Property d'ordre s** quand pour tout $J \subset \{1, \dots, N\}$ tel que $|J| = s$ et pour tout $v \in \text{Ker}(A) \setminus \{0\}$, on a :

$$\|v_J\|_1 < \|v_{J^c}\|_1 \quad (2.1)$$

On dit alors que A vérifie $NSP(s)$.

Intuitivement, la condition (2.1) dit que si v est un élément non-nul de $\text{Ker}(A)$ alors il ne peut pas exister s coordonnées de v qui contiennent plus de la moitié de sa "masse" au sens ℓ_1 . On peut réécrire (2.1) par $2\|v_J\|_1 < \|v\|_1$ et ceci pour tout $J \subset \{1, \dots, N\}$ de cardinal s . On peut en particulier choisir les s plus grands coefficients en valeur absolue de v et ainsi la condition $NSP(s)$ est équivalente à : pour tout $v \in \text{Ker}(A) \setminus \{0\}$

$$v_1^* + \dots + v_s^* < \frac{\|v\|_1}{2}$$

où $v_1^* \geq v_2^* \geq \dots \geq v_N^* \geq 0$ est la suite décroissante du réarrangement de $(|v_j|)_{j=1}^N$. En quelque sorte les vecteurs non nuls de $\text{Ker}(A)$ doivent être à l'opposé des vecteurs sparse ; on dit qu'ils sont "bien étalés".

En particulier, les vecteurs non-nuls $2s$ -sparse ne peuvent pas être dans $\text{Ker}(A)$ vu que si $v \in \Sigma_{2s}$ alors $v_1^* + \dots + v_s^* \geq v_{s+1}^* + \dots + v_{2s}^* = \|v\|_1 - (v_1^* + \dots + v_s^*)$. En particulier la Null Space Property implique la condition nécessaire et suffisante de reconstruction exacte par ℓ_0 .

Proposition 2.2. Soit $A \in \mathbb{R}^{m \times N}$. Si A vérifie $NSP(s)$ alors $\text{Ker}(A) \cap \Sigma_{2s} = \{0\}$.

Mais la condition $NSP(s)$ est plus forte que la condition $\text{Ker}(A) \cap \Sigma_{2s} = \{0\}$ du chapitre précédent. Les vecteurs qui sont "approximativement" s -sparse ne peuvent pas non plus être dans $\text{Ker}(A)$ sous $NSP(s)$; par exemple, si A vérifie $NSP(s)$ alors

$$v = \left(\underbrace{1, \dots, 1}_s, \underbrace{\frac{s}{N-s}, \dots, \frac{s}{N-s}}_{N-s} \right)^\top \notin \text{Ker}(A) \quad (2.2)$$

car $\|v_J\|_1 = \|v_{J^c}\|_1$ pour $J = \{1, \dots, s\}$. En d'autres termes, v n'est pas assez "étalé", il concentre plus de la moitié de sa masse sur seulement s coordonnées.

On montre maintenant que la Null Space Property $NSP(s)$ est une condition nécessaire et suffisante pour que A satisfasse $RE(s)$.

Théorème 2.3. Soit $A \in \mathbb{R}^{m \times N}$. Les deux assertions suivantes sont équivalentes :

1. A vérifie $RE(s)$
2. A vérifie $NSP(s)$.

Démonstration. Si A vérifie $RE(s)$ alors pour tout $x \in \Sigma_s$, x est l'unique élément de $x + \text{Ker}(A)$ de norme ℓ_1 minimale et donc pour tout $v \in \text{Ker}(A) \setminus \{0\}$, $\|x\|_1 < \|x + v\|_1$. On a alors pour tout $v \in \text{Ker}(A) \setminus \{0\}$ et $J \subset \{1, \dots, N\}$ tel que $|J| \leq s$ que pour $x = -v_J \in \Sigma_s$,

$$\|v_J\|_1 = \|x\|_1 < \|x + v\|_1 = \|v_{J^c}\|_1.$$

Ceci étant vrai pour tout $v \in \text{Ker}(A) \setminus \{0\}$ et $J \subset \{1, \dots, N\}$ tel que $|J| \leq s$, on en déduit que A vérifie $NSP(s)$.

Réciproquement, on suppose que A vérifie $NSP(s)$. Soit $x \in \Sigma_s$. On note $J = \text{supp}(x)$. Soit $v \in \text{Ker}(A) \setminus \{0\}$. On a

$$\|x + v\|_1 = \|v_{J^c}\|_1 + \|v_J + x\|_1 \geq \|x\|_1 + \|v_{J^c}\|_1 - \|v_J\|_1 > \|x\|_1.$$

Donc A vérifie $RE(s)$. ■

La condition $NSP(s)$ est donc une condition nécessaire et suffisante pour qu'une matrice A vérifie la propriété $RE(s)$. D'un point de vue théorique, il suffit maintenant de construire des matrices de mesures A vérifiant $NSP(s)$ et ayant un nombre minimal de lignes pour résoudre le problème de CS sur Σ_s via la procédure du Basis Pursuit.

Cependant la condition $NSP(s)$ n'est pas facile à vérifier directement. Concrètement, on ne sait pas construire des matrices A pour lesquelles on vérifie directement $NSP(s)$. On va donc avoir recours à deux autres conditions sur A qui vont impliquer $NSP(s)$. Ces deux conditions sont suffisantes – contrairement à NSP qui est aussi nécessaire – pour avoir $RE(s)$. Pour les cas qui nous intéressent dans ce cours, ces conditions seront celles qu'on cherchera à démontrer.

La première de ces deux conditions est liée à des résultats en théorie asymptotique des espaces de Banach. Elle concerne ce qui est appelé les *sections euclidiennes* de B_1^N . On renvoie le lecteur intéressé au chapitre dédié à ces problématiques pour plus de détails. Pour le moment, la condition qui suit dit que le noyau de A va intersecter la boule unité B_1^N de la norme ℓ_1^N d'une manière très particulière.

Pour fixer les notations, on rappelle que pour tout $p \geq 1$,

$$B_p^N = \{x \in \mathbb{R}^N : \|x\|_p \leq 1\} \text{ et on a } \frac{1}{\sqrt{N}} B_2^N \subset B_1^N \subset B_2^N,$$

autrement dit : pour tout $x \in \mathbb{R}^N$, $\|x\|_2 \leq \|x\|_1 \leq \sqrt{N} \|x\|_2$. Les deux bornes sont atteintes par $(1, 0, \dots, 0)^\top$ et $(1, \dots, 1)^\top$ respectivement. En particulier,

$$\text{diam}(B_1^N, \ell_2^N) := \sup_{x \in B_1^N} \|x\|_2 = 1$$

et cette borne est atteinte en $(1, 0, \dots, 0)^\top$.

Le diamètre ℓ_2^N est extrémal sur les axes canoniques de \mathbb{R}^N qui forment en quelques sorte les vecteurs les plus parcimonieux de \mathbb{R}^N vu qu'ils n'ont qu'une seule coordonnée non nulle.

Sous $NSP(s)$, le noyau $\text{Ker}(A)$ ne contient pas, en particulier, de vecteurs parcimonieux de Σ_{2s} – mis à part 0 – $\text{Ker}(A)$ est donc loin des axes canoniques et de tous les espaces de

dimensions $2s$ engendrés par ces vecteurs. Intuitivement, $\text{Ker}(A)$ va être positionné selon des *directions diagonales* de \mathbb{R}^N . Or pour ces vecteurs, qu'on dit *well-spread* ou *bien étalé*, le ratio de leur norme ℓ_2 sur leur norme ℓ_1 est beaucoup plus petit que 1. On peut alors s'attendre à ce que $NSP(s)$ soit vérifié si la norme ℓ_2 des vecteurs de B_1^N qui sont dans le noyau $\text{Ker}(A)$ est bien plus petite que 1. C'est ce qui est représenté dans la Figure 2 et formalisé dans la définition suivante.

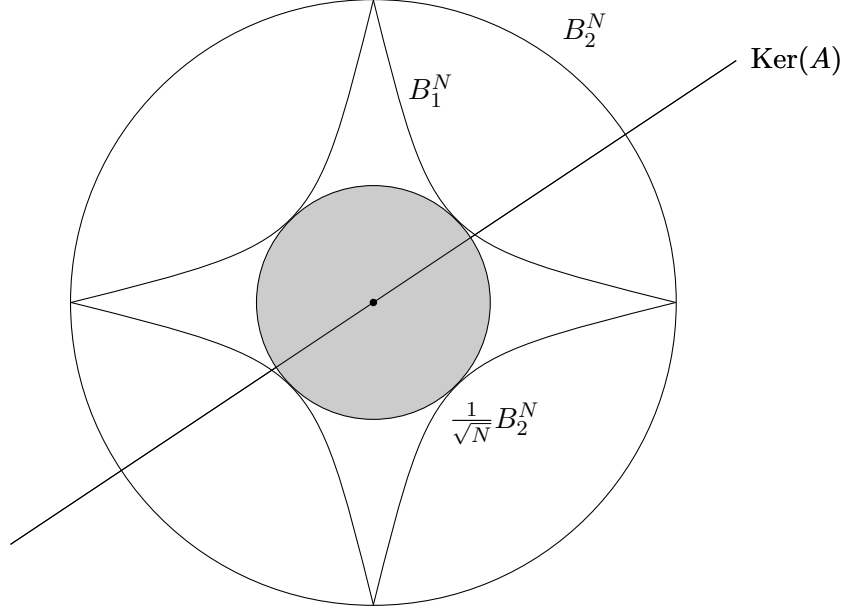


FIGURE 1 – On a $\text{diam}(B_1^N, \ell_2^N) = 1$. Mais si $\text{diam}(B_1^N \cap \text{Ker}(A), \ell_2^N) < 1/(2\sqrt{s})$ alors $NSP(s)$ est vérifiée par A .

Définition 2.4. Soit $A \in \mathbb{R}^{m \times N}$ et $1 \leq s \leq N$. On dit que A vérifie la **propriété de Gelfand d'ordre s** quand

$$\text{diam}(B_1^N \cap \text{Ker}(A), \ell_2^N) < \frac{1}{2\sqrt{s}}.$$

On dit alors que A vérifie *Gelfand(s)*.

La dimension du noyau $\text{Ker}(A)$ est au moins $N - m$. On a habituellement $m \ll N$, donc $\text{Ker}(A)$ est un espace linéaire de très grande dimension presque égale à N . Or, on a vu que $\text{diam}(B_1^N, \ell_2^N) = 1$, ici la condition *Gelfand(s)* "demande" que $\text{diam}(B_1^N \cap \text{Ker}(A), \ell_2^N) \leq 1/(2\sqrt{s})$, c'est-à-dire que la restriction de B_1^N à $\text{Ker}(A)$ ait un diamètre beaucoup plus petit que 1 même si $\text{Ker}(A)$ est de très grande dimension. C'est donc une condition très contraignante sur $\text{Ker}(A)$, mais elle est suffisante pour prouver la Null Space Property.

Théorème 2.5. Soit $A \in \mathbb{R}^{m \times N}$ et $1 \leq s \leq N$. Si A vérifie *Gelfand(s)* alors A vérifie $NSP(s)$.

Démonstration. On suppose que A vérifie *Gelfand(s)* alors pour tout $v \in \text{Ker}(A) \setminus \{0\}$, on a $\|v\|_2 < \|v\|_1 / (2\sqrt{s})$.

Soit $v \in \text{Ker}(A) \setminus \{0\}$ et $J \subset \{1, \dots, N\}$ tel que $|J| = s$. On a

$$\|v_J\|_1 \leq \sqrt{s} \|v_J\|_2 \leq \sqrt{s} \|v\|_2 < \sqrt{s} \frac{\|v\|_1}{2\sqrt{s}} = \frac{\|v\|_1}{2}.$$

■

Nous introduisons, finalement, une dernière propriété sur la matrice de mesures. C'est cette propriété qui a eu le plus de succès en CS pour démontrer $RE(s)$.

Définition 2.6. Soit $A \in \mathbb{R}^{m \times N}$ et $1 \leq s \leq N$. On dit que A vérifie la **Restricted Isometry Property d'ordre s** quand pour tout $x \in \Sigma_s$,

$$\frac{1}{2} \|x\|_2 \leq \frac{\|Ax\|_2}{\sqrt{m}} \leq \frac{3}{2} \|x\|_2$$

On dit alors que A vérifie $RIP(s)$.

La condition RIP est la plus populaire en CS. Si A vérifie $RIP(s)$ alors elle agit comme une isométrie (le terme correct serait une *isomorphie* à cause des constantes $1/2$ et $3/2$) sur l'ensemble des vecteurs Σ_s .

On rappelle que A est une matrice de compression, elle a donc un noyau $\text{Ker}(A)$ de très grande dimension. En particulier, A ne peut pas être une isométrie sur tout l'espace \mathbb{R}^N . Ce que demande $RIP(s)$ est que A agisse comme une isométrie restreinte à un sous-ensemble de \mathbb{R}^N , ici Σ_s .

Remarque 2.7. Les conditions $RE(s)$, $NSP(s)$ et $Gelfand(s)$ ne font intervenir A qu'au travers de son noyau. En particulier, ces propriétés sont inchangées pour toute renormalisation de A , vu que $\text{Ker}(\lambda A) = \text{Ker}(A)$ pour tout $\lambda \neq 0$. Dans le cas de la condition $RIP(s)$, la matrice A intervient directement et la normalisation joue donc un rôle dans sa définition.

Dans la Définition 2.6, nous avons fait le choix de considérer que la matrice A est la matrice des vecteurs mesure, c'à-d la matrice dont les lignes sont les vecteurs de mesure X_1, \dots, X_m :

$$A = \begin{pmatrix} X_1^\top \\ \vdots \\ X_m^\top \end{pmatrix}$$

Pour ce type de matrice (sans renormalisation) on présente la condition $RIP(s)$ où le terme $\|Ax\|_2^2 / \sqrt{m}$ est comparé à $\|x\|_2$.

Il existe aussi une autre normalisation classique en CS qui consiste à diviser par $1/\sqrt{m}$ la matrice A :

$$\Gamma = \frac{1}{\sqrt{m}} A = \frac{1}{\sqrt{m}} \begin{pmatrix} X_1^\top \\ \vdots \\ X_m^\top \end{pmatrix}.$$

Pour cette renormalisation, $RIP(s)$ consistera à comparer $\|\Gamma x\|_2$ (sans facteur $1/\sqrt{m}$) à $\|x\|_2$: pour tout $x \in \Sigma_s$, $(1/2) \|x\|_2 \leq \|\Gamma x\|_2 \leq (3/2) \|x\|_2$.

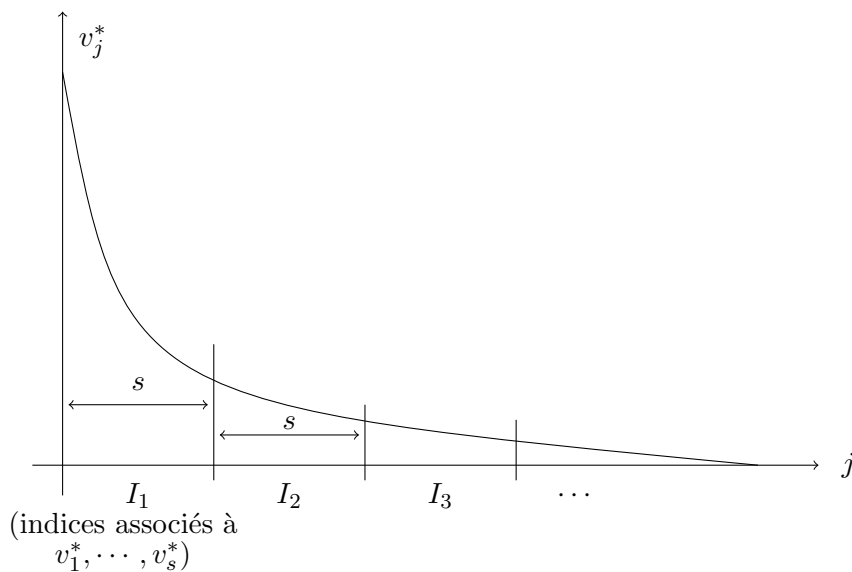
Un résultat clef en CS est que pour construire une matrice de mesures satisfaisant la condition $RE(s)$ il suffit de vérifier que A satisfait $RIP(s)$.

Théorème 2.8. Soit $A \in \mathbb{R}^{m \times N}$ et $1 \leq s \leq N$. Si A vérifie $RIP(s)$ alors A vérifie $Gelfand(s/65)$.

Remarque 2.9. *A aucun moment, on ne cherchera à optimiser les constantes numériques. On privilégie la simplicité des arguments à leurs précision concernant les constantes numériques. Ici le paramètre $s/65$ n'est pas optimal (il peut en fait être ramené à $s/2$). On donne de l'importance aux bons ordres de grandeurs et à la simplicité de l'argumentation avant l'exactitude des constantes numériques.*

Démonstration. Soit A une matrice de $\mathbb{R}^{m \times N}$ vérifiant $RIP(s)$. Soit $v \in \text{Ker}(A) \setminus \{0\}$. Montrons que $\|v\|_2 < \|v\|_1 / (2\sqrt{s/65})$.

On décompose v selon l'amplitude des valeurs absolues de ses coordonnées en blocs de taille s (cf. Figure 2). On rappelle que $v_1^* \geq v_2^* \geq \dots \geq v_N^*$ est la suite décroissante du réarrangement de $(|v_j|)_{j=1}^N$. On note I_1 l'ensemble des indices des s plus grandes coordonnées de $(|v_j|)_{j=1}^N$, I_2 les s suivantes, etc..



Pour tout $j \geq 2$, on a

$$\|v_{I_j}\|_2 \leq \sqrt{s} \|v_{I_j}\|_\infty \leq \frac{\|v_{I_{j-1}}\|_1}{\sqrt{s}}. \quad (2.3)$$

On traite d'abord le premier block. D'après la condition $RIP(s)$, A agit comme une isométrie sur tous les vecteurs de Σ_s , de plus $Av = 0$, on a alors

$$\|v_{I_1}\|_2 \leq 2 \frac{\|Av_{I_1}\|_2}{\sqrt{m}} = \frac{2}{\sqrt{m}} \|A(v_{I_1} - v)\|_2 = \frac{2}{\sqrt{m}} \left\| A \left(\sum_{j \geq 2} v_{I_j} \right) \right\|_2 \leq 2 \sum_{j \geq 2} \frac{\|Av_{I_j}\|_2}{\sqrt{m}} \leq 3 \sum_{j \geq 2} \|v_{I_j}\|_2.$$

Finalement, en utilisant la relation (2.3) sur chaque terme de la somme précédente, on obtient

$$\|v_{I_1}\|_2 \leq \frac{3}{\sqrt{s}} \sum_{j \geq 2} \|v_{I_{j-1}}\|_1 = \frac{3}{\sqrt{s}} \|v\|_1.$$

On traite maintenant les blocs suivants. On applique la formule (2.3) à chaque terme :

$$\|v_{I_2} + \dots\|_2 \leq \sum_{j \geq 2} \|v_{I_j}\|_2 \leq \frac{1}{\sqrt{s}} \sum_{j \geq 2} \|v_{I_{j-1}}\|_1 = \frac{\|v\|_1}{\sqrt{s}}.$$

On a finalement :

$$\|v\|_2 \leq \|v_{I_1}\|_2 + \|v_{I_2} + \dots\|_2 \leq \frac{4}{\sqrt{s}} \|v\|_1 < \frac{1}{2\sqrt{s/65}} \|v\|_1.$$

■

Conclusion : On a démontré la suite d'implication suivante :

$$RIP(65s) \implies \text{Gelfand}(s) \implies NSP(s) \iff RE(s).$$

Il suffit donc de construire une matrice vérifiant $RIP(s)$ pour avoir une matrice de mesures permettant la reconstruction de tout vecteur de Σ_s par le Basis Pursuit.

3 Matrices aléatoires sous-gaussiennes et condition RIP

Dans cette section, on construit des matrices vérifiant $RIP(s)$ pour un nombre optimal de mesures en $m \sim s \log(eN/s)$. Ce nombre de mesures est optimal vu que $RIP(s)$ implique $RE(c_0s)$ (pour une constante absolue c_0) et que $RE(c_0s)$ implique nécessairement que $m \gtrsim s \log(eN/s)$. On ne peut donc pas construire de matrice vérifiant $RIP(s)$ ayant moins de $\sim s \log(eN/s)$ lignes (à constante absolue près).

Il se trouve que, jusqu'à ce jour, les seules matrices qui ont été construites vérifiant $RIP(s)$ pour un nombre optimal de lignes en $s \log(eN/s)$ sont des matrices aléatoires.

Les résultats que nous établirons dans la suite sont donc du type *avec grande probabilité la matrice A vérifie $RIP(s)$* . Ce sont donc des résultats qui ne sont vrais qu'avec grande probabilité : on n'est donc pas absolument sûr que les matrices construites ainsi vérifient RIP. De plus, vérifier si une matrice donnée satisfait RIP est un problème NP-hard.

Il est un peu décevant d'avoir des résultats qui n'ont lieu qu'en probabilité cependant :

1. les probabilités qu'on obtiendra seront très proches de 1
2. si on s'autorise un nombre de mesures plus grand que $s \log(ed/s)$ alors on peut construire des matrices déterministes vérifiant $RIP(s)$ pour $m = s^2$. Si on le souhaite et si possible, on peut donc perdre sur le nombre de mesures à prendre (car $s^2 \gg s \log(eN/s)$ en général) mais ne plus considérer que des mesures déterministes.

Remarque 3.1. *La construction de matrices déterministes vérifiant $RIP(s)$ pour un nombre de lignes de l'ordre de $s \log(epN/s)$ est un des problèmes mathématiques les plus difficile en CS. Une telle construction donnerait des preuves déterministes à des résultats importants en mathématiques (Théorème de Dvoretzky, Kashin, etc.) qui n'ont pour le moment que des preuves faisant appel à l'aléatoire.*

Il faut préciser ce qu'on entend par matrice déterministe. Premièrement, la réalisation d'une matrice aléatoire n'est pas une construction déterministe. Ce qu'on veut, c'est une procédure ne faisant appel à aucun moment à de l'aléatoire qui soit en mesure de remplir les $m \times N$ entrées de la matrice A et telle qu'en sortie cette matrice vérifie $RIP(s)$ et $m \sim s \log(eN/s)$. Ce problème est de nos jours toujours ouvert.

3.1 Matrices aléatoires Gaussiennes et concentration

L'exemple standard de matrices aléatoires est celui des matrices Gaussiennes. Elle se construisent à partir d'une famille de variables aléatoires Gaussiennes sur \mathbb{R} i.i.d. centrées et réduites.

Définition 3.2. Une matrice aléatoire Gaussienne standard G est une matrice de taille $m \times N$

$$G = \begin{pmatrix} g_{11} & \cdots & g_{1N} \\ \cdots & \cdots & \cdots \\ g_{m1} & \cdots & g_{mN} \end{pmatrix}$$

où g_{11}, \dots, g_{mN} sont mN variables aléatoires Gaussiennes i.i.d. centrées réduites.

Cette définition est bien consistante avec la définition classique : un vecteur aléatoire G est Gaussien quand l'image de G par toutes formes linéaires sur cette espace est une variable aléatoire Gaussienne sur \mathbb{R} . Ainsi un **vecteur aléatoire Gaussien standard** de \mathbb{R}^N est un vecteur de \mathbb{R}^N dont les coordonnées sont des variables aléatoires gaussiennes i.i.d. centrées réduites. La distribution d'un tel vecteur sera noté $\mathcal{N}_N(0, I_N)$ où, ici, la matrice de covariance est I_N , la matrice identité de \mathbb{R}^N .

La seule propriété des matrices aléatoires Gaussiennes que nous utiliserons dans ce cours est une propriété de déviation.

Pour comprendre ce qu'est une propriété de déviation, on commence d'abord par expliquer ce qu'est une propriété de concentration.

Un résultat de concentration s'énonce de la manière suivante : par exemple, si g est une variable aléatoire Gaussienne réelle $\mathcal{N}(\mu, \sigma^2)$ alors pour tout $t > 0$,

$$\mathbb{P}[|g - \mu| \geq t\sigma] \leq 2 \exp(-t^2/2). \quad (3.1)$$

L'idée est que si g est distribuée selon une loi $\mathcal{N}(\mu, \sigma^2)$ alors elle sera avec probabilité au moins $1 - 2 \exp(-t^2/2)$ dans l'intervalle centré en μ et de longueur $2t\sigma$. La variable g est donc "concentrée" autour de sa moyenne μ .

Toutes les variables aléatoires ne sont pas concentrées autour de leur moyenne. Elles n'ont même pas forcément toute une moyenne : le cas typique est celui des variables de Cauchy qui sont absolument continues par rapport à la mesure de Lebesgue et de densité $f(t) = [\pi(1+t^2)]^{-1}$ pour $t \in \mathbb{R}$.

D'autres variables se concentrent autour de leur moyenne mais en quelques sorte "moins" que les variables Gaussiennes. Par exemple, une variables exponentielle symétrique e , càd de densité $f(t) = (2\sigma)^{-1} \exp(-|t - \mu|/\sigma)$ pour $t \in \mathbb{R}$, vérifie : pour tout $t > 0$,

$$\mathbb{P}[|e - \mu| \geq t\sqrt{2}\sigma] = \exp(-\sqrt{2}t). \quad (3.2)$$

On dit que e se concentre "en $\exp(-t)$ " (autour de sa moyenne qui vaut μ ici – sa variance valant $2\sigma^2$).

On peut en fait trouver n'importe quel type de concentration en " $\exp(-t^\alpha)$ " pour $0 < \alpha \leq 2$ ou en " $1/t^p$ " pour $p \geq 1$. Le comportement qui est le plus rechercher est le comportement dit **sous-gaussien** càd en " $\exp(-t^2)$ " comme dans (3.1). On dira aussi qu'une variable qui se concentre en " $\exp(-t)$ " est **sous-exponentielle**.

On en vient maintenant à la propriété de déviation. Contrairement à la concentration, la déviation ne s'intéresse qu'au comportement de la queue de distribution càd à $t \mapsto \mathbb{P}[|X| \geq t]$ quand t est "grand". En particulier, le comportement de X autour de sa moyenne n'a pas d'importance pour sa propriété de déviation. De même que pour la concentration, on parlera de **déviation sous-gaussienne** quand pour tout $t \geq c_0$, $\mathbb{P}[|X| \geq t] \leq 2 \exp(-c_1 t^2)$ pour $c_0 > 0$ assez grand et de **déviation sous-exponentielle** quand $t \mathbb{P}[|X| \geq t] \leq 2 \exp(-c_1 t^2)$ pour tout $t \geq c_0$.

Une manière assez utile de caractériser le type de déviation d'une variable aléatoire est donnée par la vitesse de croissance de ses moments.

Proposition 3.3. Soit X une variable aléatoire réelle et $\alpha \geq 1$. Il y a équivalence entre les deux propriétés suivantes :

1. il existe une constante absolue $c_0 > 0$ telle que pour tout $p \geq \alpha$, $\|X\|_{L_p} \leq c_0 p^{1/\alpha}$,
2. il existe deux constantes absolues c_1, c_2 telles que pour tout $t \geq c_1$, $\mathbb{P}[|X| \geq t] \leq \exp(-t^\alpha/c_2^\alpha)$.

Démonstration. On suppose que la condition sur la croissance des moments en $p^{1/\alpha}$ a lieu. On a par l'inégalité de Markov que pour tout $t > 0$,

$$\mathbb{P}[|X| \geq t] \leq \inf_{p>0} \frac{\|X\|_{L_p}^p}{t^p} \leq \inf_{p \geq \alpha} \frac{c_0^p p^{p/\alpha}}{t^p} = \inf_{p \geq \alpha} \exp\left(p \log\left(\frac{c_0 p^{1/\alpha}}{t}\right)\right).$$

Pour $t \geq ec_0 \alpha^{1/\alpha}$, on obtient le résultat désiré pour $p = [t/(ec_0)]^\alpha$ dans l'inégalité précédente.

Réciproquement, on suppose que X a une déviation en "exp(-t $^\alpha$)". On commence par montrer que pour $c = 2 \max(c_1, c_2)$ on a $\mathbb{E} \exp[(|X|/c)^\alpha] \leq e$. On rappelle que pour tout $a > 0$, on a

$$\int_a^\infty \alpha t^{\alpha-1} \exp(-t^\alpha b) dt = \frac{1}{b} \exp(-ba^\alpha)$$

et que si X est une v.a.r. positive et $f : \mathbb{R}^+ \mapsto \mathbb{R}$ est une fonction croissante et \mathcal{C}^1 alors par Fubini

$$\mathbb{E} f(X) - f(0) = \int_0^\infty f'(t) \mathbb{P}[|X| \geq t] dt.$$

On a alors :

$$\begin{aligned} \mathbb{E} \exp\left(\frac{|X|^\alpha}{c^\alpha}\right) - 1 &= \int_0^{+\infty} \alpha u^{\alpha-1} \exp(u^\alpha) \mathbb{P}[|X| > cu] du \\ &\leq \int_0^{c_1/c} \alpha u^{\alpha-1} \exp(u^\alpha) du + \int_{c_1/c}^{+\infty} \alpha u^{\alpha-1} \exp\left[u^\alpha \left(1 - \frac{c^\alpha}{c_2^\alpha}\right)\right] du \\ &= \exp\left(\frac{c_1^\alpha}{c^\alpha}\right) - 1 + \frac{1}{(c/c_2)^\alpha - 1} \exp\left[-\left(\frac{c^\alpha}{c_2^\alpha} - 1\right) \left(\frac{c_1}{c}\right)^\alpha\right] \\ &\leq \exp\left(\frac{c_1^\alpha}{c^\alpha}\right) - 1 + \exp\left[-\left(\frac{c_1}{c}\right)^\alpha\right] \leq e - 1 \end{aligned}$$

car $\exp(x) + \exp(-x) \leq e$ quand $0 \leq x \leq 1/2$. Par ailleurs, pour tout entier q et tout $x \geq 0$, $\exp(x) \geq x^q/q!$ donc

$$\mathbb{E}|X|^{\alpha q} = c^{\alpha q} \mathbb{E}\left[\left(\frac{|X|^\alpha}{c^\alpha}\right)^q\right] \leq c^{\alpha q} q! \mathbb{E} \exp\left(\frac{|X|^\alpha}{c^\alpha}\right) \leq eq! c^{\alpha q} \leq eq^q c^{\alpha q}.$$

Soit $p \geq \alpha$. Soit q un entier tel que $q\alpha \leq p < (q+1)\alpha$. On a :

$$\|X\|_{L_p} \leq \|X\|_{L_{(q+1)\alpha}} \leq e^{1/(q+1)\alpha} c(q+1)^{1/\alpha} \leq e^{1/p} c(2p/\alpha)^{1/\alpha} \leq 2ecp^{1/\alpha}.$$

■

Une variable aléatoire a donc une déviation sous-gaussienne (resp. sous-exponentielle) si et seulement si ses moments d'ordre p croissent moins vite que \sqrt{p} (resp. p).

Dans la suite, on s'intéressera au carré de variables aléatoires car la condition $RIP(s)$ fait intervenir $(1/m) \|Ax\|_2^2$ qui est une moyenne de m carrés de variables aléatoires. On a donc besoin de connaître les propriétés de déviation de ce type de variables. Pour cela, on se sert de la caractérisation de la déviation par la vitesse de croissance des moments donnée dans la Proposition 3.3.

Proposition 3.4. *Soit X une variable aléatoire réelle sous-gaussienne. Alors X^2 est sous-exponentielle.*

Démonstration. D'après la Proposition 3.3, comme X est sous-gaussien, pour tout $p \geq 2$, $\|X\|_{L_p} \leq c_0\sqrt{p}$ et donc $\|X^2\|_{L_p} = \|X\|_{L_{2p}}^2 \leq 2c_0^2p$. Donc, d'après la Proposition 3.3, X^2 est bien sous-exponentielle. ■

Il y a un type de variables aléatoires qu'on s'attend à voir se concentrer (et donc à dévier) de manière sous-gaussienne. Ce sont les moyennes empiriques de variables i.i.d.. En effet, si $(X_m)_m$ est une suite de variables i.i.d. ayant deux moments alors, pour $\bar{X}_m = \frac{1}{m} \sum_{i=1}^m X_i$, le TCL dit que $\sqrt{m}(\bar{X}_m - \mathbb{E}X_1)$ est asymptotiquement gaussien : pour tout $t > 0$, quand $m \rightarrow +\infty$,

$$\mathbb{P}[\sqrt{m}|\bar{X}_m - \mathbb{E}X_1| \geq t\sigma] \longrightarrow \mathbb{P}[|g| > t]$$

où σ est l'écart-type de X_1 . Les déviations de $\sqrt{m}(\bar{X}_m - \mathbb{E}X_1)/\sigma$ vont donc être similaires à celles d'une Gaussienne standard quand $m \rightarrow +\infty$.

Le TCL est un résultat asymptotique, il n'est vrai que pour des grandes valeurs de m . Les résultats que nous voulons établir en CS sont pour un nombre fixe d'observations m . On souhaite obtenir des résultats de concentration de \bar{X}_m à m fixé. Ce type de résultat est dit non-asymptotique.

Proposition 3.5 (Inégalité de Bernstein). *Soit Z_1, \dots, Z_m des variables aléatoires réelles i.i.d. de même loi que Z telles que pour tout $p \geq 1$,*

$$\|Z\|_{L_p} \leq C_0p \tag{3.3}$$

où C_0 est une constante absolue.

Alors, il existe une constante absolue $c_1 > 0$ telle que pour tout $t > 0$,

$$\mathbb{P}\left[\left|\frac{1}{m} \sum_{i=1}^m Z_i - \mathbb{E}Z\right| \geq tC_0\right] \leq 2 \exp(-c_1 m \min(t^2, t)).$$

Démonstration. Sans perte de généralité, on suppose que les Z_i sont centrées. De plus,

$$\mathbb{P}[|\bar{Z}_m| \geq tC_0] = \mathbb{P}[\bar{Z}_m \geq tC_0] + \mathbb{P}[-\bar{Z}_m \geq tC_0]$$

donc, quitte à changer Z en $-Z$, il suffit de majorer $\mathbb{P}[\bar{Z}_m \geq tC_0]$.

Une manière classique de démontrer des inégalités de concentration est de passer par la transformée de Laplace. On a par l'inégalité de Markov :

$$\mathbb{P}\left[\frac{1}{m} \sum_{i=1}^m Z_i \geq t\right] \leq \inf_{\lambda > 0} \frac{\mathbb{E} \exp(\lambda \bar{Z}_m)}{\exp(\lambda t)} = \inf_{\lambda > 0} \left(e^{-\lambda t} \prod_{i=1}^m \mathbb{E} \exp\left(\frac{\lambda Z_i}{m}\right) \right). \tag{3.4}$$

On est donc amené à majorer $\mathbb{E} \exp(\mu Z)$ quand Z est une variable centrée sous-exponentielle et $\mu > 0$. En utilisant le développement de Taylor de la fonction exponentielle, on a

$$\mathbb{E} \exp(\mu Z) - 1 \leq \sum_{k \geq 2} \frac{\mu^k \mathbb{E}|Z|^k}{k!} \leq \sum_{k \geq 2} \frac{\mu^k C_0^k k^k}{k!}$$

et comme $k! \geq \sqrt{2\pi k}(k/e)^k$, on a pour tout $\mu > 0$ tel que $e\mu C_0 < 1$,

$$\mathbb{E} \exp(\mu Z) \leq 1 + \sum_{k \geq 2} (e\mu C_0)^k = 1 + \frac{(e\mu C_0)^2}{1 - (e\mu C_0)} \leq \exp\left(\frac{(e\mu C_0)^2}{1 - (e\mu C_0)}\right)$$

car $1 + t \leq \exp(t)$ pour tout $t \in \mathbb{R}$.

On aura ainsi dans (3.4),

$$\begin{aligned} \mathbb{P}\left[\frac{1}{m} \sum_{i=1}^m Z_i \geq t\right] &\leq \inf_{0 < e\lambda C_0 < m} \left[e^{-\lambda t} \exp\left(\frac{m[e(\lambda/m)C_0]^2}{1 - [e(\lambda/m)C_0]}\right) \right] \\ &= \exp\left[\inf_{0 < e\lambda C_0 < m} \left(-\lambda t + \frac{(e\lambda C_0)^2}{m - (e\lambda C_0)}\right)\right] \leq \exp\left[\inf_{0 < e\lambda C_0 < m/2} \left(-\lambda t + \frac{2(e\lambda C_0)^2}{m}\right)\right] \end{aligned}$$

Maintenant, si $t \leq 2eC_0$, on prend $\lambda = mt/(2eC_0^2) \leq m/(2eC_0)$, on obtient alors

$$\mathbb{P}\left[\frac{1}{m} \sum_{i=1}^m Z_i \geq t\right] \leq \exp\left(\frac{-mt^2}{4e^2 C_0^2}\right).$$

Quand $t > 2eC_0$, on aura pour $\lambda = m/(2eC_0)$,

$$\mathbb{P}\left[\frac{1}{m} \sum_{i=1}^m Z_i \geq t\right] \leq \exp\left(\frac{-mt}{4eC_0}\right).$$

■

3.2 RIP est satisfaite par les matrices Gaussiennes

La section précédente va nous fournir les outils probabilistes pour démontrer qu'avec grande probabilité la condition $RIP(s)$ est satisfaite par les matrices aléatoires Gaussiennes pour le nombre minimal de lignes en $m \sim s \log(eN/s)$.

Théorème 3.6. *Soit $G \in \mathbb{R}^{m \times N}$ une matrice aléatoire Gaussienne standard. Il existe deux constantes absolues c_0 et c_1 telles que : avec probabilité au moins $1 - 2\exp(-c_0 m)$, pour tout $x \in \Sigma_s$,*

$$\frac{1}{2} \|x\|_2 \leq \frac{\|Gx\|_2}{\sqrt{m}} \leq \frac{3}{2} \|x\|_2$$

quand $m \geq c_1 s \log(eN/s)$.

On va en fait montrer le résultat dans un cadre plus général. On définit

$$\Gamma = \frac{1}{\sqrt{m}} A \text{ où } A = \begin{pmatrix} X_1^\top \\ \vdots \\ X_m^\top \end{pmatrix}$$

et X_1, \dots, X_m sont m variables aléatoires i.i.d. distribuées comme X qui est :

1. *isotropique* : $\mathbb{E}\langle X, x \rangle^2 = \|x\|_2^2$ pour tout $x \in \mathbb{R}^N$
2. *sous-gaussienne* : $\|\langle X, x \rangle\|_{L_p} \leq C_0 \|x\|_2 \sqrt{p}$ pour tout $x \in \mathbb{R}^N$ et tout $p \geq 1$.

Nous montrons dans ce qui suit que Γ vérifie $RIP(s)$ avec grande probabilité ; càd, pour cette normalisation, pour tout $x \in \Sigma_s$, $(1/2) \|x\|_2 \leq \|\Gamma x\|_2 \leq (3/2) \|x\|_2$.

On aura bien le résultat pour les matrices Gaussiennes vu que si \mathbf{g} est un vecteur aléatoire gaussien de \mathbb{R}^N standard alors, il est

1. isotropique : soit $x \in \mathbb{R}^N$,

$$\mathbb{E}\langle \mathbf{g}, x \rangle^2 = \mathbb{E}\left(\sum_{j=1}^N g_j x_j\right)^2 = \sum_{j=1}^N x_j^2 = \|x\|_2^2$$

2. sous-gaussien : pour tout $x \in \mathbb{R}^N$, $\langle \mathbf{g}, x \rangle$ a pour loi $\mathcal{N}(0, \|x\|_2^2)$ donc $\|\langle \mathbf{g}, x \rangle\|_{L_p} = \|x\|_2 \|g\|_{L_p} \leq C_0 \|x\|_2 \sqrt{p}$ pour tout $p \geq 1$.

Avant de passer à la preuve de RIP pour les matrices sous-gaussiennes, on rappelle deux résultats. Le premier étant un argument volumique portant sur la complexité des ensembles convexes.

Lemme 3.7. Soit s un entier. On note par B_2^s la boule euclidienne unité de \mathbb{R}^s . Soit $0 < \epsilon < 1$. Il existe un $\Lambda_\epsilon \subset B_2^s$ tel que :

1. pour tout $x \in B_2^s$ il existe $y \in \Lambda_\epsilon$ tel que $\|x - y\|_2 \leq \epsilon$,
2. $|\Lambda_\epsilon| \leq (5/\epsilon)^s$.

Démonstration. Lemme 3.7 se démontre par un argument volumique (cf. feuille d'exercices). ■

Lemme 3.8. Soit $B \in \mathbb{R}^{s \times s}$ une matrice symétrique et $0 < \epsilon < 1/2$. Soit Λ_ϵ un ϵ -net de S_2^{s-1} (la sphère euclidienne unité de \mathbb{R}^s). Alors, on a

$$\|B\| = \sup_{x \in S_2^{s-1}} |\langle Bx, x \rangle| \leq (1 - 2\epsilon)^{-1} \max_{y \in \Lambda_\epsilon} |\langle By, y \rangle|$$

où on rappelle que $\|B\| = \sup_{\|x\|_2=1} \|Bx\|_2$ est la norme d'opérateur de $B : \ell_2^s \mapsto \ell_2^s$.

Démonstration. Comme B est symétrique, il existe une matrice orthogonale $P \in \mathcal{O}(s)$ et une matrice diagonale $D = \text{diag}(\lambda_1, \dots, \lambda_s)$ telles que $B = PDP^\top$. On a

$$\begin{aligned} \|B\| &= \sup_{\|x\|_2=1} \|Bx\|_2 = \sup_{\|x\|_2=1} \|Dx\|_2 = \sup_{\|x\|_2=1} \sqrt{\sum_{j=1}^s \lambda_j^2 x_j^2} \\ &= \sup_{\|y\|_1=1} \sqrt{\sum_{j=1}^s \lambda_j^2 y_j} = \sqrt{\max(\lambda_1^2, \dots, \lambda_s^2)} = \max(|\lambda_1|, \dots, |\lambda_s|) \end{aligned}$$

et d'un autre côté,

$$\begin{aligned} \sup_{x \in S_2^{s-1}} |\langle Bx, x \rangle| &= \sup_{\|x\|_2=1} |\langle DP^\top x, P^\top x \rangle| = \sup_{\|x\|_2=1} |\langle Dx, x \rangle| \\ &= \sup_{\|x\|_2=1} \left| \sum_{j=1}^s \lambda_j x_j^2 \right| = \sup_{\|y\|_1=1} \left| \sum_{j=1}^s \lambda_j y_j \right| = \max(|\lambda_1|, \dots, |\lambda_s|). \end{aligned}$$

On a donc bien $\|B\| = \sup_{x \in S_2^{s-1}} |\langle Bx, x \rangle|$.

Pour l'inégalité de droite, on prend $x \in S_2^{s-1}$ et $y \in \Lambda_\epsilon$ tel que $\|x - y\|_2 \leq \epsilon$. On a

$$|\langle Bx, x \rangle| = |\langle By, y \rangle + \langle Bx, x - y \rangle + \langle B(x - y), y \rangle| \leq |\langle By, y \rangle| + 2\|x - y\|_2 \|B\|.$$

Comme $\|x - y\|_2 \leq \epsilon$, le résultat suit. ■

Démonstration du Théorème 3.6. On rappelle que Σ_s est l'ensemble de tous les vecteurs s -sparse de \mathbb{R}^N et S_2^{N-1} est la sphère euclidienne unité de ℓ_2^N . Pour montrer que Γ satisfait $RIP(s)$, il suffit de montrer que

$$\sup_{x \in \Sigma_s \cap S_2^{N-1}} \left| \|\Gamma x\|_2^2 - 1 \right| \leq \frac{1}{2}. \quad (3.5)$$

On peut écrire $\Sigma_s \cap S_2^{N-1} = \cup_{|J|=s} S^J$ où S^J est l'ensemble de tous les vecteurs de S_2^{N-1} dont le support est inclus dans J .

Selon Lemme 3.7, pour tout $J \subset \{1, \dots, N\}$ tel que $|J| = s$, il existe Λ_ϵ^J , un ϵ -net de S^J de cardinalité au plus $(5/\epsilon)^s$. Maintenant, si on applique Lemme 3.8 à la matrice symétrique $B = \Gamma_J^\top \Gamma_J - I_s$ où Γ_J est la matrice extraite de Γ dont les indices de colonnes sont dans J et I_s est l'identité de \mathbb{R}^s , on a

$$\begin{aligned} \sup_{x \in S^J} \left| \|\Gamma x\|_2^2 - 1 \right| &= \sup_{x \in S^J} \left| x_J^\top (\Gamma_J^\top \Gamma_J - I_s) x_J \right| = \sup_{x_J \in S_2^{s-1}} |\langle B x_J, x_J \rangle| \\ &\leq (1 - 2\epsilon)^{-1} \max_{y \in \Lambda_\epsilon^J} |\langle B y, y \rangle| = (1 - 2\epsilon)^{-1} \max_{x \in \Lambda_\epsilon^J} \left| \|\Gamma x\|_2^2 - 1 \right|. \end{aligned}$$

Ainsi, on a

$$\begin{aligned} \sup_{x \in \Sigma_s \cap S_2^{N-1}} \left| \|\Gamma x\|_2^2 - 1 \right| &= \max_{|J|=s} \sup_{x \in S^J} \left| \|\Gamma x\|_2^2 - 1 \right| \\ &\leq \max_{|J|=s} \left[(1 - 2\epsilon)^{-1} \max_{x \in \Lambda_\epsilon^J} \left| \|\Gamma x\|_2^2 - 1 \right| \right] \leq (1 - 2\epsilon)^{-1} \max_{x \in \Lambda_\epsilon} \left| \|\Gamma x\|_2^2 - 1 \right| \end{aligned} \quad (3.6)$$

où $\Lambda_\epsilon = \cup_{|J|=s} \Lambda_\epsilon^J$. En particulier, le cardinal de Λ_ϵ vérifie

$$|\Lambda_\epsilon| \leq \binom{N}{s} \left(\frac{5}{\epsilon} \right)^s \leq \left(\frac{5eN}{s\epsilon} \right)^s \quad (3.7)$$

où on a utilisé l'inégalité $\binom{N}{s} \leq (eN/s)^s$ (cf. feuille d'exercices).

On applique maintenant l'inégalité de Bernstein (cf. Proposition 3.5) pour tout les points $x \in \Lambda_\epsilon$: soit $x \in \Lambda_\epsilon$, avec probabilité au moins $1 - 2\exp(-c_0 m)$, $\left| \|\Gamma x\|_2^2 - 1 \right| \leq 1/8$. Ensuite, on utilise une borne de l'union sur Λ_ϵ : avec probabilité au moins $1 - 2|\Lambda_\epsilon| \exp(-c_0 m)$, pour tout $x \in \Lambda_\epsilon$, $\left| \|\Gamma x\|_2^2 - 1 \right| \leq 1/8$. Alors quand $m \gtrsim s \log(eN/(s\epsilon))$, on a $\max_{x \in \Lambda_\epsilon} \left| \|\Gamma x\|_2^2 - 1 \right| \leq 1/8$. Le résultat découle de (3.6) pour $\epsilon = 3/8$. ■

Conclusion : Avec grande probabilité, les matrices aléatoires Gaussiennes standard vérifient $RIP(s)$ pour le nombre optimal de mesures (lignes) en $m \sim s \log(eN/s)$. On peut donc utiliser les lignes de ces matrices (cà-d des vecteurs Gaussiens standard i.i.d. de \mathbb{R}^N) comme vecteurs de mesures.

Références

- [1] Emmanuel Candes and Terence Tao. Reflections on compressed sensing. *IEEE Information Theory Society Newsletter*, 2008.
- [2] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4) :1289–1306, 2006.