

# Parcimonie, Echantillonnage Compressif

B. Torrèsani

Master 2 Mathématiques et Applications,  
Probabilités-Statistique,  
Année 2010-11





# Table des matières

<b>0</b>	<b>Introduction</b>	<b>5</b>
0.1	Généralités . . . . .	5
0.2	Notations : normes et quasi-normes : . . . . .	5
0.3	Notations : Fourier . . . . .	5
<b>1</b>	<b>Représentation des signaux, échantillonnage, codage</b>	<b>7</b>
1.1	Echantillonnage, échantillonnage généralisé . . . . .	7
1.1.1	De Cauchy et Dirichlet à Shannon . . . . .	7
1.1.2	Sous-échantillonnage, repliement . . . . .	10
1.2	Représentations Hilbertiennes . . . . .	10
1.2.1	Rappels sur les bases Hilbertienne, et applications . . . . .	11
1.2.2	Signaux aléatoires, bases de Karhunen-Loève . . . . .	13
1.2.3	Retour sur MP3... et JPEG . . . . .	15
1.2.4	Un exemple étrange . . . . .	16
1.2.5	Approximation non-linéaire et parcimonie . . . . .	18
1.3	Exercices . . . . .	19
1.3.1	Maths . . . . .	19
1.3.2	Applications numériques . . . . .	20
<b>2</b>	<b>Principes d'incertitude et décomposition parcimonieuse</b>	<b>21</b>
2.1	Parcimonie, redondance . . . . .	21
2.1.1	Bases orthonormées, repères . . . . .	21
2.1.2	Pourquoi choisir un repère plutôt qu'une base? . . . . .	22
2.2	Algorithmes de décomposition parcimonieuse . . . . .	23
2.2.1	Méthode des repères : $p = 2$ . . . . .	23
2.2.2	Basis pursuit (BP) : $p = 1$ . . . . .	23
2.2.3	Matching pursuit (MP) . . . . .	24
2.2.4	Matching pursuit orthogonal . . . . .	26
2.2.5	LASSO, ou basis pursuit denoising . . . . .	27
2.3	Principes d'incertitude et identifiabilité . . . . .	28
2.3.1	Inégalités dans le domaine continu . . . . .	28
2.3.2	Inégalités en dimension finie . . . . .	30
2.4	Propriétés de noyau (NSP) . . . . .	36
2.4.1	NSP et identification . . . . .	36
2.4.2	Le cas $p = 1$ : lien avec la cohérence . . . . .	37
2.4.3	Stabilité : instance optimalité . . . . .	37
2.5	Exercices . . . . .	39
2.5.1	Maths . . . . .	39
2.5.2	Numérique . . . . .	39
<b>3</b>	<b>Echantillonnage Compressif</b>	<b>41</b>
3.1	Echantillonnage compressif . . . . .	41
3.2	Propriété d'isométrie restreinte . . . . .	42

*Table des matières*

3.3	Matrices aléatoires . . . . .	44
3.3.1	RIP et matrices aléatoires . . . . .	44
3.3.2	Matrices Gaussiennes . . . . .	47
3.3.3	Une classe de matrices aléatoires répondant au signalement... . . . . .	48
3.4	Une conclusion... temporaire . . . . .	49

# 0 Introduction

*Ce texte contient des notes d'un cours donné au master Mathématiques et Applications, Aix-Marseille, durant l'hiver 2011. Ces notes ont été rédigées au fur et à mesure de l'avancement du cours, et sont donc très loin d'être parfaites. Elles évolueront au fur et à mesure que l'auteur (et les lecteurs) y trouveront des défauts ou erreurs, et des améliorations possibles.*

## 0.1 Généralités

Par convention, on notera sous la forme  $f : t \rightarrow f(t)$  les fonctions d'une (ou plusieurs) variable(s) continue(s), et  $x : n \rightarrow x[n]$  les suites, en réservant les notations du type  $e_n$  pour l'indexation des éléments d'une base par exemple.

On précise ci-dessous quelques notations qui seront utiles pour la suite.

## 0.2 Notations : normes et quasi-normes :

1. Vecteurs, suites :

$$\|x\|_p = \left( \sum_n |x[n]|^p \right)^{1/p}$$

définit une norme pour  $p \geq 1$ , et une quasi-norme pour  $p < 1$ . Rappelons qu'une quasi-norme sur un espace linéaire  $E$  est une application  $v \in E \rightarrow \|v\| \in \mathbb{C}$  telle que  $\|\lambda v\| = |\lambda| \|v\|$ ,  $\|v\| = 0$  implique  $v = 0$ , et il existe une constante  $c$  telle que  $\|x + y\| \leq c(\|x\| + \|y\|)$  pour tous  $x, y \in E$ .

Il est possible de démontrer que la  $q$ -ième puissance de la quasi-norme  $\ell^q$  induit une métrique sur  $\mathbb{C}^N$ , définie par

$$d(x, y) = \|x - y\|_q^q.$$

2. Fonctions :

$$\|f\|_p = \left( \int |f(t)|^p dt \right)^{1/p}$$

## 0.3 Notations : Fourier

1. Transformation de Fourier finie (TFF) :  $x = \{x[0], \dots, x[N-1]\} \in \mathbb{C}^N \rightarrow \hat{x} \in \mathbb{C}^N$

$$\hat{x}[k] = \sum_{n=0}^{N-1} x[n] e^{-2i\pi kn/N}, \quad k = 0, \dots, N-1.$$

Transformation inverse :

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} \hat{x}[k] e^{2i\pi kn/N}.$$

Formule de Parseval :

$$\langle x, y \rangle = \frac{1}{N} \langle \hat{x}, \hat{y} \rangle, \quad \forall x, y \in \mathbb{C}^N$$

2. Transformation de Fourier discrète (TFD) :  $x \in \ell^2(\mathbb{Z}) \rightarrow \hat{x} \in L^2([-\pi, \pi])$

$$\hat{x}(\omega) = \sum_{n=-\infty}^{\infty} x[n]e^{-in\omega}, \quad \omega \in [-\pi, \pi].$$

Transformation inverse : coefficients de Fourier

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{x}(\omega)e^{in\omega} d\omega, \quad n \in \mathbb{Z}.$$

Formule de Parseval :

$$\langle x, y \rangle_{\ell^2(\mathbb{Z})} = \frac{1}{2\pi} \langle \hat{x}, \hat{y} \rangle_{L^2([-\pi, \pi])}, \quad \forall x, y \in \ell^2(\mathbb{Z}).$$

3. Transformation intégrale de Fourier :  $x \in L^2(\mathbb{R}) \rightarrow \hat{x} \in L^2(\mathbb{R})$

$$\hat{x}(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t} dt.$$

Transformation inverse :

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{x}(\omega)e^{i\omega t} d\omega,$$

où l'égalité est à prendre au sens de la convergence dans  $L^2(\mathbb{R})$ .

Formule de Plancherel :

$$\langle x, y \rangle_{L^2(\mathbb{R})} = \frac{1}{2\pi} \langle \hat{x}, \hat{y} \rangle_{L^2(\mathbb{R})}, \quad \forall x, y \in L^2(\mathbb{R}).$$

# 1 Représentation des signaux, échantillonnage, codage

Un signal peut être défini comme à peu près n'importe quel support susceptible de stocker ou transmettre des informations. Il peut s'agir d'un son ou d'une onde électromagnétique (les signaux analogiques les plus courants), ou d'une série de nombres stockés sur un disque ou dans la mémoire d'un ordinateur. Les deux questions qui se posent le plus généralement sont

- Comment représenter un signal, ou une classe de signaux, de façon la plus efficace possible ?
- Comment extraire de signaux les informations pertinentes ?

La technique de base (qui est par exemple à la base du format “standard” pour le codage des signaux audio sur les CD) porte le nom d'échantillonnage. On verra par la suite que se substituera à la notion d'échantillonnage celle d'acquisition (*sensing* en anglais), qui consiste à faire des mesures sur les signaux, mesures à partir desquelles on cherchera à reconstituer le signal le plus fidèlement possible.

## 1.1 Echantillonnage, échantillonnage généralisé

L'un des points essentiels du traitement numérique des signaux est celui de l'échantillonnage, que l'on peut formuler ainsi : étant donné un signal, quelles mesures doit-on faire pour le caractériser, au moins approximativement (avec une erreur contrôlée) ?

Si un signal est modélisé comme une fonction (éventuellement aléatoire) d'une variable réelle, disons  $f$ , l'approche la plus immédiate consiste à en faire des mesures “ponctuelles”, c'est à dire à considérer des valeurs  $f(n\delta)$ ,  $n \in \mathbb{Z}$ , pour un certain *pas d'échantillonnage*  $\delta$ . On appellera *fréquence d'échantillonnage* l'inverse  $\eta = 1/\delta$  du pas d'échantillonnage.

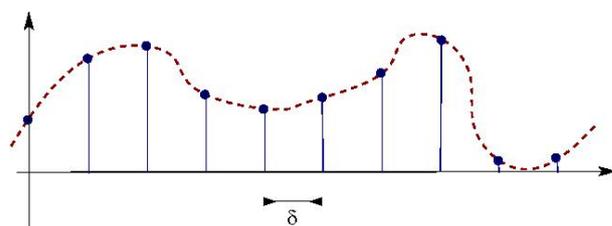


FIGURE 1.1: Echantillonnage

L'opération

$$f \in \mathcal{E} \mapsto \{f(n/\eta), n \in \Lambda\}$$

est appelée *échantillonnage*, ou *conversion analogique-numérique*. Ici,  $\mathcal{E}$  est un espace de fonctions à bien choisir pour que l'opération d'échantillonnage ait un sens, et  $\Lambda$  est un index adapté, à préciser également.

### 1.1.1 De Cauchy et Dirichlet à Shannon

Le cas le plus simple est celui des signaux analogiques de longueur finie. Prenons le modèle  $\mathcal{H} = L^2([0, T])$ . La théorie des séries de Fourier nous dit que l'application

$$x \in L^2([0, T]) \mapsto \{c_n(x), n \in \mathbb{Z}\}$$

est multiple d'une transformation unitaire : elle est bijective, et préserve (à une constante près) la norme et le produit Hermitien. Plus précisément,  $\forall x \in L^2([0, T])$ , en notant

$$c_n(x) = \frac{1}{T} \int_0^T x(t) e^{-2i\pi nt/T} dt$$

les coefficients de Fourier correspondants, on a la formule de Parseval

$$\frac{1}{T} \langle c_n(x), c_n(y) \rangle_{\ell^2(\mathbb{Z})} = \langle x, y \rangle_{L^2([0, T])}, \quad \forall x, y \in L^2([0, T]),$$

et la convergence de la série de Fourier au sens de  $L^2$  : en notant

$$x_N(t) = \sum_{n=-N}^N c_n(x) e^{2i\pi nt/T}$$

la série partielle, on a

$$\lim_{N \rightarrow \infty} \|x - x_N\| = 0, \quad \forall x \in L^2([0, T]).$$

On notera

$$\mathcal{P}_N([0, T]) = \{f \in L^2([0, T]), c_n(f) = 0 \forall |n| > N\}$$

l'espace des polynômes trigonométriques de degré inférieur ou égal à  $N$ . Cet espace se prête bien à l'échantillonnage

**Théorème 1 (Cauchy-Dirichlet)** Soit  $x \in \mathcal{P}_N([0, T])$ . Alors  $x$  est entièrement caractérisé par ses valeurs ponctuelles

$$x[n] = x\left(\frac{nT}{2N+1}\right), \quad n = 0, \dots, 2N$$

au sens où il existe une formule d'interpolation

$$x(t) = \frac{1}{2N+1} \sum_{n=0}^{2N} x[n] \mathcal{D}\left(\frac{2\pi}{T} \left(t - \frac{nT}{2N+1}\right)\right),$$

où  $\mathcal{D}$  est le noyau de Dirichlet

$$\mathcal{D}(t) = \frac{\sin[(N+1/2)t]}{\sin(t/2)}.$$

**Preuve :** Il faut tout d'abord remarquer qu'un polynôme trigonométrique est continu par construction, de sorte que les valeurs ponctuelles  $x[n]$  ont un sens. De plus, la série de Fourier est finie et donc converge point par point, on a donc

$$x[n] = \sum_{k=-N}^N c_k(x) e^{2i\pi kn/(2N+1)}.$$

Calculons maintenant

$$\sum_{n=0}^{2N} x[n] e^{-2i\pi \ell n/(2N+1)} = \sum_{n=0}^{2N} \sum_{k=-N}^N c_k(x) e^{2i\pi(k-\ell)n/(2N+1)} = (2N+1) c_\ell(x).$$

Donc

$$x(t) = \sum_{k=-N}^N c_k(x) e^{2i\pi kt/T} = \frac{1}{2N+1} \sum_{k=-N}^N \sum_{n=0}^{2N} x[n] e^{-2i\pi kn/(2N+1)} e^{2i\pi kt/T},$$

et il reste à évaluer la somme

$$\sum_{k=-N}^N e^{-ik\alpha} = e^{iN\alpha} \sum_{k=0}^{2N} e^{-ik\alpha} = e^{iN\alpha} \frac{1 - e^{-i(N+1/2)\alpha}}{1 - e^{-i\alpha}} = \frac{\sin((N+1/2)\alpha)}{\sin(\alpha/2)},$$

ce qui finit la démonstration, en prenant  $\alpha = 2\pi n/(2N+1) - 2\pi t/T$ . ♠

Le cadre traditionnel du théorème d'échantillonnage des fonctions à support non borné est l'espace des signaux à bande limitée, ou espace de Paley-Wiener

$$PW_{\omega_0} = \{x \in L^2(\mathbb{R}), \hat{x}(\omega) = 0 \text{ pour tout } \omega \notin [-\omega_0, \omega_0]\} \quad (1.1)$$

Il est facile de voir que  $PW_{\omega_0}$  est un espace de fonctions continues, de sorte que les valeurs ponctuelles des fonctions de  $PW_{\omega_0}$  sont bien définies. On peut alors introduire l'opérateur d'échantillonnage  $E$ , associé à la fréquence d'échantillonnage  $\eta$  : si  $x \in PW_{\omega_0}$ ,

$$(Ex)_n = x\left(\frac{n}{\eta}\right), \quad n \in \mathbb{Z}. \quad (1.2)$$

**Théorème 2 (Kotel'nikov, Raabe, Whittaker, Nyquist, Shannon, ...)** Soit  $x \in PW_{\omega_0}$ , et soit  $\eta > 0$  la fréquence d'échantillonnage. On considère la suite des échantillons  $x[n] = x(n/\eta)$ .

1. Si  $\pi\eta < \omega_0$ , la suite des échantillons  $x[n]$  ne permet pas de déterminer la fonction  $x$  sans hypothèse supplémentaire.
2. Si  $\pi\eta > \omega_0$ , alors il existe une infinité de formules d'interpolation de  $x$  à partir des échantillons. Quel que soit  $\varphi$  telle que  $\hat{\varphi} \in C^\infty$ ,  $\hat{\varphi}(\omega) = 0$  pour tout  $\omega \notin [-\pi\eta, \pi\eta]$  et  $\hat{\varphi}(\omega) = 1$  pour tout  $\omega \in [-\omega_0, \omega_0]$ . Alors on a

$$x(t) = \sum_{n=-\infty}^{\infty} \frac{1}{\eta} x\left(\frac{n}{\eta}\right) \varphi\left(t - \frac{n}{\eta}\right). \quad (1.3)$$

3. Si  $\pi\eta = \omega_0$ , alors il n'existe plus qu'une seule formule de reconstruction de  $x$  à partir des échantillons :

$$x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{\eta}\right) \frac{\sin(\pi\eta(t - n/\eta))}{\pi\eta(t - n/\eta)}. \quad (1.4)$$

La preuve du théorème utilise le résultat intermédiaire suivant, qui se démontre par un calcul explicite de coefficients de Fourier.

**Lemme 1** Soit

$$\Gamma(\omega) = \sum_k \hat{x}(\omega + 2k\pi\eta).$$

$\Gamma \in L^2([-\pi\eta, \pi\eta])$ , et ses coefficients de Fourier sont donnés par

$$c_n(\Gamma) = \frac{1}{\eta} x\left(\frac{-n}{\eta}\right).$$

Preuve du théorème (résumée) : La connaissance des échantillons  $x(n/\eta)$  est donc équivalente à la connaissance de la fonction  $\Gamma$ . Donc si il est possible de reconstruire  $\hat{x}$  à partir de  $\Gamma$ , il sera possible de reconstruire  $x$  à partir des échantillons. Supposons que  $\hat{x}$  ait son support inclus dans  $[-\omega_0, \omega_0]$ . Si  $\omega_0 > \pi\eta$ , la restriction de  $\Gamma$  à l'intervalle  $[-\omega_0, \omega_0]$  redonne  $\hat{x}$  ; on peut alors écrire

$$\hat{x}(\omega) = \Gamma(\omega)w(\omega),$$

pour toute fonction  $w$  telle que  $w(\omega) = 1$  sur  $[-\omega_0, \omega_0]$  et  $w(\omega) = 0$  à l'extérieur de  $[-\pi\eta, \pi\eta]$ . Dans ce cas,  $w$  peut être choisie  $C^\infty$ . Une transformation de Fourier inverse redonne  $x$  à partir de  $\Gamma$ , et donc des échantillons  $x(n/\eta)$ .

Si  $\omega_0 = \pi\eta$ , la situation est la même, sauf que  $w$  est nécessairement l'indicatrice de  $[-\omega_0, \omega_0]$ .

Si  $\pi\eta < \omega_0$ , l'argument ne fonctionne plus. ♠

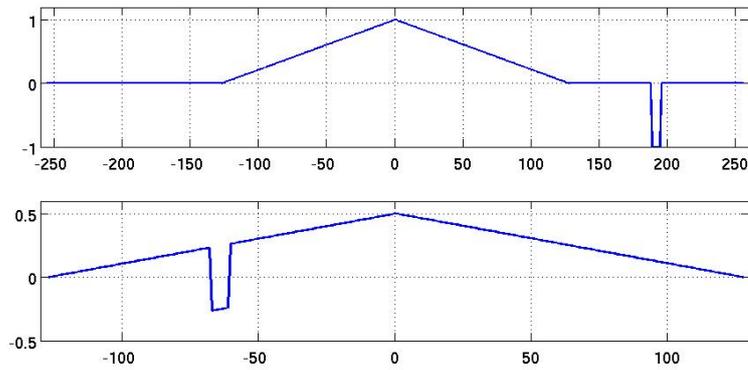


FIGURE 1.2: Illustration du phénomène de repliement de spectre

### 1.1.2 Sous-échantillonnage, repliement

On a jusqu'à présent traité uniquement le cas des signaux analogiques, modélisés comme des fonctions d'une variable réelle. Des problématiques similaires peuvent se poser dans le cas de signaux numériques.

Par exemple, étant donné un signal numérique  $x \in \ell^2(\mathbb{Z})$ , sa TFD est une fonction  $2\pi$ -périodique, caractérisée par sa restriction à l'intervalle  $[0, 2\pi]$ . Considérons un signal *sous-échantillonné*  $y$  défini par

$$y[n] = x[2n] .$$

Il est possible de démontrer que

$$\hat{y}(\omega) = \frac{1}{2} \left( \hat{x} \left( \frac{\omega}{2} \right) + \hat{x} \left( \frac{\omega}{2} + \pi \right) \right) ,$$

de sorte que cette opération de sous-échantillonnage mélange des morceaux de  $\hat{x}$ . Une illustration de ce phénomène, appelé **repliement de spectre** se trouve en FIGURE 1.2. On y voit en particulier que la partie "centrale" de la transformée de Fourier de  $x$  (figure du haut) a été correctement reproduite dans  $\hat{y}$  (figure du bas), à une dilatation près, mais que le "créneau" négatif de  $\hat{x}$  présent dans les valeurs positives de l'axe fréquentiel se retrouve "replié" dans les parties négatives de l'axe fréquentiel de  $\hat{y}$ , au milieu de la première partie. On comprend bien dans ces conditions, reconstruire  $\hat{x}$  (et donc  $x$ ) à partir de  $\hat{y}$  est impossible, sans hypothèse supplémentaire.

Similairement, on montre qu'un sous-échantillonnage d'un facteur  $p$  induit  $p-1$  termes de repliement.

**Remarque 1** Afin de pallier ce problème, le sous-échantillonnage doit toujours être précédé d'une opération permettant d'éliminer les termes de repliement. Dans l'exemple considéré, il s'agira de faire en sorte que  $\hat{x}(\omega/2 + \pi) = 0$  pour tout  $\omega \in [-\pi, \pi]$ . Ceci peut se faire en multipliant  $\hat{x}$  par une fonction  $m$  telle que  $m(\omega) = 0$  pour tout  $|\omega| \geq \pi/2$ . Cette opération, qui peut s'écrire

$$Tx[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} m(\omega) e^{in\omega} \hat{x}(\omega) d\omega$$

est appelée **filtrage passe-bas**. En pratique, il est préférable d'utiliser une fonction  $m$  (appelée fonction de transfert) plus régulière que l'indicatrice évoquée ci-dessus.

## 1.2 Représentations Hilbertiennes

La représentation d'une suite par ses valeurs peut être vue comme sa représentation par les coefficients de son développement sur la base canonique, aussi appelée base de Kronecker. Les techniques

modernes de codage des signaux utilisées par exemple dans les codeurs par transformation de type JPEG ou MP3 utilisent plutôt la représentation par les coefficients du développement du signal sur des bases plus adaptées.

### 1.2.1 Rappels sur les bases Hilbertienne, et applications

#### Bases Hilbertiennes

Dans cette section,  $\mathcal{H}$  est un espace de Hilbert séparable. Un système orthonormal dans  $\mathcal{H}$  est une famille  $\{e_\lambda \in \mathcal{H}, \lambda \in \Lambda\}$ , telle que

$$\langle e_\lambda, e_\mu \rangle = \begin{cases} 1 & \text{si } \lambda = \mu \\ 0 & \text{sinon} \end{cases} \quad (1.5)$$

Ici,  $\Lambda$  est un index au plus dénombrable (si  $\Lambda$  est infini, on prendra  $\Lambda = \mathbb{Z}$ ).

**Théorème 3 (Riesz-Fischer)** Soit  $\{e_\lambda, \lambda \in \Lambda\}$  un système orthonormal dans  $\mathcal{H}$ , et soit  $F \subset \mathcal{H}$  l'espace des combinaisons linéaires finies des  $e_\lambda, \lambda \in \Lambda$ .

1. Pour tout  $x \in \mathcal{H}$ , on a l'inégalité de Bessel :

$$\sum_{\lambda \in \Lambda} |\langle x, e_\lambda \rangle|^2 \leq \|x\|^2. \quad (1.6)$$

2. L'application  $x \in \mathcal{H} \rightarrow \{\langle x, e_\lambda \rangle, \lambda \in \Lambda\}$  est linéaire, continue de  $\mathcal{H}$  dans  $\ell^2(\Lambda)$  et sa restriction à  $\overline{F}$  est une isométrie de  $\overline{F}$  sur  $\ell^2(\Lambda)$ .

Les bases orthonormées (aussi appelées bases Hilbertiennes) peuvent être caractérisées de la façon suivante.

**Théorème 4** Soit  $\{e_\lambda, \lambda \in \Lambda\}$  un système orthonormal dans  $\mathcal{H}$ . Les assertions suivantes sont équivalentes :

1.  $\{e_\lambda, \lambda \in \Lambda\}$  est une base orthonormée de  $\mathcal{H}$ .
2. La famille  $\{e_\lambda, \lambda \in \Lambda\}$  est complète dans  $\mathcal{H}$ .
3. Pour tout  $x \in \mathcal{H}$ , on a la formule de Parseval :

$$\|x\|^2 = \sum_{\lambda \in \Lambda} |\langle x, e_\lambda \rangle|^2. \quad (1.7)$$

4. Pour tous  $x, y \in \mathcal{H}$ , on a

$$\langle x, y \rangle = \sum_{\lambda \in \Lambda} \langle x, e_\lambda \rangle \langle e_\lambda, y \rangle. \quad (1.8)$$

Ainsi, le choix d'une base orthonormée dans un espace de Hilbert établit de fait une transformation linéaire unitaire

$$v \in \mathcal{H} \mapsto \{\langle v, e_\lambda \rangle, \lambda \in \Lambda\} \in \ell^2(\Lambda)$$

appelée application coefficient.

Par ailleurs, étant donné un sous-ensemble  $\Lambda_0 = \{\lambda_1, \dots, \lambda_K\}$  de l'index  $\Lambda$ , en notant  $\mathcal{H}_0$  le sous-espace de  $\mathcal{H}$  engendré par  $\{e_\lambda, \lambda \in \Lambda_0\}$ , la base permet de donner une forme explicite au projecteur orthogonal  $\Pi_{\mathcal{H}_0} : \mathcal{H} \rightarrow \mathcal{H}_0$  :

$$\Pi_{\mathcal{H}_0} v = \sum_{k=1}^K \langle v, e_{\lambda_k} \rangle e_{\lambda_k}.$$

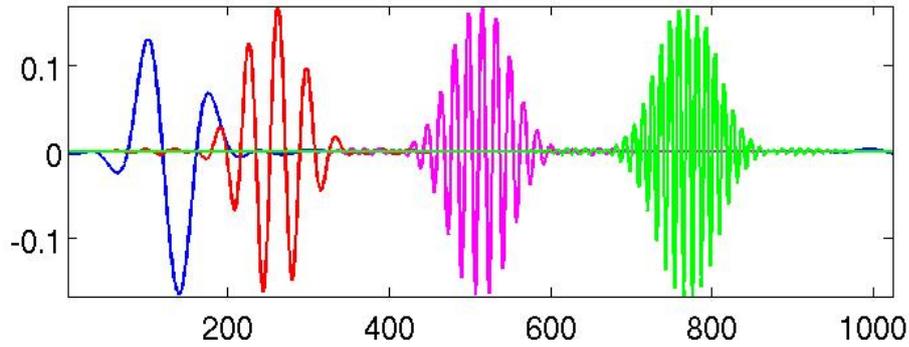


FIGURE 1.3: Quelques vecteurs de base MDCT

### Application au codage des signaux : codage par transformation

**Le codeur MP3 pour les sons** Le codeur MP3 utilise une base de Fourier locale [6], définie comme suit. Dans  $\ell^2(\mathbb{Z})$ , les vecteurs de base sont indexés par deux indices : un indice temporel  $k \in \mathbb{Z}$  et un indice fréquentiel  $\ell = 0, \dots, L - 1$ . Les vecteurs de base  $u^{(k,\ell)} \in \ell^2(\mathbb{Z})$  ont pour composantes

$$u^{(k,\ell)}[n] = \frac{1}{\sqrt{L}} w[n - kL] \cos\left(\pi \left(\ell + \frac{1}{2}\right) \frac{n - kL}{L}\right),$$

où  $w$  est une suite, appelée fenêtre, essentiellement localisée dans l'intervalle  $\{0, \dots, L - 1\}$ , et à laquelle on impose certaines conditions de symétrie (voir ci-dessous). Ces suites sont essentiellement des sinusoides, localisées à l'aide de copies translatées de la fenêtre  $w$ , comme on peut le voir en FIGURE 1.3.

Les conditions imposées à la fenêtre  $w$  sont les suivantes (voir par exemple [21, 23, 16] pour plus de détails) : on suppose qu'il existe un entier positif  $\eta < L/2$  tel que

- $w[n] = 1$  pour tout  $n = \eta, \dots, L - 1 - \eta$ ,
- $w[n] = 0$  pour tout  $n \neq -\eta, \dots, L - 1 + \eta$ ,
- $w[L + 1 - n] = w[n]$  pour tout  $n$ ,
- et une condition de normalisation pour assurer que la base est normée.

**Remarque 2** Dans  $\mathbb{R}^N$  ou  $\mathbb{C}^N$ , la construction est la même, à deux petites différences près :

- L'indice temporel varie maintenant dans un index fini :  $k = 0, \dots, K - 1$
- Les fenêtres pour  $k = 0$  et  $k = K - 1$  doivent être modifiées, pour tenir compte des bords.

**Remarque 3** Il existe des constructions similaires de bases MDCT dans  $L^2([a, b])$  et  $L^2(\mathbb{R})$ , les conditions imposées aux fenêtres sont similaires.

Etant donné un signal  $x \in \mathbb{R}^N$ , on effectue alors les opérations suivantes :

- Décomposition de  $x$  sur une base MDCT :

$$x = \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} \langle x, u^{(k,\ell)} \rangle u^{(k,\ell)}$$

- Projection sur un sous-espace de dimension inférieure, équivaut à un "forçage à zéro" des coefficients de la décomposition sur le sous-espace complémentaire. Ici le sous-espace retenu est donné par les vecteurs de base  $w^{k\ell}$  pour des indices fréquentiels  $\ell$  inférieurs à un certain  $L_0$  fixé :

$$x^{\text{app}} = \sum_{k=0}^{K-1} \sum_{\ell=0}^{L_0-1} \langle x, u^{(k,\ell)} \rangle u^{(k,\ell)},$$

l'erreur d'approximation introduite ici étant donnée par

$$\|x - x^{\text{app}}\|^2 = \sum_{k=0}^{K-1} \sum_{\ell=L_0}^{L-1} |\langle x, u^{(k,\ell)} \rangle|^2 ,$$

et peut être contrôlée si  $|\langle x, u^{(k,\ell)} \rangle|$  décroît suffisamment vite en fonction de  $\ell$ .

- Quantification et codage des coefficients retenus : les coefficients sélectionnés  $\langle x, u^{(k,\ell)} \rangle$  sont des nombres réels (ou complexes), mais doivent être stockés avec une précision finie (codés sur un nombre fini de bits). Ceci résulte en une approximation supplémentaire, appelés **quantification**, définie comme une fonction non-linéaire

$$Q : x \in \mathbb{R} \longmapsto Q(x) \in \{x_0, \dots, x_{2^R-1}\} ,$$

$R$  étant le nombre de bits utilisés pour encoder  $x$  (appelé **taux**).

### 1.2.2 Signaux aléatoires, bases de Karhunen-Loève

On s'intéresse ici à des signaux aléatoires, et à leur représentation sur une base.

**Définition 1** 1. Soit  $U = \mathbb{R}$  ou  $U = [0, T]$  ; soit  $(\mathcal{A}, \mathcal{F}, \mathbb{P})$  un espace probabilisé. Un signal aléatoire sur  $(\mathcal{A}, \mathcal{F}, \mathbb{P})$  indexé par  $U$  est une collection de variables aléatoires  $X_t : a \in \mathcal{A} \rightarrow X_t(a)$  indexée par  $t \in U$ .

2.  $X$  est du second ordre si pour tout  $t \in U$ ,  $X_t \in L^2(\Omega, \mathbb{P})$  (en d'autres termes,  $\mathbb{E}\{|X_t|^2\} < \infty$ ).
3.  $X$  est continu en moyenne quadratique si  $X$  est du second ordre et si la fonction  $(t, s) \rightarrow \mathbb{E}\{X_t \bar{X}_s\}$  est continue.

Etant donné un signal aléatoire du second ordre, il est clair que  $\mathbb{E}\{X_t\}$  existe pour tout  $t \in U$ . On supposera que le signal est centré

$$\mathbb{E}\{X_t\} = 0 , \quad \forall t \in U ,$$

et on s'intéresse à sa fonction d'autocovariance  $C_X$  définie par

$$C_X(t, s) = \mathbb{E}\{X_t \bar{X}_s\} .$$

**Lemme 2**  $C_X$  est une fonction semi-définie positive : pour tous  $t_1, \dots, t_n \in U$  et  $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ ,

$$\sum_{m,n=1}^n \alpha_m \bar{\alpha}_n C_X(t_m, t_n) \geq 0 .$$

$C_X$  est le noyau d'un opérateur  $\mathcal{C}_X$ , appelé opérateur de covariance, qui est donc semi-défini positif.

Supposons que  $X$ , défini sur  $[0, T]$  soit du second ordre, continu en moyenne quadratique, et soit  $\{\psi_n, n \in \mathbb{Z}\}$  une base orthonormée de  $L^2(\mathbb{R})$ . Il est possible de montrer que  $X$  admet une décomposition sur la base, de la forme

$$X = \sum_{n=-\infty}^{\infty} Z_n \psi_n ,$$

où les  $Z_n$  sont des variables aléatoires définies par

$$Z_n = \langle X, \psi_n \rangle = \int_0^T X_t \bar{\psi}_n(t) dt ,$$

et sont des variables aléatoires du second ordre ( $\mathbb{E}\{|Z_n|^2\} < \infty$ ). On montre alors, par un argument de type “Fubini”, que  $\mathbb{E}\{Z_n\} = 0$  pour tout  $n$ , et que

$$\mathbb{E}\{Z_n \bar{Z}_m\} = \int_0^T \int_0^T C_x(t, s) \bar{\psi}_n(t) \psi_m(s) dt ds = \langle \mathcal{C}_X \psi_m, \psi_n \rangle ,$$

ce qui conduit aux éléments de matrice de l’opérateur de covariance dans la base considérée. Dès lors, il est naturel de s’attendre à ce qu’une base diagonalisant l’opérateur  $\mathcal{C}_X$  puisse jouer un rôle intéressant dans cette histoire.

Considérons le cas où  $U = [0, T]$ , et supposons que  $C_X$  soit continue en  $t$  et  $s$ . Rappelons le théorème de Mercer :

**Théorème 5 (Mercer)** *Soit  $K$  un noyau Hermitien semi-défini positif,  $[a, b] \times [a, b] \rightarrow \mathbb{C}$ , continu. Soit  $T$  l’opérateur intégral sur  $L^2([a, b])$  défini par*

$$Tf(t) = \int_a^b K(t, s) f(s) ds .$$

*Alors il existe une base orthonormée  $\{e_k\}$  de fonctions propres de  $T$ , de valeurs propres  $\lambda_k$  réelles non-négatives*

$$T e_k = \lambda_k e_k ,$$

*et on a*

$$K(t, s) = \sum_{k=1}^{\infty} \lambda_k e_k(t) \bar{e}_k(s) ,$$

*où la somme converge absolument et uniformément. De plus, les fonctions propres  $e_k$  sont continues sur  $[a, b]$ .*

Revenons au signal aléatoire  $X$ .  $\mathcal{C}_X$  admet donc une base orthonormée de fonctions propres  $\varphi_i$ , telle que  $\mathcal{C}_X \varphi_i = \lambda_i \varphi_i$ . Cette base est appelée **base de Karhunen-Loève** (voir [20] pour plus de détails).

**Théorème 6** *Soit  $X$  un signal aléatoire du second ordre indexé par  $[0, T]$ , centré, et continu en moyenne quadratique. Les variables aléatoires  $Z_i$  définies par*

$$Z_i = \int_0^T X_t \bar{\varphi}_i(t) dt , \quad i = 1, 2, \dots$$

*sont du second ordre, centrées et décorréélées :*

$$\mathbb{E}\{Z_i\} = 0 , \quad \mathbb{E}\{Z_i \bar{Z}_j\} = \lambda_i \delta_{ij} .$$

*Par ailleurs, on a*

$$\mathbb{E}\{X_t \bar{Z}_i\} = \lambda_i \varphi_i(t) ,$$

*et on obtient le développement suivant de  $X_t$  :*

$$X_t = \sum_{i=1}^{\infty} Z_i \varphi_i(t) ,$$

*où la série est convergente dans  $L^2(\mathcal{A}, \mathbb{P})$  pour tout  $t \in [0, T]$ .*

Cette décomposition est parfois appelée *décomposition biorthogonale*, car elle fait intervenir une base orthonormée (au sens de  $L^2([0, T])$ )  $\{\varphi_i, i = 1, \dots, \infty\}$  et une famille orthogonale (au sens de  $L^2(\mathcal{A}, \mathbb{P})$ ) de variables aléatoires.

**Exemple 1** On dit qu'un signal aléatoire du second ordre indexé par  $U = [0, T]$  est cycliquement stationnaire si

$$\mathbb{E}\{X_t\} = \mathbb{E}\{X_0\} \forall t, \quad \mathbb{E}\{X_{(t+\delta) \bmod T} \overline{X_{(s+\delta) \bmod T}}\} = \mathbb{E}\{X_t \overline{X_s}\}, \quad \forall t, s, \delta \in [0, T].$$

On peut montrer que dans ce cas, la base de Karhunen-Loève associée est une base trigonométrique :

$$\varphi_k(t) = \frac{1}{\sqrt{T}} \exp\{2i\pi kt/T\}.$$

La base de Karhunen-Loève est importante en pratique car elle fournit des approximations optimales de plus basse dimension.

**Proposition 1 (Eckart-Young)** Soit  $\{X_t, t \in [0, T]\}$  un signal aléatoire du second ordre sur  $(\mathcal{A}, \mathcal{F}, \mathbb{P})$ , continu en moyenne quadratique. Soit  $\{\varphi_n, n = 1, 2, \dots\}$  la base de Karhunen-Loève associée. Soit  $\mathcal{E}_N$  un sous-espace de  $L^2([0, T])$  de dimension  $N$ , et soit  $\Pi_{\mathcal{E}_N}$  le projecteur orthogonal associé. Alors l'espace de dimension  $N$  qui minimise l'erreur de projection, en moyenne quadratique

$$\mathbb{E}\{\|X - \Pi_{\mathcal{E}_N} X\|^2\}$$

n'est autre que le sous-espace de  $L^2([0, T])$  engendré par les  $N$  premiers vecteurs de la base de Karhunen-Loève. L'approximation optimale (en ce sens) de  $X$  est donc

$$X^{(M)} = \sum_{m=1}^M \langle X, \varphi_m \rangle \varphi_m,$$

et l'erreur, mesurée en moyenne quadratique, vaut

$$\mathbb{E}\{\|X - X^{(M)}\|^2\} = \sum_{m=M}^{\infty} \lambda_m.$$

**Remarque 4** La construction de la base de Karhunen-Loève effectuée ci-dessus est bien plus simple dans le cas de signaux aléatoires de longueur finie. En effet, la fonction d'autocovariance définit une matrice Hermitienne, semi-définie positive, et dont la diagonalisation est donc élémentaire.

Le cas de signaux à temps continu et support non borné peut par contre être plus complexe, la diagonalisation pouvant conduire à des difficultés techniques demandant un traitement plus élaboré.

### 1.2.3 Retour sur MP3... et JPEG

Compte tenu de la Proposition 1, si on accepte l'idée de modéliser les signaux audio comme des réalisations d'un signal aléatoire, on s'attend à ce que l'erreur d'approximation due à la projection sur un sous-espace soit minimisée en utilisant la base de Karhunen-Loève associée à ce signal aléatoire.

L'idée sous-jacente aux codeurs audio de la famille MPEG est que les signaux audio ne peuvent pas être considérés stationnaires, mais que si on les analyse sur des durées suffisamment courtes (de l'ordre de 20 à 40 millisecondes), ils peuvent raisonnablement être approximatés par des signaux stationnaires. D'où l'idée d'utiliser, sur ces laps de temps, la base de Karhunen-Loève associée aux signaux stationnaires, qui n'est autre que la base trigonométrique. Ceci conduit donc au choix des bases trigonométriques locales.

Similairement, le codeur JPEG utilisé pour compresser des images fixes repose sur la constatation empirique que des imagerie de taille environ 8 pixels par 8 pixels peuvent raisonnablement être considérées comme stationnaires. Ceci conduit alors à représenter chaque imagerie par les coefficients de sa projection sur un sous-espace engendré par un sous-ensemble d'une base trigonométrique locale 2D.

Pour plus de détails sur le codage de signaux par transformation et les choix les plus appropriés de base orthonormée, on pourra par exemple consulter [16, 21].

### 1.2.4 Un exemple étrange

Comme on l'a vu, la base de Karhunen-Loève s'introduit assez naturellement lorsqu'il s'agit de coder des signaux aléatoires. Cependant, ça n'est pas obligatoirement le choix optimal, comme le montre l'exemple suivant dû à Y. Meyer [18].

Considérons le signal aléatoire  $X_t$  construit de la façon suivante : soit  $U_0 \sim \mathcal{U}([0, 1])$  une variable aléatoire, et soit

$$X_t = \begin{cases} t & \text{si } 0 \leq t < U_0 \\ t - 1 & \text{si } U_0 < t \leq 1 \end{cases}$$

On voit facilement que  $X_t$  est continu sauf en  $U_0$ . Un calcul explicite permet de montrer que

$$\mathbb{E}\{X_t\} = 0 \quad \forall t \in [0, 1], \quad \mathbb{E}\{X_t X_s\} = \min(s, t) - st.$$

Notons que ces deux moments sont identiques à ceux du pont Brownien (centré)

$$P_t = W_t - tW_1,$$

ces deux processus admettent donc la même base de Karhunen-Loève.

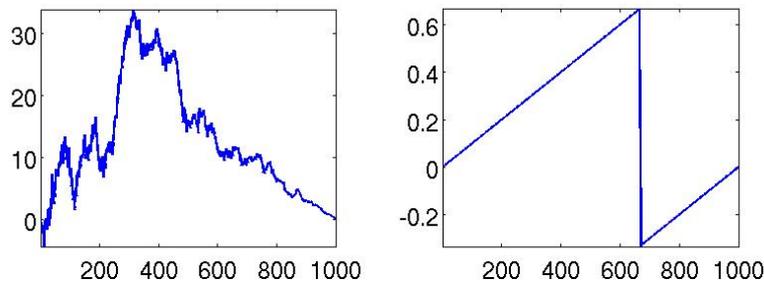


FIGURE 1.4: Deux signaux aléatoires possédant la même base de Karhunen-Loève : réalisation d'un pont Brownien (gauche) et d'une fonction "rampe" (droite).

### Base de Karhunen-Loève

**Lemme 3** *La base de Karhunen-Loève associée à l'autocovariance*

$$C_X(t, s) = \min(s, t) - st, \quad t, s \in [0, 1]$$

est donnée par

$$\varphi_n(t) = \sqrt{2} \sin(n\pi t),$$

et les valeurs propres correspondantes sont

$$\lambda_n = \frac{1}{(\pi n)^2}, \quad n \in \mathbb{Z}^+.$$

Supposons maintenant que l'on cherche à approximer le signal  $X$  le plus précisément possible par sa projection sur un sous-espace de dimension  $N$ . En notant  $E_N$  un tel sous-espace, et  $\Pi_N$  le projecteur orthogonal sur  $E_N$ , calculons

$$\mathbb{E}\{\|\Pi_N X\|^2\} = \sum_{m=1}^{\infty} \lambda_m \|\Pi_N \varphi_m\|^2.$$

Les  $\lambda_m$  étant positifs ou nuls, et ordonnés par ordre décroissant, on peut voir que le maximum de cette expression est obtenu pour  $\|\Pi_N \varphi_m\| = 1$  pour tout  $m = 1, \dots, N$ <sup>1</sup>. Ainsi, le sous-espace de dimension  $N$  de  $L^2(\mathbb{R})$  qui maximise  $\mathbb{E}\{\|\Pi_N X\|^2\}$  n'est autre que le sous-espace engendré par les  $N$  premiers éléments de la base de Karhunen-Loève. L'approximation optimale de  $X$  est donc de la forme

$$\Pi_N X = \sum_{m=1}^N Z_m \varphi_m ,$$

et l'erreur d'approximation peut ainsi être estimée par

$$\mathbb{E}\{\|X - \Pi_N X\|^2\} = \sum_{n=N+1}^{\infty} \lambda_n = \sum_{n=N+1}^{\infty} \frac{1}{\pi^2 n^2} \leq \frac{1}{\pi^2} \sum_{n=N+1}^{\infty} \int_{n-1}^n \frac{dx}{x^2} = \frac{1}{\pi^2} \int_N^{\infty} \frac{dx}{x^2} \leq \frac{1}{\pi^2 N} ,$$

ce qui donne une convergence très lente.

**Ondelettes** Considérons maintenant une base orthonormée d'ondelettes de  $L^2([0, 1])$ ,  $\{\psi_{jk}, j \in \mathbb{Z}^+, k = 0, \dots, 2^j - 1\}$ . Il s'agit essentiellement de copies translatées et dilatées  $2^{j/2}\psi(2^j t - k)$  d'une ondelette de base  $\psi$ , que l'on doit modifier sur les bords pour qu'elles restent supportées dans  $[0, 1]$ . Il est possible de construire de telles familles de sorte que les propriétés suivantes soient satisfaites : (voir [4, 16] pour plus de détails)

- La famille  $\{\psi_{jk}, j \in \mathbb{Z}^+, k = 0, \dots, 2^j - 1\}$  est une base orthonormée de  $L^2([0, 1])$ .
- Les ondelettes  $t \rightarrow \psi_{jk}(t)$  sont de classe  $C^q([0, 1])$ .
- Elles ont  $q$  moments nuls :  $\int_0^1 t^p \psi_{jk}(t) dt = 0$  pour tout entier  $p \leq q - 1$ , quels que soient  $j, k$ . Les ondelettes sont donc aveugles aux polynômes de degré inférieur ou égal à  $q$ .

Il est en fait possible de construire une base d'ondelettes telle que pour chaque valeur de la variable d'échelle  $j$ , seul un petit nombre de coefficients d'ondelettes soit non-nul. Par exemple, supposons que  $\psi$  possède  $q = 2$  moments nuls. Alors tous les coefficients d'ondelettes  $\langle X_t, \psi_{jk} \rangle$  tels que le support de  $\psi_{jk}$  ne contienne pas  $U_0$  sont nuls. Il est possible de démontrer que

- Le nombre de telles ondelettes à chaque échelle est borné par une (petite) constante  $K$ .
- A l'échelle  $j$ , ces coefficients sont bornés par  $|\langle X_t, \psi_{jk} \rangle| \leq C^{-j/2}$ .

Mettant ces estimations ensembles, on montre qu'en se limitant aux  $M = KJ$  coefficients correspondants aux plus petites valeurs de  $j$ , l'erreur d'approximation est de la forme

$$E \leq C' 2^{-\alpha J} ,$$

qui donne une décroissance bien plus rapide que celle obtenue dans le cas précédent. Il faut noter qu'elle est également indépendante de la réalisation de  $U_0$ .

**Comparaison** La comparaison faite ci-dessus peut être visualisée en analysant la décroissance des coefficients de Karhunen-Loève et d'ondelettes d'une réalisation de la fonction étudiée (voir **Figure. 1.5**, où on peut voir la très nette différence entre les deux vitesses de décroissance).

Ceci dit, elle n'est en fait pas vraiment *fair play* :

- Dans le premier cas, l'approximation sur la base de Karhunen-Loève est obtenue par projection sur un sous-espace de dimension  $N$ , toujours le même quelle que soit la réalisation de  $U_0$ .
- Dans le second cas, l'approximation est adaptative : le choix des vecteurs de base (ondelettes) retenus dépend de  $U_0$ , et peut donc s'adapter. Il ne s'agit pas d'une projection sur un sous-espace de  $L^2([0, 1])$ .

La différence entre ces deux types d'approche (projection et approximation par les vecteurs de base "les plus significatifs") tient à une notion un peu différente, appelée **parcimonie**.

---

1. C'est le théorème d'Eckart-Young, qui peut notamment se montrer en utilisant un développement sur une base orthonormée arbitraire

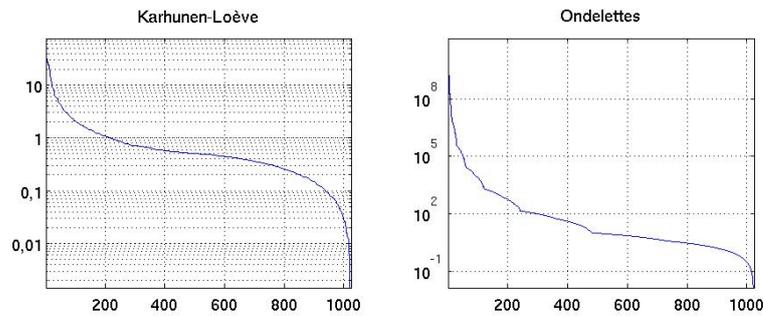


FIGURE 1.5: Décroissance des coefficients du développement de la fonction “rampe” sur la base de Karhunen-Loève (gauche) et une base d’ondelettes (droite).

### 1.2.5 Approximation non-linéaire et parcimonie

Etant donnée une base  $\mathcal{B} = \{\psi_k, k = 1, 2, \dots\}$  d’un espace de Hilbert  $\mathcal{H}$ , et un vecteur  $x \in \mathcal{H}$ , on distingue deux formes différentes d’approximation de  $x$  de “dimension”  $N$  :

- Approximation linéaire : projection de  $x$  sur le sous-espace engendré par  $N$  éléments fixés de  $\mathcal{B}$ , par exemple les premiers :

$$x_N^{(L)} = \sum_{n=1}^N \langle x, \psi_n \rangle \psi_n ,$$

et l’erreur d’approximation linéaire est donnée par

$$\sigma_N^{(L)} = \sum_{n \geq N+1} |\langle x, \psi_n \rangle|^2 .$$

- Approximation non-linéaire : notons  $n_1, n_2, \dots$  les indices tels que

$$|\langle x, \psi_{n_1} \rangle| \geq |\langle x, \psi_{n_2} \rangle| \geq |\langle x, \psi_{n_3} \rangle| \geq \dots ,$$

on définit alors

$$x_N^{(NL)} = \sum_{k=1}^N \langle x, \psi_{n_k} \rangle \psi_{n_k} ,$$

et l’erreur d’approximation non-linéaire est donnée par

$$\sigma_N^{(NL)} = \sum_{k \geq N+1} |\langle x, \psi_{n_k} \rangle|^2 .$$

On vérifie facilement (en conséquence de la formule de Parseval) que

$$\sigma_N^{(NL)} \leq \sigma_N^{(L)} ,$$

mais la question principale est la suivante : *existe-t-il des hypothèses assurant que la vitesse de convergence de l’approximation non-linéaire soit supérieure à la vitesse de convergence de l’approximation linéaire ?*

La réponse est positive, et correspond à la notion de *parcimonie*. En quelques mots, on dit d’un signal qu’il est représenté de façon parcimonieuse dans une base (ou un repère, voir plus loin) s’il peut être caractérisé, à une bonne approximation près, par une petite fraction des coefficients de son développement sur la base. On dit qu’il est compressible s’il existe une base où un repère dans lequel il est représenté de façon parcimonieuse.

Comme on va le voir, on associe souvent la propriété de parcimonie à l’appartenance de la suite des coefficients à un espace de type  $\ell^p$ .

## 1.3 Exercices

### 1.3.1 Maths

1. Démontrer le Théorème 2, dans le cas critique  $\omega_0 = \pi\eta$ . On pourra commencer par construire une transformation unitaire entre  $PW_{\omega_0}$  et  $L^2([-\omega_0, \omega_0])$ , puis utiliser la base trigonométrique de  $L^2([-\omega_0, \omega_0])$ .

2. Montrer que le système trigonométrique, constitué des  $N$  vecteurs  $e^k$ ,  $k = 0, \dots, N-1$  de composantes

$$e^k[n] = \frac{1}{\sqrt{N}} e^{2i\pi kn/N}$$

est une base orthonormée de  $\mathbb{C}^N$ . Expliciter les coefficients de la décomposition d'un signal  $x \in \mathbb{C}^N$  sur cette base.

3. Montrer que le système DCT-II, constitué des  $N$  vecteurs  $e^k$ ,  $k = 0, \dots, N-1$  définis par

$$e^k[n] = \sqrt{\frac{2}{N}} \cos\left(\pi\left(k + \frac{1}{2}\right)\frac{n}{N}\right), \quad n = 1, \dots, N-1, \quad e^k[0] = \frac{1}{\sqrt{N}}$$

est une base orthonormée de  $\mathbb{R}^N$ .

4. Même question pour le système DCT-IV, défini par

$$e^k[n] = \sqrt{\frac{2}{N}} \cos\left(\pi\frac{\left(k + \frac{1}{2}\right)\left(n + \frac{1}{2}\right)}{N}\right), \quad n = 0, \dots, N-1.$$

5. Soit  $x \in \mathbb{C}^N$ , soit  $p$  un diviseur de  $N$ . Soit  $y \in \mathbb{C}^{N/p}$  défini par

$$y[n] = x[pn], \quad n = 0, \dots, (N/p) - 1.$$

Exprimer  $\hat{y}$  en fonction de  $\hat{x}$ .

6. Démontrer la Proposition 1. On pourra considérer génériquement une famille orthonormée de  $N$  vecteurs et le sous-espace qu'elle engendre, et montrer que l'optimum est obtenu quand cette famille est constituée des  $N$  premiers vecteurs de la base de Karhunen-Loève (si  $\lambda_N$  est non-dégénérée). Cette démonstration peut être trouvée dans [16].

7. Montrer que la base de Karhunen-Loève du mouvement Brownien sur  $[0, 1]$ , d'autocovariance

$$C_W(t, s) = \min(t, s)$$

est donnée par

$$\varphi_k(t) = \sqrt{2} \sin(\pi(k + 1/2)t),$$

et expliciter les valeurs propres correspondantes.

8. Démontrer le Lemme 2.

9. Démontrer le Lemme 3

10. Un signal aléatoire sur  $[0, 1]$  est cycliquement stationnaire en moyenne d'ordre deux si sa fonction d'autocovariance  $C_X$  est telle que

$$C_X((t + \delta)[\text{mod } 1], (s + \delta)[\text{mod } 1]) = C_X(t, s), \quad \forall t, s, \delta \in [0, 1].$$

Calculer la base de Karhunen-Loève d'un signal cycliquement stationnaire en moyenne d'ordre deux.

### 1.3.2 Applications numériques

1. L'un des outils de base en analyse harmonique numérique est la transformation de Fourier rapide, algorithme qui permet de calculer la TFF d'un vecteur de longueur  $N$  en  $CN \log(N)$  opérations, où  $C$  est une constante. Sous *Matlab* cet algorithme est implémenté dans la fonction `fft`, et la fonction `ifft` effectue une TFF inverse (avec le même algorithme).
  - Tester la fonction `fft` sur des signaux sinusoidaux. Par exemple :

```
f0=64;  
X = sin(2*pi*f0*[1:512]);  
Xchap = fft(X);  
plot(abs(Xchap));  
plot(fftshift(abs(Xchap)));
```

La fonction `fftshift` permet de réordonner un vecteur (transformé de Fourier) de sorte que l'élément de fréquence nulle se trouve au milieu du vecteur.
  - On pourra faire plusieurs essais en variant la valeur de `f0`
2. Pour visualiser les effets du sous-échantillonnage, on pourra procéder comme ci-dessus, et ne garder qu'un échantillon de  $X$  sur deux :

```
Y = X(1:2:N);
```

et effectuer les mêmes opérations.  
Même questions pour un sous-échantillonnage d'un facteur  $p$ .
3. Implémenter l'exemple simple (et "étrange") ci dessus. Calculer ses coefficients de Fourier (en utilisant "fft").

## 2 Principes d'incertitude et décomposition parcimonieuse

### 2.1 Parcimonie, redondance

La parcimonie est devenue une notion centrale en traitement des signaux. Voici quelques exemples

- Les appareils photo prennent maintenant des photos en couleur (codées sur les 3 couleurs fondamentales R, G, B) d'une dizaine de millions de pixels. Chaque pixel étant codé sur 1 octet, cela représenterait environ 30 Mo par photo. Les fichiers, généralement codés au format JPEG, sont bien plus petits que cela.
- Sur un CD audio standard, une heure de musique (échantillonnée à 44100 valeurs par seconde et par canal, chaque valeur étant codée sur 2 octets) représente environ  $44100 \times 2 \times 3600 \times 2$ , soit environ 630 Mo. Le format MP3 produit des fichiers bien plus petits.
- En imagerie médicale, on est de nouveau confronté à de gigantesques masses de données, qui doivent être codées de façon efficace.
- ...

La clé qui permet un codage efficace des signaux est la notion de parcimonie. Supposons qu'un signal soit constitué d'un vecteur dans  $\mathbb{R}^N$ , et soit  $\mathcal{B} = \{\varphi_1, \dots, \varphi_N\}$  une base orthonormée de  $\mathbb{R}^N$ . On dira que le signal  $x \in \mathbb{R}^N$  est représenté de façon parcimonieuse dans la base  $\mathcal{B}$  si  $x$  peut être caractérisé par un petit nombre  $n \ll N$  de coefficients  $\langle x, \varphi_n \rangle$  de son développement sur la base  $\mathcal{B}$ .

#### 2.1.1 Bases orthonormées, repères

On utilisera souvent dans ce qui suit des bases orthonormées élémentaires telles que

- La base canonique, ou base de Kronecker  $\Delta = \{\delta_0, \dots, \delta_{N-1}\}$

$$\delta_n[n] = 1, \quad \delta_n[k] = 0 \quad \forall k \neq n$$

- La base de Fourier finie  $\mathcal{E} = \{\epsilon_0, \dots, \epsilon_{N-1}\}$  définie par

$$\epsilon_k[n] = \frac{1}{\sqrt{N}} e^{2i\pi kn/N}, \quad n = 0, \dots, n-1$$

**Remarque 5** Le cadre des bases orthonormées est souvent insuffisant quand il s'agit de représenter des signaux un tant soit plus complexes. Prenons l'exemple de  $x = \epsilon_{k_0} + \delta_{n_0}$ , pour  $k_0$  et  $n_0$  fixés.  $x$  n'est représenté parcimonieusement ni dans la base canonique ni dans la base de Fourier ; par contre il est représenté parcimonieusement dans un système constitué de l'union de ces deux bases.

Le problème qui se pose dans ce cas est que par rapport à ce système  $\mathcal{E} \cup \Delta$ ,  $x$  admet une infinité de développements possibles

$$x = \sum_{n=0}^{N-1} \alpha_n \delta_n + \sum_{k=0}^{N-1} \beta_k \epsilon_k .$$

Une heuristique de simplicité (ou de parcimonie) nous dit que la "bonne" représentation est celle donnée par  $\alpha_n = \delta_{n,n_0}$  et  $\beta_k = \delta_{k,k_0}$ , mais comment concrétiser cela ?

Au delà des unions de base, on utilisera plus génériquement la notion de repère (que l'on appelle parfois *structure oblique* ou *trame* :

**Définition 2** Un repère dans un espace de Hilbert séparable  $\mathcal{H}$  est une famille  $\{\varphi_\lambda, \lambda \in \Lambda\}$  d'éléments  $\varphi_\lambda \in \mathcal{H}$ , où  $\Lambda$  est un index donné, telle qu'il existe deux constantes  $0 < A \leq B < \infty$  vérifiant,  $\forall x \in \mathcal{H}$

$$A\|x\|^2 \leq \sum_{\lambda \in \Lambda} |\langle x, \varphi_\lambda \rangle|^2 \leq B\|x\|^2 .$$

**Remarque 6** On utilise parfois le nom de **dictionnaire** pour désigner un repère (bien que *stricto sensu*, un dictionnaire ne soit pas nécessairement complet, c'est à dire peut être tel que  $A = 0$ ).

On associe naturellement à un repère les opérateurs suivants :

– Opérateur d'analyse :

$$U : x \in \mathcal{H} \mapsto Ux = \{\langle x, \varphi_\lambda \rangle, \lambda \in \Lambda\} \in \ell^2(\Lambda)$$

– Opérateur de synthèse, défini comme l'adjoint  $V = U^*$  de  $U$  :

$$U^* : \alpha \in \ell^2(\Lambda) \mapsto U^*\alpha = \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda .$$

– Opérateur de repère :

$$\mathcal{R} = U^*U : x \in \mathcal{H} \mapsto \sum_{\lambda \in \Lambda} \langle x, \varphi_\lambda \rangle \varphi_\lambda .$$

On montre facilement le résultat suivant :

**Proposition 2** 1.  $\mathcal{R}$  est auto-adjoint et défini positif, son spectre est inclus dans l'intervalle  $[A, B]$ .

2. Il existe une famille  $\{\tilde{\varphi}_\lambda, \lambda \in \Lambda\}$  d'éléments de  $\mathcal{H}$ , constituant un repère de  $\mathcal{H}$  (appelé *dual canonique du repère initial*), telle que pour tout  $x \in \mathcal{H}$ , on ait

$$x = \sum_{\lambda \in \Lambda} \langle x, \varphi_\lambda \rangle \tilde{\varphi}_\lambda = \sum_{\lambda \in \Lambda} \langle x, \tilde{\varphi}_\lambda \rangle \varphi_\lambda . \quad (2.1)$$

On notera

$$\alpha_\lambda^0 = \langle x, \tilde{\varphi}_\lambda \rangle = \langle \mathcal{R}^{-1}x, \varphi_\lambda \rangle$$

les coefficients de la décomposition obtenus avec le repère dual canonique.

### 2.1.2 Pourquoi choisir un repère plutôt qu'une base ?

Considérons l'exemple simple suivant : dans  $\mathbb{C}^N$ ,

$$x = \delta_{n_0} + \epsilon_{k_0} ,$$

pour un certain  $n_0$  et un certain  $k_0$ . Clairement, chacune des deux “composantes” de  $x$  est représentée parcimonieusement (par un unique coefficient non nul) dans la base de Kronecker et la base de Fourier respectivement. Cependant, aucune des deux bases ne donne une représentation parcimonieuse de  $x$  (aucun coefficient n'est nul).

Par contre, il est facile de montrer que l'union de ces deux bases est un repère de  $\mathbb{C}^N$ , et que  $x$  peut être représenté par seulement 2 coefficients non-nuls dans ce repère. Par contre, la décomposition de  $x$  dans ce repère n'est pas unique, il va donc falloir travailler pour développer une méthode qui puisse sélectionner la “bonne” décomposition.

Plus généralement, l'heuristique est la suivante : plus un repère est “riche” (c'est à dire, redondant), plus un  $x$  donné peut être représenté parcimonieusement dans ce repère. Par contre, plus difficile sera la tâche de sélectionner la meilleure représentation.

## 2.2 Algorithmes de décomposition parcimonieuse

**Problème :** étant donné  $x \in \mathcal{H}$ , et un repère de  $\mathcal{H}$ , trouver la décomposition

$$x = \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda$$

telle que le nombre de coefficients  $\alpha_\lambda$  non nuls soit le plus faible possible. En introduisant la quasi-norme

$$\|\alpha\|_0 = \#\{\lambda \in \Lambda : \alpha_\lambda \neq 0\}$$

on est donc amené à considérer le problème

$$(P0) \quad \min \|\alpha\|_0 \quad \text{sous contrainte} \quad x = \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda .$$

Ce problème est souvent impossible à résoudre pratiquement, et on lui substitue le problème relaxé obtenu en substituant à la quasi-norme  $\ell^0$  une (quasi-)norme  $\ell^p$  plus facile à appréhender. Rappelons que pour  $p \geq 1$ , l'application  $x \rightarrow \|x\|_p = [\sum_\lambda |x_\lambda|^p]^{1/p}$  définit effectivement une norme, qui est donc convexe. Comme on va le voir, le cas  $p = 1$  a un statut particulier, qui conduit à la parcimonie.

### 2.2.1 Méthode des repères : $p = 2$

La méthode des repères est basée sur la norme  $\ell^2$ . le problème

$$(P2) \quad \min \|\alpha\|_2 \quad \text{sous contrainte} \quad x = \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda .$$

est un problème d'optimisation d'une forme quadratique sous contrainte linéaire, qui peut en fait se résoudre exactement.

**Proposition 3** *La solution du problème (P2) est donnée par*

$$\alpha_\lambda = \langle x, \tilde{\varphi}_\lambda \rangle = \alpha_\lambda^0$$

où  $\{\tilde{\varphi}_\lambda, \lambda \in \Lambda\}$  est le repère dual du repère  $\{\varphi_\lambda, \lambda \in \Lambda\}$ .

**Preuve :** On sait déjà d'après la Proposition 2 qu'il existe une façon canonique de décomposer  $x$  sur le repère, donnée par le repère dual canonique :  $\alpha_\lambda^0 = \langle x, \tilde{\varphi}_\lambda \rangle = \langle \mathcal{R}^{-1}x, \varphi_\lambda \rangle$ . Donc  $\alpha^0 \in U(\mathcal{H})$ ,  $U$  étant l'opérateur d'analyse. Soit maintenant  $\alpha$  tel que  $x = U^*\alpha$ .  $\alpha$  se décompose de manière unique sous la forme  $\alpha = \alpha^0 + \alpha^1$ , avec  $\alpha^1 \in U(\mathcal{H})^\perp$ , et on a

$$\|\alpha\|^2 = \|\alpha^0\|^2 + \|\alpha^1\|^2 ,$$

d'où le minimum est obtenu pour  $\alpha^1 = 0$ . ♠

Malheureusement, la méthode des repères ne permet pas d'obtenir de représentation parcimonieuse en général, comme on peut le voir sur l'exemple 2 discuté ci-dessous.

### 2.2.2 Basis pursuit (BP) : $p = 1$

*Basis pursuit* utilise la formulation

$$(P1) \quad \min \|\alpha\|_1 \quad \text{sous contrainte} \quad x = \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda .$$

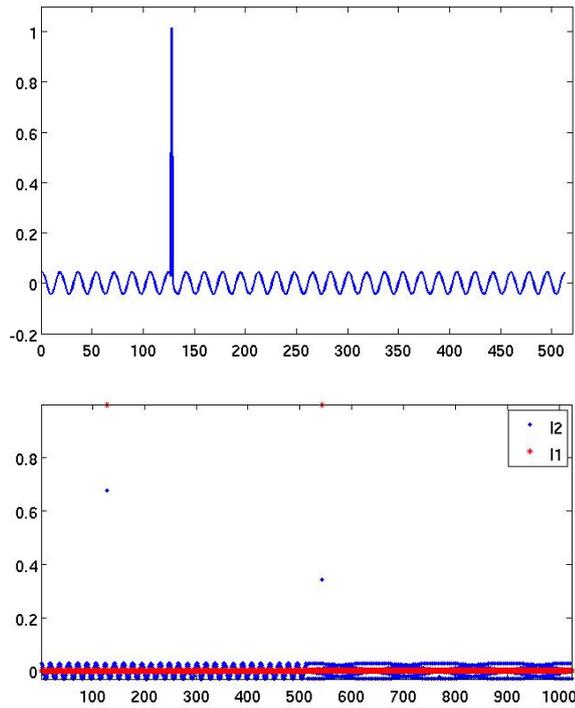


FIGURE 2.1: Décomposition de  $x = \delta_{n_0} + \epsilon_{k_0}$  sur le repère “Kronecker  $\cup$  Fourier”. Haut : le signal ; bas : coefficients de la décomposition obtenus par la méthode des repères (bleu) et Basis Pursuit (rouge).

où on a introduit la norme  $\ell^1$

$$\|\alpha\|_1 = \sum_{\lambda \in \Lambda} |\alpha_\lambda| .$$

Ce problème peut se formuler comme un problème de programmation linéaire, pour lequel il existe des algorithmes efficaces. On n’entrera pas ici dans le détail des algorithmes. On utilisera en pratique une implémentation dans le paquet logiciel `sparselab` développé sous *Matlab*, dans lequel BP est implémenté dans la fonction `SolveBP`.

**Exemple 2** On considère le cas simple  $x = \delta_{n_0} + \epsilon_{k_0}$ , pour lequel on recherche une décomposition parcimonieuse à l’aide de la méthode des repères et de Basis Pursuit. La Figure 2.1 ci dessous représente un signal, somme d’une sinusoïde  $\epsilon_{k_0}$  et d’un pulse  $\delta_{n_0}$ , ainsi que les coefficients de la décomposition sur le repère constitué de l’union de la base de Kronecker et de la base de Fourier. Comme on le voit, Basis Pursuit est capable d’identifier les deux coefficients significatifs, tous les autres étant nuls (points rouges sur la figure du bas), alors que la méthode des repères ne génère pas de parcimonie (points bleus), et les deux coefficients importants sont sous-estimés.

### 2.2.3 Matching pursuit (MP)

Le matching pursuit [17] est un algorithme itératif, de conception assez simple, qui recherche des approximations de plus en plus précises de la solution par optimisations unidimensionnelles. Supposons qu’on dispose d’un repère  $\{\varphi_\lambda, \lambda \in \Lambda\}$ , normalisé de sorte que  $\|\varphi_\lambda\| = 1$  pour tout  $\lambda$ . Le principe de l’algorithme est de rechercher à chaque itération  $k$  le vecteur  $\varphi_\lambda$  dont la contribution au résidu courant  $r^{(k-1)}$  est la plus grande, et soustraire cette contribution, générant un nouveau résidu  $r^{(k)}$ .

Dans certaines situations, cette stratégie n’est pas applicable, et on ne peut que rechercher un

vecteur  $\varphi_\lambda$  tel que

$$|\langle r^{(k)}, \varphi_\lambda \rangle| \geq \alpha |\langle r^{(k)}, \varphi_{\lambda'} \rangle|, \quad \forall \lambda',$$

où  $0 < \alpha \leq 1$  est un paramètre fixé.

- **Initialisation** :  $r^{(0)} = x$ ,  $x^{(0)} = 0$ .
- **Itération** : tant que le critère d'arrêt n'est pas atteint
  - Recherche de l'élément optimal : à l'étape  $k$ ,  $r^{(k)}$  est connu et on cherche

$$\lambda_k \in \left\{ \lambda \in \Lambda : |\langle r^{(k)}, \varphi_\lambda \rangle| \geq \alpha |\langle r^{(k)}, \varphi_{\lambda'} \rangle|, \quad \forall \lambda' \right\}$$

- Mise à jour de l'approximation au rang  $k$

$$x^{(k+1)} = x^{(k)} + \langle r^{(k)}, \varphi_{\lambda_k} \rangle \varphi_{\lambda_k}$$

et du résidu

$$r^{(k+1)} = x - x^{(k+1)} = r^{(k)} - \langle r^{(k)}, \varphi_{\lambda_k} \rangle \varphi_{\lambda_k}.$$

En sortie d'algorithme,  $x$  s'exprime sous la forme d'une série télescopique

$$x = \sum_{k=0}^K \gamma_{\lambda_k} \varphi_{\lambda_k} + r^{(K)},$$

où les coefficients  $\gamma_{\lambda_k}$  sont issus du calcul

$$\gamma_{\lambda_k} = \langle r^{(k)}, \varphi_{\lambda_k} \rangle.$$

Il est à noter qu'à chaque étape, on a

$$\|r^{(k+1)}\|^2 = \|r^{(k)}\|^2 - |\gamma_k|^2,$$

de sorte que la norme du résidu est décroissante. La convergence est donnée dans le cas général par le résultat suivant

**Théorème 7** Soit  $\mathcal{D} = \{\varphi_\lambda, \lambda \in \Lambda\}$  un repère dans l'espace de Hilbert  $\mathcal{H}$ . Alors

$$\lim_{n \rightarrow \infty} \|r^{(n)}\| = 0.$$

Si  $\mathcal{H}$  est de dimension finie, il est possible de préciser davantage la convergence, en utilisant la mesure suivante de corrélation :

$$I(\mathcal{D}) = \inf_{x \in \mathcal{H}} \sup_{\lambda \in \Lambda} \frac{|\langle x, \varphi_\lambda \rangle|}{\|x\|}.$$

$I(\mathcal{D})$  est en fait le cosinus de l'angle maximum d'un élément quelconque  $x \in \mathcal{H}$  avec un élément quelconque du repère.

**Lemme 4**

$$I(\mathcal{D}) > 0.$$

**Preuve** : Supposons que ce soit faux. Alors il est possible de trouver une suite de réels  $u_n \rightarrow 0$  et une suite de vecteurs  $x_n$  tels que  $\|x_n\| = 1$  et

$$\epsilon_n = \inf_{\varphi \in \mathcal{D}} |\langle x_n, \varphi \rangle| \leq u_n.$$

L'ensemble des vecteurs de norme 1 est une variété compacte dans  $\mathcal{H}$ . On peut donc extraire une sous-suite convergente  $x_{k_n} \rightarrow x$ , et on a donc

$$|\langle x, \varphi_\lambda \rangle| = \lim_{n \rightarrow \infty} |\langle x_{k_n}, \varphi_\lambda \rangle| \leq \lim_{n \rightarrow \infty} u_{k_n} = 0.$$

Comme le repère est complet, on a nécessairement  $x = 0$ , d'où contradiction. ♠

On en déduit le résultat suivant :

**Proposition 4** Avec les notations ci-dessous, la norme du résidu est bornée comme

$$\|r^{(k)}\| \leq \|x\| [1 - \alpha^2 I(\mathcal{D})^2]^{k/2} .$$

**Preuve :** A l'itération  $n$ , on choisit  $\lambda_n$  tel que

$$|\langle r^{(n)}, \varphi_{\lambda_n} \rangle| \geq \alpha \sup_{\lambda} |\langle r^{(n)}, \varphi_{\lambda} \rangle| \geq \alpha I(\mathcal{D}) \|r^{(n)}\| .$$

Donc

$$\|r^{(n+1)}\|^2 = \|r^{(n)}\|^2 - |\langle r^{(n)}, \varphi_{\lambda_n} \rangle|^2 \leq \|r^{(n)}\|^2 (1 - \alpha^2 I(\mathcal{D})^2) ,$$

ce qui prouve la proposition. ♠

Notons qu'en dimension finie, si  $\mathcal{D}$  est de taille finie, on peut prendre  $\alpha = 1$ .

Dans le paquet logiciel **SparseLab**, le matching poursuit est implémenté dans la fonction **SolveMP**.

### 2.2.4 Matching poursuit orthogonal

Le principe de base reste le même que celui du matching poursuit. La seule chose qui change est la mise à jour de l'approximation, et donc du résidu. Dans ce qui suit, on notera  $\Pi_E$  l'opérateur de projection orthogonale d'un espace de Hilbert sur un sous-espace  $E$ .

L'algorithme de Matching poursuit orthogonal fonctionne comme suit :

- **Initialisation** :  $r^{(0)} = x$ ,  $x^{(0)} = 0$ .
- **Itération** : tant que le critère d'arrêt n'est pas atteint
  - Recherche de l'élément optimal

$$\lambda_k = \operatorname{argmax}_{\lambda} |\langle r^{(k)}, \varphi_{\lambda} \rangle|$$

- Mise à jour de l'approximation au rang  $k$  : soit  $E_k$  le sous-espace engendré par les vecteurs sélectionnés aux  $k$  premières itérations. On pose

$$x^{(k+1)} = \Pi_{E_k} x$$

et le résidu est mis à jour par

$$r^{(k+1)} = x - x^{(k+1)}$$

**Remarque 7** 1. Il faut noter que contrairement à ce qui se passe dans le cas du matching poursuit, le résidu  $r^{(n)}$  à l'étape  $n$  appartient au sous-espace orthogonal à l'espace engendré par les  $n$  vecteurs sélectionnés. Ainsi, lorsqu'un vecteur du repère a été sélectionné, il ne le sera plus jamais. OMP effectue de fait une réduction de dimension.

2. La contrepartie est la nécessité de calculer à chaque étape une projection orthogonale sur un sous-espace de dimension de plus en plus grande, ce qui devient coûteux lorsque l'indice  $n$  augmente. En effet, on peut écrire

$$x^{(N+1)} = \sum_{n=0}^N \gamma_n \varphi_{\lambda_n} ,$$

où le vecteur  $\gamma = \{\gamma_0, \dots, \gamma_N\}$  des coefficients est donné par

$$\gamma = G^{-1} U x = (U U^*)^{-1} U x ,$$

$G = U U^* = \{\langle \varphi_{\lambda}, \varphi_{\lambda'} \rangle, \lambda, \lambda' \in \Lambda\}$  étant la **matrice de Gram** de la famille  $\{\varphi_{\lambda_0}, \dots, \varphi_{\lambda_N}\}$ , dont on peut montrer qu'elle est inversible.  $G$  est une matrice  $N \times N$ , de sorte que son inversion peut poser problème quand  $N$  est grand.

Un exemple de convergence de MP et OMP se trouve en **Figure 2.2**. Un signal parcimonieux a été généré à partir de 50 coefficients non-nuls dans un repère "Fourier+Kronecker" et les coefficients ont été identifiés par MP et OMP. Comme prévu, OMP a besoin d'exactly 50 itérations, alors que MP demande un peu plus.

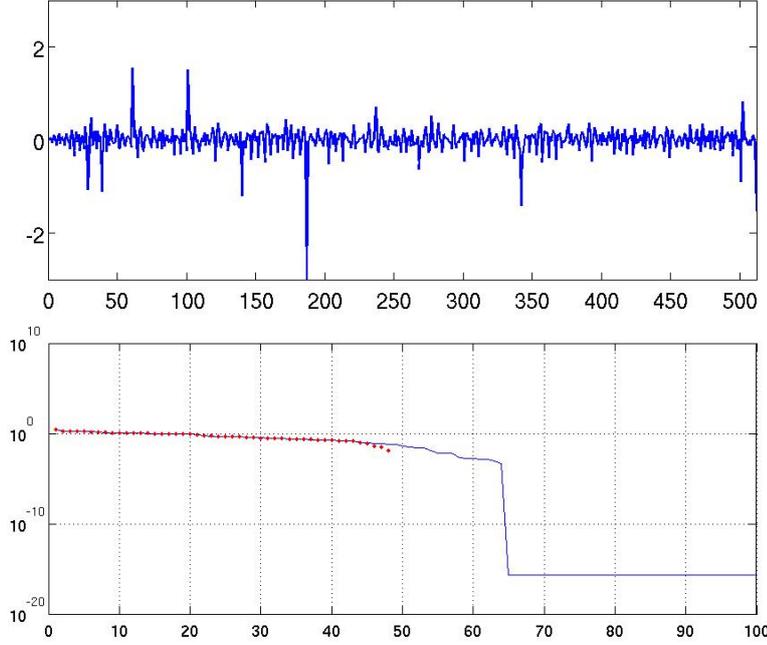


FIGURE 2.2: Exemple d'application de MP et OMP : signal parcimonieux (haut) et vitesses de convergence des deux algorithmes (bas) : MP (bleu, trait plein) et OMP (rouge, tirets).

### 2.2.5 LASSO, ou basis pursuit denoising

L'algorithme du LASSO, aussi appelé *basis pursuit denoising*, peut être vu comme une version relaxée du Basis Pursuit. L'idée est de rechercher une décomposition parcimonieuse d'un signal à une petite erreur près. Le problème de LASSO peut se formuler de trois façons équivalentes :

$$(P1a') \quad \min \|\alpha\|_1 \quad \text{sous contrainte} \quad \left\| x - \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda \right\|_2 \leq \epsilon$$

$$(P1b') \quad \min \left\| x - \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda \right\|_2 \quad \text{sous contrainte} \quad \|\alpha\|_1 \leq \delta$$

$$(P1c') \quad \min \left( \frac{1}{2} \left\| x - \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda \right\|_2^2 + \mu \|\alpha\|_1 \right)$$

pour des valeurs compatibles de  $\epsilon$ ,  $\delta$  et  $\lambda$ .

**Lemme 5** Lorsque  $\{\varphi_\lambda, \lambda \in \Lambda\}$  est une base orthonormée de l'espace considéré, le problème (P1') possède une solution explicite :

$$\alpha_\lambda = \mathbb{S}_\mu(\langle x, \varphi_\lambda \rangle) = \begin{cases} e^{i \arg(\langle x, \varphi_\lambda \rangle)} [|\langle x, \varphi_\lambda \rangle| - \mu] & \text{si } |\langle x, \varphi_\lambda \rangle| \geq \mu \\ 0 & \text{sinon} \end{cases}$$

La fonction  $\mathbb{S}_\lambda$  est appelée *seuillage doux*, par opposition au seuillage dur.

$$\mathbb{D}_\mu(z) = \begin{cases} z & \text{si } |z| \geq \mu \\ 0 & \text{sinon} \end{cases}$$

**Remarque 8** Lorsque le repère  $\{\varphi_\lambda, \lambda \in \Lambda\}$  n'est plus une base orthonormée mais est redondant, c'est à dire quand  $\text{Ker}(U^*) \neq \{0\}$ , la solution du problème n'est plus donnée par un simple seuillage doux. Cependant, le problème peut être résolu numériquement par seuillage doux itératif [7], une méthode qui tombe dans la famille des **itérations de Landweber**. Partant d'une initialisation  $\alpha^{(0)} = \{\alpha_\lambda^{(0)}, \lambda \in \Lambda\}$ , la mise à jour s'effectue via l'expression

$$\alpha_\lambda^{(n+1)} = \mathbb{S}_\mu \left( \alpha_\lambda^{(n)} - \epsilon U \left( U^* \alpha^{(n)} - x \right) \right),$$

dont on peut démontrer la convergence pour des valeurs de  $\epsilon$  suffisamment petites.

## 2.3 Principes d'incertitude et identifiabilité

Un principe d'incertitude est un résultat qui traduit généralement l'impossibilité pour un objet (vecteur, suite, fonction,...) d'être représentées de façon parcimonieuses simultanément dans des représentations différentes. Ceci se traduit généralement par une inégalité qui limite un produit de deux mesures de parcimonie, ou concentration.

Il s'avère que les principes d'incertitude jouent un rôle fondamental dans les problèmes d'approximation parcimonieuse.

### 2.3.1 Inégalités dans le domaine continu

#### L'inégalité d'Heisenberg

La première formulation connue du principe d'incertitude est associée à l'inégalité d'Heisenberg, qui utilise l'intégrale de Fourier : si  $f \in L^2(\mathbb{R})$ ,

$$\hat{f}(\nu) = \int_{-\infty}^{\infty} f(t) e^{-2i\pi\nu t} dt.$$

**Théorème 8** Soit  $f \in L^2(\mathbb{R})$ ,  $f \neq 0$ . Soient  $e_f$  et  $v_f$  la valeur moyenne et la variance de  $f$ , définies par

$$e_f = \frac{1}{\|f\|^2} \int_{-\infty}^{\infty} t |f(t)|^2 dt, \quad v_f = \frac{1}{\|f\|^2} \int_{-\infty}^{\infty} (t - e_f)^2 |f(t)|^2 dt.$$

Alors

$$v_f \cdot v_{\hat{f}} \geq \frac{1}{16\pi^2},$$

avec égalité si et seulement si  $f$  est une fonction Gaussienne, de la forme

$$f(t) = a e^{-bt^2} e^{ict+dt^2},$$

pour des constantes  $a \in \mathbb{C}$ ,  $b \in \mathbb{R}^+$  et  $c, d \in \mathbb{R}$ .

#### L'inégalité de Donoho-Stark

L'inégalité d'Heisenberg utilise une notion de variance pour mesurer la concentration de  $f$  et  $\hat{f}$ . L'inégalité de Donoho-Stark [10] utilise des mesures de concentration différentes, définies comme suit : soient  $f \in L^2(\mathbb{R})$ , et  $\epsilon > 0$ ;  $f$  est  $\epsilon$ -concentrée dans l'ensemble mesurable  $E \subset \mathbb{R}$  si

$$\left( \int_{\mathbb{R} \setminus E} |f(t)|^2 dt \right)^{1/2} \leq \epsilon \|f\|.$$

**Théorème 9** Supposons que  $f \in L^2(\mathbb{R})$  soit  $\epsilon_T$ -concentrée dans  $T$ , et que  $\hat{f}$  soit  $\epsilon_F$ -concentrée dans  $F$ ; alors

$$|T| \cdot |F| \geq (1 - (\epsilon_T + \epsilon_F))^2 ,$$

où on a noté  $|T|$  la mesure de Lebesgue de l'ensemble  $T$ .

**Preuve :** Commençons par introduire les opérateurs (linéaires) de localisation temporelle et fréquentielle. Soient  $T, F$  deux sous-ensembles mesurables de  $\mathbb{R}$ . Sans perte de généralité, on peut supposer que  $T$  et  $F$  sont de mesure finie (Dans le cas contraire le résultat est trivialement vérifié). On note

$$P_T f(t) = \begin{cases} f(t) & \text{si } t \in T \\ 0 & \text{sinon.} \end{cases} \quad P_F f(t) = \int_F \hat{f}(\nu) e^{2i\pi\nu t} d\nu , \quad f \in L^2(\mathbb{R}) .$$

$P_T$  et  $P_F$  sont deux projecteurs orthogonaux; donc  $f$  est  $\epsilon_T$ -concentrée dans  $T$  si et seulement si  $\|f - P_T f\| \leq \epsilon_T$ , et  $\hat{f}$  est  $\epsilon_F$ -concentrée dans  $F$  si et seulement si  $\|f - P_F f\| \leq \epsilon_F$ .

On a de plus

$$\|f - P_F P_T f\| = \|f - P_F f + P_F f - P_F P_T f\| \leq \|f - P_F f\| + \|f - P_T f\| ,$$

car  $\|P_F\| \leq 1$ . De plus

$$\|f - P_F P_T f\| \geq \|f\| - \|P_F P_T f\| ,$$

d'où on déduit

$$\|P_F P_T f\| \geq 1 - (\epsilon_F + \epsilon_T) .$$

$P_F P_T$  est la composition de deux projecteurs orthogonaux sur des domaines bornés, qui sont des opérateurs de Hilbert-Schmidt, c'est aussi un opérateur de Hilbert-Schmidt. Il existe donc d'après le théorème des noyaux une fonction  $\kappa \in L^2(\mathbb{R}^2)$  telle que  $P_F P_T f(t) = \int_{-\infty}^{\infty} \kappa(t, s) f(s) ds$ . Un calcul explicite donne

$$P_F P_T f(t) = \int_F e^{2i\pi\nu t} \int_T f(s) e^{-2i\pi\nu s} ds d\nu = \int_T \int_F e^{2i\pi\nu(s-t)} d\nu f(s) ds ,$$

d'où

$$\kappa(t, s) = \begin{cases} \int_F e^{2i\pi\nu(s-t)} d\nu & \text{si } s \neq T \\ 0 & \text{sinon} \end{cases}$$

Calculons maintenant

$$\int_{\mathbb{R} \times T} |\kappa(t, s)|^2 dt ds = \int_{\mathbb{R} \times T} \left( \int_F e^{2i\pi\nu(t-s)} d\nu \right) \left( \int_F e^{-2i\pi\nu'(t-s)} d\nu' \right) dt ds .$$

En notant  $u \rightarrow I_U(u)$  l'indicatrice de l'ensemble mesurable  $U$ , on a

$$\begin{aligned} \int_{-\infty}^{\infty} |\kappa(t, s)|^2 dt &= I_T(s) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_F(\nu) I_F(\nu') e^{2i\pi(\nu-\nu')(t-s)} d\nu d\nu' dt \\ &= I_T(s) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_F(\nu) I_F(\nu') e^{2i\pi(\nu-\nu')u} d\nu d\nu' du \\ &= I_T(s) \int_{-\infty}^{\infty} I_F(\nu) d\nu \\ &= |F| I_T(s) , \end{aligned}$$

d'où la norme Hilbert-Schmidt de  $P_F P_T$  vaut

$$\|P_F P_T\|_{HS}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\kappa(t, s)|^2 dt ds = |F| \cdot |T| .$$

On sait par ailleurs que

$$\|P_F P_T\| \leq \|P_F P_T\|_{HS} ,$$

on obtient donc

$$|F| \cdot |T| \leq \|P_F P_T\|^2 \leq [1 - (\epsilon_F + \epsilon_T)]^2 ,$$

ce qui est le résultat désiré. ♠

**Application : reconstruction d'un signal à bande limitée à partir d'une observation tronquée**

Ce résultat, en apparence négatif, a des conséquences positives. On considère le problème pratique suivant. Supposons que  $x \in L^2(\mathbb{R})$  soit à bande limitée à  $F$  (c'est à dire  $\text{supp}(\hat{f}) \subset F$ , et qu'on dispose d'observations de la forme

$$y(t) = \begin{cases} x(t) + n(t) & \text{si } t \notin T \\ 0 & \text{sinon ,} \end{cases}$$

où  $T$  est un sous-ensemble mesurable de  $\mathbb{R}$  et  $n \in L^2(\mathbb{R})$  est un bruit inconnu.

**Corollaire 1** *Si  $|T|.|F| < 1$ , il est possible de trouver un algorithme de reconstruction d'une approximation  $\tilde{x}$  de  $x$  à partir de  $y$ , stable au sens où il existe une constante  $C$  telle que*

$$\|\tilde{x} - x\| \leq C\|n\| .$$

**Preuve :** On considère l'opérateur

$$Q = (1 - P_T P_F)^{-1} .$$

$Q$  est bien défini si  $|T|.|F| < 1$ . Clairement,  $y = (1 - P_T)x + n$ , de sorte que, puisque  $x = P_F x$ , en définissant  $\tilde{x} = Qy$ ,

$$x - \tilde{x} = x - Qy = x - Q(x - P_T P_F x) - Qn = -Qn .$$

Par ailleurs,  $\|Q\| = \|(1 - P_T P_F)^{-1}\|$ , et comme pour tout opérateur  $A$  tel que  $\|A\| < 1$ , on a  $\|(1 - A)^{-1}\| \leq 1/(1 - \|A\|)$ , on en déduit

$$\|\tilde{x} - x\| \leq \frac{1}{1 - \sqrt{|F|.|T|}} \|n\| .$$

Pour ce qui concerne l'algorithme proprement dit, il suffit d'utiliser la série de Neumann (qui converge dès que  $\|P_T P_F\| < 1$ )

$$Qy = \sum_{n=0}^{\infty} (P_T P_F)^n y ,$$

et de poser

$$\tilde{x}^{(k)} = \sum_{n=0}^{\infty} (P_T P_F)^n y .$$

On a alors

$$\tilde{x}^{(k)} - y = P_T P_F x^{(k-1)} ,$$

ce qui suggère d'utiliser un algorithme itératif basé sur la récurrence

$$\tilde{x}^{(k)} = y + P_T P_F x^{(k-1)} ,$$

qui n'utilise que les deux opérateurs de projection orthogonale  $P_T$  et  $P_F$ . ♠

**2.3.2 Inégalités en dimension finie**

Concrètement, le traitement du signal numérique concerne des signaux de longueur finie, et donc des espaces de dimension finie. La notion de cohérence ci-dessous joue un rôle important.

**Définition 3** *Soit  $\Phi = \{\varphi_\lambda, \lambda \in \Lambda\}$  un repère d'un espace de Hilbert séparable  $\mathcal{H}$ . La cohérence de  $\Phi$  est la quantité*

$$\mu = \mu(\Phi) = \sup_{\lambda \neq \lambda'} |\langle \varphi_\lambda, \varphi_{\lambda'} \rangle| .$$

**Lemme 6** Soit  $\Phi = \{\varphi_\lambda, \lambda \in \Lambda\}$  un repère de  $\mathcal{H}$ , Hilbert de dimension  $N$ , normalisé de sorte que  $\|\varphi_\lambda\| = 1$  pour tout  $\lambda$ . Alors

$$\sqrt{\frac{|\Lambda| - N}{N(|\Lambda| - 1)}} \leq \mu(\Phi) \leq 1 .$$

**Remarque 9** Dans  $\mathcal{H} = \mathbb{C}^N$ , pour  $N$  premier différent de 2 ou 3, il est possible de trouver un repère  $\Phi$  formé de  $N^2$  vecteurs tels que

$$\mu(\Phi) = \frac{1}{\sqrt{N}} .$$

Il est constitué de vecteurs translétés en temps-fréquence

$$\varphi_{k\ell}[n] = e^{2i\pi\ell n/N} \varphi[n - k]$$

d'un vecteur appelé **séquence d'Alltop**, défini par

$$\varphi[n] = \frac{1}{\sqrt{N}} e^{2i\pi n^3/N} , \quad n = 0, \dots, N-1 ,$$

très utilisé en détection radar. Ce repère n'est pas très éloigné de l'optimum, qui est dans ce cas  $1/\sqrt{N+1}$ .

### L'inégalité de Donoho-Huo : repère « Kronecker $\cup$ Fourier »

La première version connue était une version de dimension finie de l'inégalité d'Heisenberg, concernant les représentations temporelle et fréquentielle [9]. Dans  $\mathbb{C}^N$ , on note  $\{\delta_0, \dots, \delta_{N-1}\}$  la base canonique, et  $\{\epsilon_0, \dots, \epsilon_{N-1}\}$  la base de Fourier finie. On s'intéresse à la décomposition parcimonieuse de signaux sur le dictionnaire formé de l'union de ces deux bases. Donoho et Huo ont été les premiers à faire un lien réel entre ce problème et les relations d'incertitude.

**Théorème 10 (Donoho-Stark)** Soit  $x \in \mathbb{C}^N$ ,  $x \neq 0$ , et notons  $\|x\|_0$  et  $\|\hat{x}\|_0$  les tailles des supports respectifs de  $x$  et  $\hat{x}$ . Alors

$$\begin{aligned} \|x\|_0 \cdot \|\hat{x}\|_0 &\geq N \\ \|x\|_0 + \|\hat{x}\|_0 &\geq 2\sqrt{N} \end{aligned}$$

On omettra la preuve ici, car elle est un cas particulier de l'inégalité d'Elad-Bruckstein ci-dessous.

**Exemple 3** Le signal "clôture" joue un rôle fondamental dans cette construction, dans la mesure où il apparaît comme cas limite pour ces inégalités. Soit  $N = p^2$  un carré parfait, et soit

$$x[np] = 1, \quad n = 0, \dots, p-1 , \quad x[m] = 0 \text{ si } m \neq np .$$

Il est facile de voir que

$$\hat{x}[kp] = 1, \quad k = 0, \dots, p-1 , \quad \hat{x}[\ell] = 0 \text{ si } \ell \neq kp .$$

Ainsi, dans ce cas,  $\|x\|_0 \cdot \|\hat{x}\|_0 = N$  et l'inégalité est une égalité.

**Corollaire 2 (Donoho-Huo)** Soit

$$x = \sum_{n=0}^{N-1} \alpha[n] \delta_n + \sum_{m=0}^{N-1} \beta[m] \epsilon_m \in \mathbb{C}^N .$$

Si

$$\|\alpha\|_0 + \|\beta\|_0 < \sqrt{N}$$

alors le problème (P0) a une solution unique. Cependant, il existe  $N$  et  $x \in \mathbb{C}^N$  tel que  $\|\alpha\|_0 + \|\beta\|_0 = \sqrt{N}$  et que le problème (P0) n'admette pas de solution unique.

**Preuve :** Notons pour simplifier  $\mathcal{D} = \{\varphi_k, k = 0, \dots, 2N - 1\}$  le repère constitué de l'union des deux bases (Kronecker et Fourier), et soit  $U$  l'opérateur d'analyse correspondant. Soient  $\gamma^1 = (\alpha^1, \beta^1)$  et  $\gamma^2 = (\alpha^2, \beta^2)$  tels que

$$x = \sum_{n=0}^{N-1} \alpha^1[n] \delta_n + \sum_{m=0}^{N-1} \beta^1[m] \epsilon_m = \sum_{n=0}^{N-1} \alpha^2[n] \delta_n + \sum_{m=0}^{N-1} \beta^2[m] \epsilon_m .$$

Alors

$$\gamma^1 - \gamma^2 \in \mathcal{N} = \text{Ker}(U^*) .$$

La structure de  $\mathcal{N}$  peut être caractérisée facilement. Soit  $\gamma = (u, v) \in \mathcal{N}$ . On a  $\sum_{n=0}^{N-1} u[n] \delta_n + \sum_{m=0}^{N-1} v[m] \epsilon_m = 0$ , d'où

$$v[k] = \left\langle \sum_{m=0}^{N-1} v[m] \epsilon_m, \epsilon_k \right\rangle = - \sum_{n=0}^{N-1} u[n] \langle \delta_n, \epsilon_k \rangle = - \sum_{n=0}^{N-1} u[n] \epsilon_k[n] = - \frac{1}{\sqrt{N}} \hat{u}[k] .$$

Donc  $\mathcal{N} = \{(u, -\hat{u}/\sqrt{N}), u \in \mathbb{C}^N\}$ , soit

$$\mathcal{N} = \{(u, -\tilde{u}), u \in \mathbb{C}^N\}, \quad \text{où} \quad \tilde{u} = \hat{u}/\sqrt{N} . \quad (2.2)$$

D'après l'inégalité de Donoho-Stark ci-dessus, les éléments de  $\mathcal{N}$  ont donc au moins  $2\sqrt{N}$  composantes non-nulles. On a donc

$$2\sqrt{N} \leq \|\gamma^1 - \gamma^2\|_0 \leq \|\gamma^1\|_0 + \|\gamma^2\|_0 ,$$

ce qui est incompatible avec  $\|\gamma^i\|_0 < \sqrt{N}$  pour  $i = 1, 2$ , sauf si  $\gamma^1 = \gamma^2$ .

Pour ce qui est de la non-unicité, reprenons l'exemple 3, qui montre que le signal "clôture"  $x$  admet deux représentations également parcimonieuses de longueur  $p = \sqrt{N}$  si  $N = p^2$  est un carré parfait. ♠

Ce résultat, qui porte sur le problème (P0), est de faible portée pratique, dans la mesure où le problème (P0) est difficile à résoudre numériquement, au contraire du problème (P1), auquel s'applique le résultat suivant. Introduisons tout d'abord la mesure suivante de concentration : étant données deux parties  $T, F \subset \{0, \dots, N - 1\}$ , on note

$$\mu_1(T, F) = \sup_{y \in \mathbb{C}^N, y \neq 0} \frac{\sum_{t \in T} |y[t]| + \sum_{f \in F} |\tilde{y}[f]|}{\|y\|_1 + \|\tilde{y}\|_1} . \quad (2.3)$$

**Corollaire 3** Soient  $T, F \subset \{0, \dots, N - 1\}$ , et soit  $x \in \mathbb{C}^N$  tel que

$$x = \sum_{t \in T} \alpha[t] \delta_t + \sum_{f \in F} \beta[f] \epsilon_f .$$

Si  $T, F$  sont tels que

$$\mu_1(T, F) < \frac{1}{2}$$

alors  $\gamma = (\alpha, \beta) \in \mathbb{C}^{2N}$  est l'unique solution du problème (P1).

*Preuve :* Soit donc  $x \in \mathbb{R}^N$ , soit  $\mathcal{D}$  le repère formé de l'union de la base de Fourier et de la base de Kronecker, et soit  $\gamma \in \mathbb{C}^{2N}$  tel que  $x = U^* \gamma$ ,  $U^*$  étant l'opérateur de synthèse du repère. Dire que  $\|\gamma\|_1$  est minimale revient à dire que

$$\|\gamma + \delta\|_1 - \|\gamma\|_1 > 0, \quad \forall \delta \in \mathcal{N},$$

où  $\mathcal{N} = \text{Ker}(U^*)$ . Notant  $S_\gamma \subset \{0, \dots, 2N - 1\}$  le support de  $\gamma$  et  $\bar{s}_\gamma$  l'ensemble complémentaire, on a

$$\|\gamma + \delta\|_1 - \|\gamma\|_1 = \sum_{\lambda \in \bar{S}_\gamma} |\delta_\lambda| + \sum_{\lambda \in S_\gamma} (|\gamma_\lambda + \delta_\lambda| - |\gamma_\lambda|) \geq \sum_{\lambda \in \bar{S}_\gamma} |\delta_\lambda| - \sum_{\lambda \in S_\gamma} |\delta_\lambda| .$$

Donc l'unicité sera assurée dès que

$$\sum_{\lambda \in \bar{S}_\gamma} |\delta_\lambda| > \sum_{\lambda \in S_\gamma} |\delta_\lambda| \quad \forall \delta \in \mathcal{N},$$

soit aussi

$$\|\delta\|_1 > 2 \sum_{\lambda \in \bar{S}_\gamma} |\delta_\lambda| \quad \forall \delta \in \mathcal{N}.$$

Rappelons maintenant que  $\mathcal{N} = \{(u, -\tilde{u}), u \in \mathbb{C}^N\}$ . Cette condition devient donc, en notant  $T$  et  $F$  les parties temporelle et fréquentielle du support de  $\gamma$  ( $S_\gamma = T \cup F$ )

$$\|u\|_1 + \|\tilde{u}\|_1 > 2 \left( \sum_{t \in T} |x_t| + \sum_{f \in F} |\tilde{x}_f| \right),$$

ce qui conclut la démonstration. ♠

Donoho et Huo montrent ensuite le résultat suivant [9]

**Proposition 5** *Avec les mêmes notations que ci-dessus,*

$$\mu_1(T, F) \leq \frac{|T| + |F|}{\sqrt{N} + 1}.$$

On en déduit immédiatement

**Corollaire 4** *Avec les mêmes notations que ci-dessus, si*

$$\|\alpha\|_0 + \|\beta\|_0 \leq \frac{\sqrt{N}}{2},$$

alors  $\gamma = (\alpha, \beta)$  est l'unique solution de (P1).

Comme on va le voir ci-dessous, la constante 1/2 du corollaire ci-dessus n'est pas du tout optimale, la valeur optimale étant  $\sqrt{2} - 1/2 \approx 0.914$ .

**L'inégalité d'Elad-Bruckstein : repère « union de deux bases orthonormées  $\mathcal{B}_1 \cup \mathcal{B}_2$  »**

Cette inégalité [12] généralise l'inégalité de Donoho-Huo à des représentations par les coefficients de décompositions sur des bases orthonormées.

**Théorème 11** *Soient  $\Phi = \{\varphi_0, \dots, \varphi_{N-1}\}$  et  $\Psi = \{\psi_0, \dots, \psi_{N-1}\}$  deux bases orthonormées de  $\mathbb{C}^N$ . Soit  $x \in \mathbb{C}^N$ , tel que  $\|x\| = 1$ ;  $x$  admet les décompositions*

$$x = \sum_{n=0}^{N-1} \alpha[n] \varphi_n = \sum_{n=0}^{N-1} \beta[n] \psi_n.$$

Alors

$$\|\alpha\|_0 + \|\beta\|_0 \geq 2\sqrt{\|\alpha\|_0 \cdot \|\beta\|_0} \geq \frac{2}{\mu(\Phi \cup \Psi)}.$$

**Preuve :** notons

$$\mu = \mu(\Phi \cup \Psi) = \max_{m,n} |\langle \varphi_n, \psi_m \rangle|$$

la cohérence du repère  $\Phi \cup \Psi$ . Soit  $x \in \mathbb{C}^N$ , tel que  $\|x\| = 1$ . D'après la formule de Parseval, on a

$$\|x\|^2 = 1 = \left\langle \sum_{n=0}^{N-1} \alpha[n] \varphi_n, \sum_{n=0}^{N-1} \beta[n] \psi_n \right\rangle = \sum_{m,n=0}^{N-1} \alpha[n] \overline{\beta[m]} \langle \varphi_n, \psi_m \rangle,$$

d'où

$$1 \leq \sum_{m,n=0}^{N-1} |\alpha[n]| \cdot |\beta[m]| \cdot |\langle \varphi_n, \psi_m \rangle| \leq \mu \|\alpha\|_1 \|\beta\|_1 .$$

On a de plus  $\|\alpha\| = \|\beta\| = 1$ . Notons  $A = \|\alpha\|_0$  et  $B = \|\beta\|_0$ . Sans perte de généralité on peut supposer  $\alpha[A] = \alpha[A+1] = \dots = \alpha[N-1] = 0$  et  $\beta[B] = \beta[B+1] = \dots = \beta[N-1] = 0$ , et que tous les coefficients sont réels positifs. On a donc

$$1 \leq \mu \sum_{a=0}^{A-1} \alpha[a] \sum_{b=0}^{B-1} \beta[b] .$$

Pour chercher le minimum du membre de droite, sous contrainte  $\|\alpha\| = \|\beta\| = 1$ , introduisons deux multiplicateurs de Lagrange  $\lambda_A$  et  $\lambda_B$ , et optimisons la quantité

$$\sum_0^{A-1} \alpha[a] \sum_0^{B-1} \beta[b] + \lambda_A \left( \sum_0^{A-1} \alpha[a]^2 - 1 \right) + \lambda_B \left( \sum_0^{B-1} \beta[b]^2 - 1 \right) .$$

Les équations variationnelles conduisent à

$$\alpha[a] = C^{\text{te}} = \frac{1}{\sqrt{A}} , \quad \beta[b] = C^{\text{te}} = \frac{1}{\sqrt{B}} ,$$

de sorte que l'on obtient

$$1 \leq \mu \sqrt{AB} ,$$

ce qui est la première inégalité. La seconde résulte de l'inégalité arithmético-géométrique. ♠

L'inégalité d'Elad-Bruckstein a une conséquence importante : si un signal admet une représentation suffisamment parcimonieuse dans un dictionnaire formé de l'union de deux bases orthonormées, alors il n'en existe pas de plus parcimonieuse. Plus précisément,

**Proposition 6** Soit  $\mathcal{D} = \Phi \cup \Psi$  l'union de deux bases orthonormées. Soit  $x \in \mathbb{C}^N$  et soient

$$x = \sum_0^{N-1} \alpha^1[n] \varphi_n + \sum_0^{N-1} \beta^1[m] \psi_m = \sum_0^{N-1} \alpha^2[n] \varphi_n + \sum_0^{N-1} \beta^2[m] \psi_m$$

deux décompositions de  $x$  sur  $\mathcal{D}$ . Alors, en posant  $\gamma^i = (\alpha^i, \beta^i)$

$$\|\gamma^1\|_0 + \|\gamma^2\|_0 \geq \frac{2}{\mu(\mathcal{D})}$$

**Preuve :** Avec les notations ci-dessus, on obtient facilement

$$\sum_{n=0}^{N-1} (\alpha^1[n] - \alpha^2[n]) \varphi_n = - \sum_{n=0}^{N-1} (\beta^1[n] - \beta^2[n]) \psi_n ,$$

d'où d'après le résultat précédent

$$\|\alpha^1 - \alpha^2\|_0 + \|\beta^1 - \beta^2\|_0 \geq \frac{2}{\mu(\mathcal{D})} .$$

Supposons que  $\|\alpha^1\|_0 + \|\beta^1\|_0 < 1/\mu(\mathcal{D})$  et  $\|\alpha^2\|_0 + \|\beta^2\|_0 < 1/\mu(\mathcal{D})$ , alors

$$\|\alpha^1 - \alpha^2\|_0 + \|\beta^1 - \beta^2\|_0 < \frac{2}{\mu(\mathcal{D})} ,$$

d'où contradiction. ♠

**Corollaire 5 (Unicité)** Soit  $x = \sum_{n=0}^{N-1} \alpha[n] \varphi_n + \sum_{n=0}^{N-1} \beta[n] \psi_n \in \mathbb{C}^N$ , avec  $x \neq 0$ . Si  $\|\alpha\|_0 + \|\beta\|_0 < 1/\mu(\mathcal{D})$ , alors  $(\alpha, \beta) \in \mathbb{C}^{2N}$  est l'unique solution de (P0).

**Remarque 10** Elad et Bruckstein fournissent aussi des conditions assurant que la solution de (P0) soit aussi solution de (P1). On n'abordera pas ces aspects ici, mais on y reviendra dans la section consacrée aux propriétés de noyau ci-dessous.

Comme on l'a vu, la cohérence joue un rôle fondamental dans ces bornes. Cette quantité ne peut être rendue aussi petite que l'on voudrait :

**Lemme 7** Soient  $\Phi = \{\varphi_0, \dots, \varphi_{N-1}\}$  et  $\Psi = \{\psi_0, \dots, \psi_{N-1}\}$  deux bases orthonormées de  $\mathbb{C}^N$ . Alors

$$\mu(\Phi \cup \Psi) \geq \frac{1}{\sqrt{N}} .$$

**Remarque :** Cette borne est en fait atteinte pour le cas des bases de Kronecker et de Fourier.

### L'inégalité de Donoho-Elad : repère quelconque

Les résultats d'Elad et Bruckstein permettent de généraliser ceux de Donoho et Huo au cas de la représentation sur l'union de deux bases quelconques, plus nécessairement la base canonique et la base de Fourier. On peut naturellement se poser la question de l'existence d'un résultat équivalent dans le cas d'un repère quelconque. Il est pour cela nécessaire d'introduire la notion de **rang de Kruskal**, ou **spark** d'un repère.

**Définition 4** Etant donnée une matrice  $M$ , son rang de Kruskal, ou spark, est le plus petit nombre  $\sigma$  tel qu'il soit possible de trouver  $\sigma$  colonnes de  $M$  linéairement dépendantes.

**Proposition 7** Soit  $\mathcal{D}$  un repère de  $\mathbb{C}^N$ , et soit  $\sigma(\mathcal{D})$  le rang de Kruskal de la matrice de son opérateur de synthèse  $V = U^*$ . Si  $x \in \mathbb{C}^N$  admet deux représentations

$$x = \sum_{k=0}^{N-1} \gamma^1[k] \varphi_k = \sum_{k=0}^{N-1} \gamma^2[k] \varphi_k ,$$

alors

$$\|\gamma^1\|_0 + \|\gamma^2\|_0 \geq \sigma(\mathcal{D}) .$$

**Preuve :** Soit  $\delta = \gamma^1 - \gamma^2$ . Alors  $\delta$  appartient au noyau de  $U^*$ , d'où  $\|\delta\|_0 \geq \sigma$ . Par ailleurs, la quasi-norme  $\ell^0$  vérifie l'inégalité triangulaire, donc  $\|\gamma^2 - \gamma^1\|_0 \leq \|\gamma^1\|_0 + \|\gamma^2\|_0$ , ce qui conclut la preuve.

♠

On en déduit immédiatement le résultat d'unicité suivant :

**Corollaire 6** Soit

$$x = \sum_{k=0}^{N-1} \gamma[k] \varphi_k .$$

Si  $\|\gamma\|_0 < \sigma(\mathcal{D})/2$ , alors  $x$  n'admet pas de représentation plus parcimonieuse dans  $\mathcal{D}$ .

Ainsi, le rang de Kruskal joue ici, un rôle similaire au rôle joué par la cohérence  $\mu$  dans le cas de repères constitués d'unions de bases orthonormées. Il est intéressant de comparer précisément ces quantités. C'est l'objet du théorème suivant, dont la preuve se trouve dans [8].

**Théorème 12 (Donoho-Elad)** 1. Soit  $\mu(\mathcal{D})$  la cohérence du repère  $\mathcal{D}$ . Alors

$$\sigma(\mathcal{D}) > \frac{1}{\mu(\mathcal{D})} .$$

2. Si  $\mathcal{D} = \mathcal{B}_1 \cup \mathcal{B}_2$  est l'union de deux bases orthonormées, alors

$$\sigma(\mathcal{D}) > \frac{2}{\mu(\mathcal{D})} .$$

**Remarque 11** La référence [8] fournit en outre des conditions supplémentaires permettant d'assurer que la solution du problème (P0) est obtenue en résolvant (P1), pour lequel il existe des algorithmes efficaces. On n'entrera pas ici dans ces raffinements... mais il est conseillé de jeter un oeil à l'article.

## 2.4 Propriétés de noyau (NSP)

### 2.4.1 NSP et identification

Nous avons déjà vu apparaître à divers endroits, notamment dans la démonstration du Corollaire 3, des propriétés du noyau de l'opérateur de synthèse. On se focalise ici sur ces propriétés, dans le cadre du problème général d'optimisation

$$(Pq) \quad \min \|\alpha\|_q \quad \text{sous contrainte} \quad x = \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda ,$$

où  $x \in \mathbb{C}^m$  représente l'observation,  $\alpha \in \mathbb{C}^N$  est le vecteur des coefficients du développement, supposé parcimonieux, et  $\mathcal{D} = \{\varphi_\lambda, \lambda \in \Lambda\}$  est un repère de  $\mathbb{C}^m$  constitué de  $N = |\Lambda|$  vecteurs. En représentant  $\alpha$  et  $x$  par des vecteurs colonne, ce problème s'écrit matriciellement sous la forme

$$(Pq) \quad \min \|\alpha\|_q \quad \text{sous contrainte} \quad x = V\alpha ,$$

$A \in \mathcal{M}_{m,N}$  étant la matrice de l'opérateur de synthèse  $V = U^*$  du repère.

Avant d'aller plus loin, il est utile d'introduire quelques notations. Pour un sous-ensemble  $\Lambda_0 \subset \Lambda$  de l'index  $\Lambda$ , on notera  $\overline{\Lambda_0}$  son complémentaire. Pour un vecteur  $\alpha = \{\alpha_\lambda, \lambda \in \Lambda\}$ , on notera  $\alpha_{\Lambda_0} = \{\alpha_\lambda, \lambda \in \Lambda_0\}$  sa restriction à  $\Lambda_0$ .

**Définition 5** Soit  $s$  un entier positif. Le repère  $\mathcal{D}$  (ou la matrice correspondante  $A$ ) satisfait la propriété du noyau d'ordre  $s$  par rapport à  $\ell^q$ , (notée NSP<sub>q</sub>) si pour tout  $\alpha \in \ker(A) \setminus \{0\}$  et pour tout  $\Lambda_0 \subset \Lambda$  tel que  $|\Lambda_0| \leq s$ , la condition ci-dessous est satisfaite :

$$\|\alpha_{\Lambda_0}\|_q^q < \|\alpha_{\overline{\Lambda_0}}\|_q^q .$$

Notons que cette propriété est équivalente à

$$\|\alpha_{\Lambda_0}\|_q^q < \frac{1}{2} \|\alpha\|_q^q .$$

La propriété (NSP<sub>q</sub>) est une condition suffisante permettant d'assurer l'identification de  $\alpha$ .

**Proposition 8** Avec les mêmes notations que ci-dessus, soit  $s$  un entier positif. Tout vecteur  $\alpha$  tel que  $\|\alpha\|_0 \leq s$  est identifié en résolvant (Pq) si et seulement si la matrice  $A$  satisfait la propriété du noyau NSP<sub>q</sub> d'ordre  $s$  par rapport à  $\ell^q$ .

**Preuve :** i) Supposons que pour tout  $\alpha$  tel que  $\|\alpha\|_0 \leq s$ , la résolution du problème (Pq) donne  $\alpha$  comme solution unique. Soit  $\delta \in \ker(A) \setminus \{0\}$ , et soit  $\Lambda_0 \subset \Lambda$  tel que  $|\Lambda_0| \leq s$ . Il est clair que  $A\delta_{\Lambda_0} = -A\delta_{\Lambda_0^c}$ ; comme  $\|\delta_{\Lambda_0}\|_0 \leq s$  on a nécessairement  $\|\delta_{\Lambda_0^c}\|_q > \|\delta_{\Lambda_0}\|_0$ .

ii) Supposons maintenant (NSPq) à l'ordre  $s$ , et soit  $\alpha$  tel que  $A\alpha = x$ , avec  $\|\alpha\|_0 \leq s$ . On note  $\Lambda_0$  le support de  $\alpha$ . Soit  $\beta \neq \alpha$  tel que  $A\beta = x$ , alors  $\delta = \alpha - \beta \in \ker(A) \setminus \{0\}$ . Calculons

$$\|\alpha\|_q^q = \|\alpha - \beta_{\Lambda_0} + \beta_{\Lambda_0}\|_q^q \leq \|\alpha - \beta_{\Lambda_0}\|_q^q + \|\beta_{\Lambda_0}\|_q^q < \|\delta_{\Lambda_0^c}\|_q^q + \|\beta_{\Lambda_0}\|_q^q = \|\beta_{\Lambda_0^c}\|_q^q + \|\beta_{\Lambda_0}\|_q^q = \|\beta\|_q^q ,$$

ce qui montre que  $\alpha$  minimise la quasi-norme  $\ell^q$ . ♠

**Remarque 12** En plus des résultats d'identification ci-dessus, il est également possible de démontrer des résultats complémentaires (voir [14] pour plus de détails, et des démonstrations). Par exemple :

1. Si  $q = 1$ , il est possible de trouver un vecteur à une seule composante non-nulle tel qu'il ne soit pas la solution de (Pq).
2. Si  $\alpha$  est identifié en résolvant (Pq), alors  $\alpha$  est également identifié en résolvant (Pr) pour tout  $r \leq q$ .

### 2.4.2 Le cas $p = 1$ : lien avec la cohérence

Dans le cas  $p = 1$ , il est possible de faire un lien avec la cohérence.

**Corollaire 7** Soit  $\mathcal{D}$  un repère tel que ses éléments sont normalisés ( $\|\varphi_\lambda\|_2 = 1$  pour tout  $\lambda$ ), et soit  $\mu(\mathcal{D})$  sa cohérence. Si

$$\mu(\mathcal{D}) < \frac{1}{2s-1}$$

pour un certain entier positif  $s$ , alors tout  $\alpha$  tel que  $\|\alpha\|_0 \leq s$  est identifié en résolvant (P1).

**Preuve :** Il suffit de montrer que  $\mathcal{D}$  vérifie (NSP1) d'ordre  $s$ . Pour cela, soit  $\alpha \in \ker(A)$ . Pour tout  $\ell$ , on a  $\alpha[\ell]\varphi_\ell = -\sum_{k \neq \ell} \alpha[k]\varphi_k$ , de sorte que  $|\alpha[\ell]| = |\sum_{k \neq \ell} \alpha[k]\langle \varphi_k, \varphi_\ell \rangle| \leq \mu(\mathcal{D})[\|\alpha\|_1 - |\alpha[\ell]|]$ . Donc  $|\alpha[\ell]|(1 + \mu(\mathcal{D})) \leq \mu(\mathcal{D})\|\alpha\|_1$ , et

$$\forall \Lambda_0 \subset \Lambda, \text{ tel que } |\Lambda_0| \leq s, \|\alpha_{\Lambda_0}\|_1 \leq s \frac{\mu(\mathcal{D})}{1 + \mu(\mathcal{D})} \|\alpha\|_1 .$$

Par conséquent, si

$$s \frac{\mu(\mathcal{D})}{1 + \mu(\mathcal{D})} < \frac{1}{2} ,$$

la propriété (NSP1) à l'ordre  $s$  est vérifiée, ce qui donne le résultat. ♠

**Remarque 13** En fait, il est possible de vérifier que cette borne est très restrictive. Par exemple, considérons le système d'Alltop décrit dans la Remarque 9 : un repère en dimension  $N$ , constitué de  $n = N^2$  vecteurs, de cohérence  $\mu = \sqrt{N}$ . Un petit calcul montre que si  $N$  est de l'ordre de 100 (rappelons que  $N$  doit être premier différent de 2 et 3), la cardinalité des vecteurs identifiables doit être telle que  $s \leq 5$ , ce qui est vraiment faible. En fait, cette borne est bien trop restrictive, il est possible d'en démontrer de plus fines, mais les démonstration sont bien plus techniques.

### 2.4.3 Stabilité : instance optimalité

Nous n'avons jusqu'ici considéré que le cas idéal où le signal  $x$  s'écrit **exactement** sous forme d'une combinaison linéaire parcimonieuse d'éléments du repère. En pratique, la situation est souvent différente, et il a plus de sens de modéliser le signal observé sous la forme

$$x = A\alpha + \epsilon ,$$

où  $\epsilon$  est inconnu, et peut représenter toutes sortes de perturbations (erreurs de mesure, bruits,...). Deux pistes sont alors possibles :

- Utiliser la même approche qu'auparavant (par exemple résoudre (P1) en utilisant Basis Pursuit, ou Orthogonal Matching Pursuit), et démontrer des résultats de stabilité. En notant  $\alpha^*$  la solution de (P1) fournie par l'algorithme (BP ou OMP), on peut essayer de borner  $\|\alpha - \alpha^*\|$  par une quantité dépendant de  $\epsilon$ .
- Modifier l'algorithme, en cherchant à résoudre par exemple le problème (P1') (à l'aide de LASSO, ou Matching Pursuit avec un nombre limité d'itérations).

On se limitera ici à discuter la première piste, pour laquelle on utilisera la notion d'**instance optimalité** introduite dans [5]. Pour  $\alpha \in \mathbb{R}^n$ , on note

$$\sigma_s(\alpha)_q = \inf_{z \in \Sigma_s} \|\alpha - z\|_q$$

sa distance au plus proche élément de  $\Sigma_s$ , mesurée en (quasi-)norme  $\ell^q$ .

**Définition 6** *Un algorithme de reconstruction est instance optimal si pour tout  $\alpha \in \Sigma_s$  la solution  $\alpha^*$  satisfait*

$$\|\alpha^* - \alpha\|_q \leq C \sigma_s(\alpha)_q ,$$

pour une certaine constante  $C$ .

**Proposition 9** *Soit  $0 < q \leq 1$ , et supposons la propriété du noyau (NSP $_q$ ) satisfaite à l'ordre  $s$ . Alors tout algorithme résolvant (P $_q$ ) satisfait la propriété d'instance optimalité.*

**Preuve :** Supposons (NSP $_q$ ) satisfaite à l'ordre  $s$ . Soit  $\delta \in \ker(A) \cap \mathcal{S}_q^N$ , où  $\mathcal{S}_q^N$  est la sphère unité dans  $\mathbb{C}^N$ , par rapport à la quasi-norme  $\ell^q$

$$\mathcal{S}_q^N = \{ \delta \in \mathbb{C}^N, \|\delta\|_q = 1 \} .$$

Alors on sait que pour tout  $\Lambda_0$  tel que  $|\Lambda_0| \leq s$ ,  $\|\delta_{\Lambda_0}\|_q^q < 1/2$ . Soit maintenant

$$c = 2 \sup_{|\Lambda_0| \leq s} \sup_{\delta \in \ker(A) \setminus \{0\}} \|\delta_{\Lambda_0}\|_q^q < 1 .$$

On a alors

$$\|\delta_{\Lambda_0}\|_q^q \leq \frac{c}{2} \|\delta\|_q^q - \frac{1-c}{2} \|\delta\|_q^q \implies \|\delta_{\Lambda_0}\|_q^q \leq \|\delta_{\Lambda_0^c}\|_q^q - (1-c) \|\delta\|_q^q$$

Soient maintenant  $\alpha, \alpha^*$ , et posons  $\delta = \alpha - \alpha^*$ . On a  $\|\alpha\|_q \geq \|\alpha^*\|_q$ , et

$$\|\alpha\|_q = \|\alpha_{\Lambda_0}\|_q + \|\alpha_{\Lambda_0^c}\|_q \geq \|(\alpha - \delta)_{\Lambda_0}\|_q + \|(\alpha - \delta)_{\Lambda_0^c}\|_q \geq \|\alpha_{\Lambda_0}\|_q - \|\delta_{\Lambda_0}\|_q - \|\alpha_{\Lambda_0^c}\|_q + \|\delta_{\Lambda_0^c}\|_q ,$$

donc

$$\|\delta_{\Lambda_0^c}\|_q^q \leq \|\delta_{\Lambda_0}\|_q^q + 2\|\alpha_{\Lambda_0^c}\|_q^q \leq \|\delta_{\Lambda_0}\|_q^q + 2\sigma_s(\alpha)_q^q .$$

Ainsi

$$\|\delta\|_q^q = \|\delta_{\Lambda_0}\|_q^q + \|\delta_{\Lambda_0^c}\|_q^q \leq \|\delta_{\Lambda_0}\|_q^q - (1-c) \|\delta\|_q^q + \|\delta_{\Lambda_0}\|_q^q + 2\sigma_s(\alpha)_q^q \leq c \|\delta\|_q^q + 2\sigma_s(\alpha)_q^q ,$$

d'où

$$\|\delta\|_q \leq \left( \frac{2}{1-c} \right)^{1/q} \sigma_s(\alpha)_q ,$$

ce qui conclut la démonstration. ♠

## 2.5 Exercices

### 2.5.1 Maths

1. Démontrer la Proposition 2. On pourra prendre  $\tilde{\varphi}_\lambda = \mathcal{R}^{-1}\varphi_\lambda$ .
2. Montrer que la matrice de Gram

$$G = UU^* = \{\langle \varphi_\lambda, \varphi_{\lambda'} \rangle, \lambda, \lambda' \in \Lambda\}$$

d'un repère  $\{\varphi_{\lambda_0}, \dots, \varphi_{\lambda_N}\}$  est inversible.

3. Démontrer la Proposition 4.
4. Démontrer le Théorème 8. On pourra supposer  $v_f, v_{\hat{f}} < \infty$  (sinon l'inégalité est trivialement vérifiée), et appliquer l'inégalité de Cauchy-Schwarz à

$$\int_{-\infty}^{\infty} t \frac{d}{dt} (|f(t)|^2) dt .$$

5. Démontrer le Lemme 7, et montrer que la borne est atteinte pour le case des bases de Kronecker et de Fourier.
6. Soit  $A \in \mathcal{M}_{m,N}$  une matrice satisfaisant la propriété (NSPq) à l'ordre  $s$ .
  - a) Soit  $B \in \mathcal{M}_{r,N}$ , et soit  $\tilde{A} \in \mathcal{M}_{m+r,N}$  la concaténation verticale de  $A$  et  $B$ . Montrer que  $\tilde{A}$  vérifie toujours (NSPq) à l'ordre  $s$  (indication : comparer le noyau de  $\tilde{A}$  au noyau de  $A$ ).
  - b) Soit  $C \in \mathcal{M}_{m,m}$  une matrice non-singulière, et soit  $\tilde{A} = CA \in \mathcal{M}_{m,N}$ . Montrer que  $\tilde{A}$  vérifie toujours (NSPq) à l'ordre  $s$ .
7. Montrer que la cohérence du système d'Alltop (voir Remarque 9) en dimension  $N$  (premier, différent de 2 et 3) vaut  $1/\sqrt{N}$ . Le calcul est long, et pas vraiment intéressant... il peut être trouvé dans [14], page 43.

### 2.5.2 Numérique

On va tester les diverses méthodes de reconstruction parcimonieuse proposées par la toolbox

*Sparselab*

développé par le département de statistique de l'Université de Stanford, en particulier `SolveBP`, `SolveMP` et `SolveOMP`.

Dans les exercices ci-dessous, les repères considérés sont représentés par la matrice de leur opérateur de synthèse  $U^*$  (ou parfois la matrice adjointe, qui correspond à l'opérateur d'analyse  $U$ ). On pourra télécharger l'archive se trouvant sur le site MOODLE du cours, qui contient

- Quelques exemples de script, exploitant les fonctions de `SparseLab`.
- Deux fichiers `.mat`, à charger en utilisant l'instruction `load`, et contenant les matrices associées à deux bases MDCT : ce sont les matrices des opérateurs de synthèse  $U^*$ , les matrices des opérateurs d'analyse sont donc leurs conjuguées Hermitiennes. Ces fichiers existent en deux versions : petite taille et grande taille.

1. Etude de la décomposition d'un signal simple avec une base MDCT : en prenant deux exemples de signaux, par exemple `HiSine` et `Bumps`, calculer les coefficients de leur développement sur les deux bases MDCT (individuellement). Ci dessous les instructions à suivre si on utilise les fichiers issus de `M2PS_petit.mat`.

```
>> coeff1 = wmat1'*HiSine';
```

produit les coefficients du développement de `HiSine` (qui doit être transposé car il est sous la forme d'un vecteur ligne) sur une base MDCT de  $\mathbb{R}^{1024}$  associée à 32 indices temporels et 32 indices fréquentiels. On pourra la visualiser par

```
>> coeff1 = reshape(coeff1,32,32);  
>> imagesc(abs(coeff1)); axis xy
```

`wmat2` est la matrice de l'opérateur de synthèse d'une base MDCT de  $\mathbb{R}^{1024}$  associée à 8 indices temporels et 128 indices fréquentiels.

Dans le cas où `M2PS.mat` est utilisé, les instructions sont les mêmes, mais les signaux sont de longueur 2048, `wmat1` est la matrice de l'opérateur d'analyse d'une base MDCT à 32 indices temporels et 64 indices fréquentiels, et `wmat2` est la matrice de l'opérateur d'analyse d'une base MDCT à 8 indices temporels et 256 indices fréquentiels.

2. Alternativement, les fonctions `RealDiracFourierMatrix.m` et `DiracFourierMatrix.m` permettent de générer les matrices des opérateurs de synthèse des dictionnaires  $Dirac \cup Cosinus$  et  $Dirac \cup Fourier$ . On utilisera plutôt le premier.
3. Un vecteur parcimonieux dans le repère représenté par la matrice  $V$  de son opérateur de synthèse  $U^*$  peut être construit de la façon suivante :
  - Générer un vecteur  $\alpha$  de longueur  $2N$ , contenant  $m$  coefficients non nuls. Par exemple :

```
>> alpha = zeros(2*N,1);  
>> locs = randperm(N);  
>> locs = locs(1:m);  
>> alpha(locs) = randn(m,1);
```
  - Générer le vecteur  $X = U^*\alpha$  correspondant :

```
>> X = V * alpha;
```

On pourra tracer  $X$ , pour voir...
4. Effectuer la décomposition parcimonieuse de vecteurs ainsi générés à l'aide de `SolveBP`, `SolveMP` ou `SolveOMP`. Ne pas hésiter à utiliser le `help` pour connaître la syntaxe des différentes fonctions. Dans chaque cas, on pourra calculer l'erreur de reconstruction.
5. pour ceux qui ne sont pas très à l'aise, les scripts `MPdemoscrypt.m` (Matching Pursuit et Matching Pursuit Orthogonal) et `SparseDecompTest.m` (comparaison de MP, OMP et BP) peuvent être utilisés... et modifiés.

## 3 Echantillonnage Compressif

Dans ce chapitre, on revient sur le problème d'échantillonnage, en abordant (de façon succincte) une approche récente, appelée **échantillonnage compressif** (**compressed sensing** en anglais). Le domaine est devenu très vaste durant ces dernières années, avec notamment des applications impressionnantes, par exemple l'appareil photo à pixel unique [11] ou encore en astronomie (voir [19]). On se contentera ici d'une introduction et de quelques résultats représentatifs. On pourra voir par exemple [3] pour une introduction compréhensible à l'échantillonnage compressif, et [14] pour un exposé plus complet.

### 3.1 Echantillonnage compressif

Comme on l'a vu, le concept de base de l'échantillonnage était de considérer des valeurs ponctuelles d'une fonction (ou suite) à échantillonner (ou sous-échantillonner). A cette notion basique a succédé une notion plus élaborée, qui est la notion d'acquisition, ou de mesure (*sensing* en anglais), qui a suscité des questions d'un type différent. Par exemple, une question que l'on peut se poser est la suivante : supposons que l'on sache qu'un signal  $x \in \mathbb{C}^N$  soit représenté parcimonieusement dans un repère  $\mathcal{D}$  de taille  $n$ , par exemple qu'il puisse s'écrire comme combinaison linéaire de  $s \ll n$  éléments du repère.

- Si  $\mathcal{D}$  est une base orthonormée, et si  $N = n$  est très grand, on n'a pas nécessairement la possibilité (ni l'envie) de calculer tous les coefficients du développement de  $x$  sur  $\mathcal{D}$ , sachant qu'ils sont quasiment tous nuls.
- Si  $\mathcal{D}$  n'est pas une base, il semble également inutile (et peut être impossible) de calculer toutes les décompositions possibles pour sélectionner celle qui sera parcimonieuse.

La question est alors : comment estimer de façon la plus efficace les coefficients de la décomposition parcimonieuse ?

Plus concrètement : étant donné un repère  $\mathcal{D} = \{\varphi_1, \dots, \varphi_n\}$  et un entier positif  $s$ , on note

$$\Sigma_s(\mathcal{D}) = \left\{ x \in \mathbb{C}^N, x = \sum_{\lambda \in \Lambda} \alpha_\lambda \varphi_\lambda, |\Lambda| \leq s \right\}$$

l'ensemble des signaux de  $\mathbb{C}^N$  qui peuvent s'écrire comme combinaisons linéaires d'au plus  $s$  éléments de  $\mathcal{D}$ . On dira que ces signaux sont  $s$ -parcimonieux dans  $\mathcal{D}$ . On notera en particulier  $\Sigma_s$  l'ensemble des vecteurs  $\alpha$  ne possédant pas plus de  $s$  composantes non-nulles.

**Remarque 14** Il est important de noter que  $\Sigma_s$  n'est pas un espace linéaire. On voit facilement en particulier que  $\Sigma_s + \Sigma_s = \Sigma_{2s}$ . En fait,  $\Sigma_s$  est l'union de  $C_n^s$  sous-espaces de dimension  $s$  de  $\mathbb{R}^n$ .

Etant donné un signal  $s$ -parcimonieux  $x \in \Sigma_s(\mathcal{D})$ , on considère un système d'acquisition effectuant des mesures sur  $x$  de la forme

$$y_k = L_k(x) = \langle x, \psi_k \rangle, \quad k = 1, \dots, m$$

pour certains vecteurs  $\psi_k \in \mathbb{C}^N$ . On a donc

$$y_k = \sum_{\ell=1}^n \langle \varphi_\ell, \psi_k \rangle \alpha_\ell,$$

soit sous forme matricielle

$$y = A\alpha,$$

où  $A \in \mathcal{M}_{m,n}$  est la matrice des produits scalaires des vecteurs  $\varphi_\ell$  du repère  $\mathcal{D}$  et des vecteurs de mesure  $\psi_k$ . Le problème de l'échantillonnage compressif peut se formuler ainsi : **Est-il possible de construire des systèmes de mesure  $\{\psi_1, \dots, \psi_m\}$  capables d'identifier tout vecteur  $s$ -parcimonieux  $\alpha \in \Sigma_s(\mathcal{D})$  ?** La réponse sera positive ou négative, en fonction des valeurs de  $m$  et  $s$ .

Un premier élément de réponse est donné par le résultat suivant. Supposons qu'il existe deux vecteurs  $\alpha, \alpha' \in \Sigma_s$  tels que  $y = A\alpha = A\alpha'$ . Alors  $\alpha - \alpha' \in \ker(A) \cap \Sigma_{2s}$ . Inversement, tout  $\delta \in \Sigma_{2s}$  peut toujours s'écrire sous la forme  $\delta = \alpha - \alpha'$  avec  $\alpha, \alpha' \in \Sigma_s$ . Supposons  $\delta \in \ker(A)$  et  $\alpha \neq 0$ , on a alors nécessairement  $A\alpha = A\alpha'$  et donc pas unicité pour la solution  $s$ -parcimonieuse de  $y = A\alpha$ . On a donc montré

**Lemme 8** *Une condition nécessaire et suffisante pour que tout vecteur  $s$ -parcimonieux  $\alpha$  soit identifié de façon unique à partir des mesures  $y = A\alpha$  est que*

$$\Sigma_{2s}(I) \cap \ker(A) = \{0\} .$$

Notons que ce résultat, pour intéressant qu'il soit, ne dit pas comment résoudre le problème !

**Corollaire 8** *Tout vecteur  $s$ -parcimonieux  $\alpha \in \mathbb{C}^n$  peut être identifié de façon unique à partir des mesures  $y = A\alpha \in \mathbb{C}^m$  seulement si*

$$m \geq 2s .$$

A partir de maintenant, on se focalisera sur la matrice  $A$ , et on essaiera d'identifier des conditions suffisantes assurant que le problème d'échantillonnage compressif puisse être résolu.

## 3.2 Propriété d'isométrie restreinte

La question est de savoir dans quelle condition les méthodes d'identification que nous avons vues dans le chapitre précédent, et en particulier BP, sont capables de résoudre le problème d'échantillonnage compressif.

Une propriété centrale dans la théorie de l'échantillonnage compressif est la **propriété d'isométrie restreinte** (RIP), définie de la façon suivante.

**Définition 7** *Soit  $A \in \mathcal{M}_{m,n}$ .  $A$  satisfait la propriété d'isométrie restreinte d'ordre  $s$  si il existe une constante  $\delta_s \in ]0, 1[$  telle que pour tout  $\alpha \in \Sigma_s$ , on ait l'encadrement*

$$(1 - \delta_s)\|\alpha\|_2^2 \leq \|A\alpha\|_2^2 \leq (1 + \delta_s)\|\alpha\|_2^2 .$$

En d'autres termes, la propriété RIP d'ordre  $s$  avec une valeur de  $\delta$  proche de 1 signifie  $A$  préserve approximativement la norme de tout vecteur  $s$ -parcimonieux.

On va montrer que la propriété RIP est une condition suffisante pour permettre de résoudre le problème de l'échantillonnage compressif. Avant d'en arriver là, il est utile d'introduire une version forte de la propriété (NSPq).

**Définition 8** *Une matrice  $A \in \mathcal{M}_{m,n}$  satisfait la propriété forte du noyau (SNSPq;p) d'ordre  $s$  par rapport aux normes  $\ell^q$  et  $\ell^p$  si pour tout  $\alpha \in \ker A$ ,  $\alpha \neq 0$  et tout  $\Lambda_0 \subset \{1, \dots, n\}$  tel que  $|\Lambda_0| \leq s$ , on a l'inégalité*

$$\|\alpha_{\Lambda_0}\|_p < \frac{1}{s^{1/q-1/p}} \|\alpha_{\Lambda_0}\|_q .$$

Le résultat suivant est un résultat classique de dimension finie.

**Lemme 9** *Si  $0 < q \leq p \leq \infty$ , on a*

$$\|\alpha\|_p \leq \|\alpha\|_q \leq K^{1/p-1/q} \|\alpha\|_p, \forall \alpha \in \mathbb{C}^K .$$

**Corollaire 9** Si  $A$  satisfait la propriété (SNSP $_q;p$ ) d'ordre  $s$ , alors résoudre le problème (P $_q$ ) identifie toute solution de l'équation  $x = A\alpha$  avec  $\alpha \in \Sigma_s$ .

Preuve : On sait que pour tout  $\alpha \in \ker(A) \setminus \{0\}$  et tout  $\Lambda_0$  tel que  $|\Lambda_0| \leq s$ , on a

$$\|\alpha_{\Lambda_0}\|_q \leq s^{1/p-1/q} \|\alpha_{\Lambda_0}\|_p .$$

Donc si de plus (SNSP $_q;p$ ) est satisfaite, on a bien (NSP $_q$ )

$$\|\alpha_{\Lambda_0}\|_q < \|\alpha_{\overline{\Lambda_0}}\|_q ,$$

qui entraîne le résultat d'identifiabilité. ♠

Cette préparation nous permet de démontrer le résultat principal de cette section.

**Théorème 13** Soit  $A \in \mathcal{M}_{m,n}$  une matrice satisfaisant la propriété d'isométrie restreinte d'ordre  $2s$ , et telle que

$$\delta_{2s} < \frac{1}{3} .$$

Alors  $A$  vérifie la propriété (SNSP $1;2$ ) à l'ordre  $s$  :

$$\forall \alpha \in \ker(A) \setminus \{0\} , \quad \forall |\Lambda_0| \leq s , \quad \|\alpha_{\Lambda_0}\|_2 < \frac{1}{2\sqrt{s}} \|\alpha\|_1 .$$

Preuve : On se place pour simplifier dans le cas de  $\mathbb{R}^n$ . L'extension au cas complexe est possible. Supposons satisfaite la propriété (RIP) d'ordre  $2s$ . Soit  $\alpha \in \ker(A)$ , avec  $\alpha \neq 0$ . Notons tout d'abord qu'il suffit de montrer la propriété (SNSP $1;2$ ) pour  $\Lambda_0$  constitué des  $s$  plus grandes composantes de  $\alpha$  (en valeur absolue), la propriété pour  $\Lambda_0$  arbitraire en découle ensuite.

On considère une partition du complémentaire  $\overline{\Lambda_0}$  de la forme

$$\overline{\Lambda_0} = \Lambda_1 \cup \Lambda_2 \cup \Lambda_3 \cup \dots ,$$

où

- $\Lambda_1$  est le sous-ensemble de  $\overline{\Lambda_0}$  contenant les  $s$  plus grandes composantes de  $\alpha_{\overline{\Lambda_0}}$
- $\Lambda_2$  est le sous-ensemble de  $\overline{\Lambda_0}$  contenant les  $s$  plus grandes composantes suivantes de  $\alpha_{\overline{\Lambda_0}}$
- ...

Comme  $\alpha \in \ker(A)$ , on a

$$A\alpha_{\Lambda_0} = - \sum_{k \geq 1} A\alpha_{\Lambda_k} ,$$

et la propriété (RIP) implique

$$\|\alpha_{\Lambda_0}\|_2^2 \leq \frac{1}{1 - \delta_{2s}} \|A\alpha_{\Lambda_0}\|_2^2 = \frac{1}{1 - \delta_{2s}} \sum_{k \geq 1} \langle A\alpha_{\Lambda_0}, -A\alpha_{\Lambda_k} \rangle .$$

Introduisons maintenant les vecteurs unitaires  $u_k$  tels que  $\alpha_{\Lambda_0} = \|\alpha_{\Lambda_0}\|_2 u_0$  et  $\alpha_{\Lambda_k} = -\|\alpha_{\Lambda_k}\|_2 u_k$ . On peut donc écrire, grâce à l'identité de polarisation (voir exercice 3 ci-dessous)

$$\begin{aligned} \langle A\alpha_{\Lambda_0}, -A\alpha_{\Lambda_k} \rangle &= \|\alpha_{\Lambda_0}\|_2 \cdot \|\alpha_{\Lambda_k}\|_2 \cdot \langle Au_0, Au_k \rangle \\ &= \frac{1}{4} \|\alpha_{\Lambda_0}\|_2 \cdot \|\alpha_{\Lambda_k}\|_2 \cdot [\|A(u_0 + u_k)\|_2^2 - \|A(u_0 - u_k)\|_2^2] \\ &\leq \frac{1}{4} \|\alpha_{\Lambda_0}\|_2 \cdot \|\alpha_{\Lambda_k}\|_2 \cdot [(1 + \delta_{2s})\|u_0 + u_k\|_2^2 - (1 - \delta_{2s})\|u_0 - u_k\|_2^2] \\ &\leq \frac{1}{4} \|\alpha_{\Lambda_0}\|_2 \cdot \|\alpha_{\Lambda_k}\|_2 \cdot 4\delta_{2s} , \end{aligned}$$

on a donc

$$\|\alpha_{\Lambda_0}\|_2 \leq \frac{\delta_{2s}}{1 - \delta_{2s}} \sum_{k \geq 1} \|\alpha_{\Lambda_k}\|_2 .$$

Nous n'avons pas encore utilisé la définition des  $\Lambda_k$ , il est temps... On remarque que si  $i \in \Lambda_k$  et  $j \in \Lambda_{k-1}$ , on a

$$|\alpha_i| \leq \alpha_j \quad \Longrightarrow \quad |\alpha_i| \leq \frac{1}{s} \|\alpha_{\Lambda_{k-1}}\|_1 \quad \Longrightarrow \quad \|\alpha_{\Lambda_k}\|_2^2 \leq \frac{1}{s} \|\alpha_{\Lambda_{k-1}}\|_1^2 ,$$

et donc

$$\sum_{k \geq 1} \|\alpha_{\Lambda_k}\|_2 \leq \frac{1}{\sqrt{s}} \sum_{k \geq 0} \|\alpha_{\Lambda_k}\|_1 = \frac{1}{\sqrt{s}} \|\alpha\|_1 .$$

On a donc montré que

$$\|\alpha_{\Lambda_0}\|_2 \leq \frac{\delta_{2s}}{1 - \delta_{2s}} \frac{1}{\sqrt{s}} \|\alpha\|_1 .$$

Donc si  $\delta_{2s} < 1/3$ ,  $\delta_{2s}/(1 - \delta_{2s}) < 1/2$  et on a bien vérifié la propriété (SNSP1;2). ♠

### 3.3 Matrices aléatoires

Un problème majeur de la théorie de l'échantillonnage compressif tient à la difficulté de construire pratiquement des matrices satisfaisant la propriété d'isométrie restreinte, et de vérifier celle-ci. En fait, on ne connaît pas de construction déterministe générique produisant de bonnes matrices de mesure. Par contre, comme on va le voir, des matrices aléatoires dans une classe bien choisie vérifient avec très grande probabilité la propriété RIP. Les résultats ci-dessous s'appuient fortement sur des techniques développées dans le cadre du lemme de Johnson-Lindenstrauss [15], qui démontre l'existence de projections d'espaces Euclidiens de grande dimension sur des espaces de dimension plus faible préservant (à une petite erreur près) les distances entre un certain nombre de points si celui-ci n'est pas trop élevé. Le lien avec l'échantillonnage compressif a été éclairci dans [1], et la démonstration ci-dessous est fortement inspirée de cet article.

#### 3.3.1 RIP et matrices aléatoires

**Théorème 14** *Soit  $A \in \mathcal{M}_{m,n}$  une matrice aléatoire, telle que sa distribution de probabilités satisfait pour tout  $\alpha \in \mathbb{R}^n$ , l'inégalité de concentration*

$$\mathbb{P}\{ \left| \|A\alpha\|_2^2 - \|\alpha\|_2^2 \right| > \epsilon \|\alpha\|_2^2 \} \leq 2e^{-C(\epsilon)m} , \quad \forall \epsilon \in ]0, 1[ ,$$

où  $C(\epsilon)$  est une constante positive ne dépendant que de  $\epsilon$ . Alors pour tout  $\delta \in ]0, 1[$  il existe des constantes  $C_0(\delta)$  et  $C_1(\delta)$  telles que

$$\mathbb{P}\{ \exists \alpha \in \Sigma_s : \left| \|A\alpha\|_2^2 - \|\alpha\|_2^2 \right| > \delta \|\alpha\|_2^2 \} \leq 2e^{-C_0(\delta)m} ,$$

dès que

$$m \geq sC_1(\delta) \ln(en/s) .$$

Ce résultat montre en particulier que (à un logarithme près), le nombre de mesures à effectuer est proportionnel au nombre de coefficients à identifier.

La preuve du théorème repose sur le résultat suivant :

**Lemme 10** Dans  $\mathbb{R}^n$ , muni d'une norme quelconque  $\|\cdot\|$ , soit  $\mathcal{S} = \{z \in \mathcal{E}, \|z\| = 1\}$  la sphère unité. Soit  $\delta > 0$ . Il existe un sous-ensemble  $\mathcal{U} = \{u_1, \dots, u_k\}$  de  $\mathcal{S}$  tel que pour tout  $x \in \mathcal{S}$ ,

$$\min_{u \in \mathcal{U}} \|x - u\| \leq \delta ,$$

et que la cardinalité de  $\mathcal{U}$  peut être contrôlée comme

$$|\mathcal{U}| \leq \left(1 + \frac{2}{\delta}\right)^n .$$

**Preuve du lemme :** Soient  $u_1, \dots, u_I \in \mathcal{S}$  une famille maximale telle que  $\|u_i - u_j\| > \delta$  pour tous  $i, j = 1, \dots, I$  (c'est à dire que lui ajouter un vecteur, quel qu'il soit, viole cette propriété). Alors pour tout  $\beta \in \mathcal{S}$ , on a

$$\min_{i=1, \dots, I} \|u_i - \beta\| \leq \delta .$$

Soit  $\mathcal{B}$  la boule unité par rapport à la norme considérée. Remarquons tout d'abord que les ensembles  $u_i + \frac{\delta}{2}\mathcal{B}$  sont deux à deux disjoints. En effet, soit  $v \in (u_i + \frac{\delta}{2}\mathcal{B}) \cap (u_j + \frac{\delta}{2}\mathcal{B})$ ; on a alors

$$\|u_i - u_j\| \leq \|u_i - v\| + \|u_j - v\| \leq \delta ,$$

d'où contradiction.

Notons aussi que pour tout  $v \in \mathcal{B}$ , on a pour tout  $i = 1, \dots, I$

$$\left\|u_i + \frac{\delta}{2}v\right\| \leq \|u_i\| + \frac{\delta}{2}\|v\| + \frac{\delta}{2} ,$$

d'où

$$u_i + \frac{\delta}{2}\mathcal{B} \subset \mathcal{B} + \frac{\delta}{2}\mathcal{B} , \quad \forall i = 1, \dots, I .$$

Ainsi, on a

$$\sum_{i=1}^I \text{Vol} \left(u_i + \frac{\delta}{2}\mathcal{B}\right) = I \left(\frac{\delta}{2}\right)^n \leq \text{Vol} \left(\left(1 + \frac{\delta}{2}\right)\mathcal{B}\right) = \left(1 + \frac{\delta}{2}\right)^n ,$$

d'où on déduit finalement

$$I \leq \left(1 + \frac{2}{\delta}\right)^n ,$$

qui conclut la démonstration. ♠

**Preuve du théorème :** soit  $A \in \mathcal{M}_{m,n}$ , on notera  $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$  ses colonnes. Supposons l'inégalité de concentration satisfaite. Soit  $s$  un entier positif fixé, soit  $\Lambda_0$  un sous-ensemble de  $\Lambda = \{1, \dots, n\}$ , de cardinalité  $|\Lambda_0| = s$ , et soit  $\mathcal{E}_{\Lambda_0}$  l'espace linéaire engendré par les vecteurs  $\mathbf{a}_\lambda$ ,  $\lambda \in \Lambda_0$  et  $\mathcal{S}_{\Lambda_0}$  la sphère unité dans  $\mathcal{E}_{\Lambda_0}$

$$\mathcal{E}_{\Lambda_0} = \left\{ \sum_{\lambda \in \Lambda_0} \alpha[\lambda] \mathbf{a}_\lambda , \alpha \in \mathbb{R}^s \right\} , \quad \mathcal{S}_{\Lambda_0} = \{u \in \mathcal{E}_{\Lambda_0} , \|u\|_2 = 1\} .$$

D'après le lemme ci-dessus, on sait qu'il existe un sous-ensemble  $\mathcal{U}$  de  $\mathcal{S}_{\Lambda_0}$  tel que pour tout  $z \in \mathcal{S}_{\Lambda_0}$ , vérifiant

$$\min_{u \in \mathcal{U}} \|z - u\|_2 \leq \frac{\delta}{\delta + 4} ,$$

et que

$$|\mathcal{U}| \leq \left(1 + \frac{2}{\delta/(\delta + 4)}\right)^s = \left(3 + \frac{8}{\delta}\right)^s .$$

Il résulte de l'inégalité de concentration que

$$\mathbb{P}\left\{\exists u \in \mathcal{U} : \left| \|Au\|_2^2 - \|u\|_2^2 \right| \geq \frac{\delta}{4} \|u\|_2^2 \right\} \leq 2 \left(3 + \frac{8}{\delta}\right)^s \exp\{-C(\delta/4)m\} .$$

Montrons tout d'abord que si pour tout  $u \in \mathcal{U}$  on a  $\left| \|Au\|_2^2 - \|u\|_2^2 \right| \leq (\delta/4) \|u\|_2^2$  alors tout  $\alpha \in \mathcal{E}_{\Lambda_0}$  satisfait l'encadrement (RIP) avec constante  $\delta$ . Supposons donc que

$$\forall u \in \mathcal{U}, \quad \left| \|Au\|_2^2 - \|u\|_2^2 \right| \leq \frac{\delta}{4} \|u\|_2^2 .$$

Soit  $\delta'$  le plus petit réel positif tel que pour tout  $\alpha \in \mathcal{E}_{\Lambda_0}$ , on ait  $\|A\alpha\|_2^2 \leq (1 + \delta') \|\alpha\|_2^2$ . Nous allons tout d'abord montrer que  $\delta' \leq \delta$ .

Soit  $\alpha \in \mathcal{S}_{\Lambda_0}$ , et soit  $u \in \mathcal{U}$  tel que  $\|u - \alpha\|_2 \leq \delta/(\delta + 4)$ . On a alors

$$\|A\alpha\|_2 \leq \|A(\alpha - u)\|_2 + \|Au\|_2 \leq \sqrt{1 + \delta'} \|\alpha - u\|_2 + \|u\|_2 \leq \sqrt{1 + \delta'} \frac{\delta}{\delta + 4} + \sqrt{1 + \delta/4} .$$

Par définition de  $\delta'$ , on a

$$\sqrt{1 + \delta'} \leq \sqrt{1 + \delta'} \frac{\delta}{\delta + 4} + \sqrt{1 + \delta/4} ,$$

d'où on déduit

$$1 + \delta/4 \geq (1 + \delta') \left(1 + \frac{\delta}{4 + \delta}\right)^2 ,$$

et

$$1 + \delta' \leq \left(1 + \frac{\delta}{4}\right)^3 \leq 1 + \delta ,$$

d'où  $\delta' \leq \delta$ . On a donc, pour tout  $\alpha \in \mathcal{E}_{\Lambda_0}$ ,

$$\|A\alpha\|_2^2 - \|\alpha\|_2^2 \leq \delta \|\alpha\|_2^2 .$$

Reste maintenant à montrer l'autre inégalité  $\|Au\|_2^2 - \|u\|_2^2 \geq -\delta \|u\|_2^2/4$ . Avec les mêmes notations, on a

$$\begin{aligned} \|A\alpha\|_2^2 &\geq (\|Au\|_2 - \|A(\alpha - u)\|_2)^2 \geq \|Au\|_2^2 - 2\|Au\|_2 \|A(\alpha - u)\|_2 \\ &\geq (1 - \delta/4) - 2\sqrt{1 + \delta'} \sqrt{1 - \delta/4} \frac{\delta}{4 + \delta} \\ &\geq (1 - \delta/4) - 2(1 + \delta/2) \sqrt{1 - \delta/4} \frac{\delta}{4 + \delta} \\ &\geq 1 - \delta/4 - \delta/2(1 + \delta/2) \geq 1 - \delta . \end{aligned}$$

Nous avons donc montré que l'encadrement RIP pour tout  $u \in \mathcal{U}$ , avec constante d'isométrie restreinte  $\delta/4$  implique l'encadrement RIP pour tout  $\alpha \in \mathcal{E}_{\Lambda_0}$ , avec constante d'isométrie restreinte  $\delta$ .

Finalement, rappelons que  $\Sigma_s$  est l'union de  $C_n^s$  sous-espaces de la forme  $\mathcal{E}_{\Lambda_0}$ . En notant que

$$C_n^s \leq \left(\frac{en}{s}\right)^s ,$$

on en déduit

$$\begin{aligned} \mathbb{P}\{\exists \alpha \in \Sigma_s : \left| \|A\alpha\|_2^2 - \|\alpha\|_2^2 \right| \geq \delta \|\alpha\|_2^2\} &\leq 2 \left(\frac{en}{s}\right)^s \left(3 + \frac{8}{\delta}\right)^s \exp\{-C(\delta/4)m\} \\ &\leq 2 \exp\{s [\ln(en/s) + \ln(3 + 8/\delta)] - C(\delta/4)m\} . \end{aligned}$$

On aura donc

$$\mathbb{P}\{\exists \alpha \in \Sigma_s : \left| \|A\alpha\|_2^2 - \|\alpha\|_2^2 \right| \geq \delta \|\alpha\|_2^2\} \leq 2e^{-C(\delta/4)m/2}$$

dès que

$$s \left[ \ln \left( \frac{en}{s} \right) + \ln \left( 3 + \frac{8}{\delta} \right) \right] \leq \frac{1}{2} C(\delta/4) m ,$$

qui est assuré dès que

$$m \geq s \cdot \ln \left( \frac{en}{s} \right) \cdot C_1(\delta) ,$$

où on a pris

$$C_1(\delta) = \frac{2(1 + \ln(3 + 8/\delta))}{C(\delta/4)} .$$



### 3.3.2 Matrices Gaussiennes

Le cas le plus simple de matrice satisfaisant les inégalités de concentration exigées sont des matrices Gaussiennes i.i.d. Rappelons que si  $X$  est une variable aléatoire  $\mathcal{N}(0, \sigma^2)$  alors pour tout  $t \in \mathbb{R}$ , on a

$$\mathbb{E} \{ e^{tX} \} = e^{\sigma^2 t^2 / 2} .$$

De plus, on vérifie facilement que

$$\mathbb{E} \{ e^{tX^2} \} = \frac{1}{\sqrt{1 - 2t\sigma^2}} .$$

Soit  $A \in \mathcal{M}_{m,n}$  telle que ses coefficients de matrice  $a_{ij}$  soient des variables aléatoires  $\mathcal{N}(0, \sigma^2)$  indépendantes et identiquement distribuées. On note, pour  $\alpha \in \mathbb{R}^n$ ,

$$X_i = \sum_{j=1}^n a_{ij} \alpha_j , \quad i = 1, \dots, m ,$$

les  $X_i$  sont alors Gaussiennes  $\mathcal{N}(0, \sigma^2 \|\alpha\|^2)$ , et i.i.d. On pose pour simplifier  $\gamma = \sigma^2 \|\alpha\|^2$ , et

$$Y_i = X_i^2 - \mathbb{E} \{ X_i^2 \} = X_i^2 - \gamma ,$$

de sorte que la quantité qui nous intéresse s'écrit

$$\|A\alpha\|_2^2 - m\sigma^2 \|\alpha\|^2 = \sum_{i=1}^m Y_i ,$$

et nous allons chercher à borner

$$\mathbb{P} \left\{ \left| \|A\alpha\|_2^2 - m\sigma^2 \|\alpha\|^2 \right| \geq \epsilon m \sigma^2 \|\alpha\|^2 \right\} = \mathbb{P} \left\{ \sum_{i=1}^m Y_i > \epsilon m \gamma \right\} + \mathbb{P} \left\{ \sum_{i=1}^m Y_i < -\epsilon m \gamma \right\} .$$

Bornons tout d'abord le premier terme. Pour tout  $t \in \mathbb{R}^+$ ,

$$\mathbb{P} \left\{ \sum_{i=1}^m Y_i > \epsilon m \gamma \right\} = \mathbb{P} \left\{ e^{t \sum_{i=1}^m Y_i} > e^{t \epsilon m \gamma} \right\} \leq \frac{\mathbb{E} \left\{ e^{t \sum_{i=1}^m Y_i} \right\}}{e^{t \epsilon m \gamma}} = \prod_{i=1}^m \left[ \mathbb{E} \left\{ e^{t Y_i} \right\} e^{-t \epsilon \gamma} \right] ,$$

où on a utilisé l'inégalité de Markov, puis l'indépendance des  $Y_i$ . Calculons

$$p(t) = \mathbb{E} \left\{ e^{t Y_i} \right\} e^{-t \epsilon \gamma} = \mathbb{E} \left\{ e^{t X_i^2} \right\} e^{-t(1+\epsilon)\gamma} = \frac{1}{\sqrt{1 - 2\gamma t}} e^{-t(1+\epsilon)\gamma} .$$

La majoration étant valable pour tout  $t$ , cherchons la valeur de  $t$  qui minimise la borne. Le minimum de  $p(t)$  est obtenu pour

$$p'(t) = \gamma(1 - 2\gamma t)^{-3/2} - (1 + \epsilon)\gamma(1 - 2\gamma t)^{-1/2} = 0 ,$$

qui correspond à

$$1 - 2\gamma t_* = \frac{1}{1 + \epsilon} \quad \Longrightarrow \quad t_* = \frac{1}{2\gamma} \frac{\epsilon}{1 + \epsilon} .$$

On a alors

$$\ln(p(t_*)) = \frac{1}{2} \ln(1 + \epsilon) - \frac{\epsilon}{2} \leq \frac{1}{2} \left( \epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{3} \right) - \frac{\epsilon}{2} \leq -\frac{\epsilon^2}{4} + \frac{\epsilon^3}{6} .$$

On en déduit

$$\mathbb{P} \left\{ \sum_{i=1}^m Y_i > \epsilon m \gamma \right\} \leq e^{-m(\epsilon^2/4 - \epsilon^3/6)} .$$

Le second terme se traite de façon similaire : on calcule

$$\mathbb{P} \left\{ \sum_{i=1}^m Y_i < -\epsilon m \gamma \right\} = \mathbb{P} \left\{ e^{-t \sum_{i=1}^m Y_i} > e^{t \epsilon m \gamma} \right\} \leq \frac{\mathbb{E} \left\{ e^{-t \sum_{i=1}^m Y_i} \right\}}{e^{t \epsilon m \gamma}} = \prod_{i=1}^m [\mathbb{E} \{ e^{-t Y_i} \} e^{-t \epsilon \gamma}] ,$$

puis

$$q(t) = \mathbb{E} \{ e^{-t Y_i} \} e^{-t \epsilon \gamma} = \mathbb{E} \left\{ e^{-t X_i^2} \right\} e^{t(1-\epsilon)\gamma} = \frac{1}{\sqrt{1 + 2\gamma t}} e^{t(1-\epsilon)\gamma} .$$

Le même calcul que précédemment montre que la valeur de  $t$  qui minimise cette quantité est

$$1 + 2\gamma t_* = \frac{1}{1 - \epsilon} \quad \Longrightarrow \quad t_* = \frac{1}{2\gamma} \frac{\epsilon}{1 - \epsilon} ,$$

ce qui conduit à la borne

$$q(t_*) \leq e^{-\epsilon^2/4 - \epsilon^3/6} \leq e^{-(\epsilon^2/4 - \epsilon^3/6)} ,$$

et donc

$$\mathbb{P} \left\{ \sum_{i=1}^m Y_i < -\epsilon m \gamma \right\} \leq e^{-m(\epsilon^2/4 - \epsilon^3/6)} .$$

En prenant  $\sigma^2 = 1/m$ , nous avons donc montré :

**Théorème 15** Soit  $A \in \mathcal{M}_{m,n}$  une matrice réelle, dont les coefficients sont des variables aléatoires  $\mathcal{N}(0, 1/m)$  indépendantes identiquement distribuées. Alors  $A$  satisfait l'inégalité de concentration

$$\mathbb{P} \{ \left| \|A\alpha\|_2^2 - \|\alpha\|_2^2 \right| > \epsilon \|\alpha\|_2^2 \} \leq 2e^{-Cm(\epsilon^2/4 - \epsilon^3/6)} , \quad \forall \epsilon \in ]0, 1[ .$$

### 3.3.3 Une classe de matrices aléatoires répondant au signalement...

Comme on le voit dans le calcul ci-dessus, l'essentiel de l'argument est basé sur le contrôle des premiers moments des variables aléatoires  $X_i$ , et des quantités  $\mathbb{E} \left\{ e^{t X_i^2} \right\}$ . Il est possible de démontrer les mêmes estimations que ci-dessus pour une classe plus large de matrices, à savoir les matrices dont les coefficients sont des variables aléatoires strictement sous-gaussiennes.

**Définition 9** 1. Une variable aléatoire  $X$  est sous-Gaussienne si il existe  $a \in \mathbb{R}$  tel que

$$\mathbb{E} \left\{ e^{tX} \right\} \leq e^{a^2 t^2 / 2}$$

pour tout  $t \in \mathbb{R}$ .

2. En notant

$$\tau(X) = \inf \left\{ a \in \mathbb{R} : \mathbb{E} \{ e^{tX} \} \leq e^{a^2 t^2 / 2} \quad \forall t \in \mathbb{R} \right\} ,$$

$X$  est strictement sous-Gaussienne si

$$\mathbb{E} \{ X^2 \} = \tau(X)^2 .$$

**Lemme 11** 1. Si  $X$  est sous-Gaussienne, alors

$$\mathbb{E} \{ X \} = 0 , \quad \mathbb{E} \{ X^2 \} \leq \tau(X)^2 .$$

2. Si  $X$  est strictement sous-Gaussienne, alors

$$\mathbb{E} \{ X^3 \} = 0 , \quad \mathbb{E} \{ X^4 \} \leq 3 \mathbb{E} \{ X^2 \}^2 .$$

**Lemme 12** 1. Soient  $X_1, \dots, X_n$  des variables aléatoires sous-Gaussiennes indépendantes. Alors  $X_1 + X_2 + \dots + X_n$  est également sous-Gaussienne, et

$$\tau(X)^2 \leq \tau(X_1)^2 + \tau(X_2)^2 + \dots + \tau(X_n)^2 .$$

2. Si  $X_1, \dots, X_n$  sont des variables aléatoires strictement sous-Gaussiennes indépendantes, alors  $X_1 + X_2 + \dots + X_n$  est strictement sous-Gaussienne.

**Exemple 4** On dispose de nombreux exemples de variables aléatoires sous-Gaussiennes. Par exemple

1. Une variable aléatoire Gaussienne centrée est strictement sous-Gaussienne.
2. Une variante uniformément distribuée sur  $[-1, 1]$  est strictement sous-Gaussienne.

Les ingrédients ci-dessus sont suffisants pour adapter la preuve du théorème précédent au cas strictement sous-Gaussien.

**Théorème 16** Soit  $A \in \mathcal{M}_{m,n}$  une matrice réelle, dont les coefficients sont des variables aléatoires indépendantes identiquement distribuées strictement sous-gaussiennes, de variance  $\sigma^2 = 1/m$ . Alors  $A$  satisfait l'inégalité de concentration

$$\mathbb{P} \left\{ \left| \|A\alpha\|_2^2 - \|\alpha\|_2^2 \right| > \epsilon \|\alpha\|_2^2 \right\} \leq 2e^{-Cm(\epsilon^2/4 - \epsilon^3/6)} , \quad \forall \epsilon \in ]0, 1[ .$$

Les détails de la démonstration peuvent être trouvés dans [14].

### 3.4 Une conclusion... temporaire

Nous avons vu dans ces pages comment la notion de parcimonie a fait évoluer la notion d'échantillonnage, depuis ses origines (calcul de valeurs ponctuelles d'un signal) jusqu'à l'échantillonnage compressif, qui revient à essayer d'effectuer un nombre le plus petit possible de mesures sur le signal.

Les difficultés inhérentes à l'échantillonnage compressif tiennent à la difficulté de vérifier les différentes propriétés qui assurent l'unicité d'une décomposition parcimonieuse. En grandes dimensions, les propriétés liées à la cohérence donnent des bornes assez éloignées de l'optimum, alors que les propriétés de type RIP sont invérifiables, car elles demandent d'analyser un trop grand nombre d'ensembles  $\Sigma_s$  (qui sont des unions de sous-espaces). Le côté positif est lié aux derniers résultats discutés, qui montrent en fait que les matrices d'acquisition vérifiant la propriété RIP sont en fait une majorité écrasante : une matrice tirée au hasard avec une distribution assez standard vérifie la condition, avec une énorme probabilité... mais on ne peut pas le vérifier... Beaucoup de travaux à l'heure actuelle portent sur des tentatives de constructions déterministes (l'état de l'art à l'heure actuelle est une construction de J. Bourgain et ses collaborateurs [2]), mais celles-ci sont encore très loin du compte.

Il est important de signaler que l'échantillonnage compressif n'est pas la seule alternative à l'échantillonnage classique, et il faut en particulier mentionner l'échantillonnage à taux d'innovation fini [22], qui repose sur des idées très différentes, ainsi que différentes techniques d'échantillonnage irrégulier (voir par exemple [13] et les références incluses).

## Exercices

1. Démontrer le Corollaire 8
2. Démontrer le Lemme 9.
3. Démontrer l'identité de polarisation dans le cas d'un espace Hilbertien réel :

$$\langle u, v \rangle = \frac{1}{4} [\|u + v\|^2 - \|u - v\|^2] ,$$

et dans le cas d'un espace de Hilbert complexe

$$\langle u, v \rangle = \frac{1}{4} [\|u + v\|^2 - \|u - v\|^2 + i\|u + iv\|^2 - i\|u - iv\|^2] .$$

4. Adapter la preuve du Théorème 13 au cas d'un espace Hilbertien complexe.
5. Démontrer un analogue du théorème 13 en utilisant (SNSP<sub>q,2</sub>) à la place de (SNSP<sub>1,2</sub>).
6. Démontrer le Lemme 11 ; on pourra utiliser le développement en série entière de l'exponentielle.  
Démontrer le Lemme 12.

# Bibliographie

- [1] R. G. BARANIUK, M. A. DAVENPORT, R. DEVORE et M. WAKIN : A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28:189–206, 2008.
- [2] J. BOURGAIN, S. DILWORTH, K. FORD, S. KONYAGIN et D. KUTZAROVA : Explicit constructions of rip matrices and related problems. *Duke Mathematical Journal*, à paraître, 2011.
- [3] E. CANDÈS et M. WAKIN : Introduction to compressive sampling. *IEEE Signal Processing Magazine*, March:21–30, 2008.
- [4] A. COHEN : *Numerical analysis of wavelet methods*. Studies in mathematics and its applications. Elsevier, Amsterdam, 2003.
- [5] A. COHEN, R. DEVORE et W. DAHMEN : Compressed sensing and best k-term approximation. *Journal of the American Mathematical Society*, 22:211–231, 2009.
- [6] R. R. COIFMAN et Y. MEYER : Remarques sur l’analyse de Fourier à fenêtre. *C. R. Acad. Sci. Paris Sér. I Math.*, 312(3):259–261, 1991.
- [7] I. DAUBECHIES, M. DEFRISE et C. DE MOL : An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications in Pure and Applied Mathematics*, 57(11):1413–1457, 2004.
- [8] D. L. DONOHO et M. ELAD : Optimally sparse representation in general (nonorthogonal) dictionaries via  $l_1$  minimization. *the Proceedings of the National Academy of Sciences*, 100:2197–2202, March 2003.
- [9] D. L. DONOHO et X. HUO : Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47:2845–2862, 2001.
- [10] D. L. DONOHO et P. B. STARK : Uncertainty principles and signal recovery. *SIAM Journal of Applied Mathematics*, 49(3):906–931, June 1989.
- [11] M. F. DUARTE, M. A. DAVENPORT, D. TAKHAR, J. N. LASKA, T. SUN, K. K. F. et R. G. BARANIUK : Single pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 2008. <http://dsp.rice.edu/cscamera>.
- [12] M. ELAD et A. BRUCKSTEIN : A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Transactions On Information Theory*, 48:2558–2567, 2002.
- [13] H. G. FEICHTINGER et K. GRÖCHENIG : Theory and practice of irregular sampling. *In Wavelets : mathematics and applications*, Stud. Adv. Math., p. 305–363. CRC, Boca Raton, FL, 1994.
- [14] S. FOUCART : Lectures on compressed sensing, 2009. Graduate course in the Department of Mathematics at Vanderbilt, [www.math.vanderbilt.edu/~foucart/TeachingFiles/S09/CSNotes.pdf](http://www.math.vanderbilt.edu/~foucart/TeachingFiles/S09/CSNotes.pdf).
- [15] W. JOHNSON et J. LINDENSTRAUSS : Extensions of lipschitz mappings into a hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.
- [16] S. MALLAT : *A wavelet tour of signal processing, the sparse way*. Academic Press, Burlington, MA, USA, 2009.
- [17] S. MALLAT et Z. ZHANG : Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41:3397–3415, 1993.
- [18] Y. MEYER : *Oscillating patterns in image processing and nonlinear evolution equations*, vol. 22 de *University Lecture Series*. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.

- [19] Y. MEYER : Les mathématiques embarquées dans la mission herschel, 2010. <http://www.cmla.ens-cachan.fr/fileadmin/Groupes/Seminaire/slides20092010/Mission.pdf>.
- [20] C. SOIZE : *Méthodes Mathématiques en Analyse du Signal*, vol. 553 de *Enseignement de la Physique*. Masson, Paris, 1993.
- [21] M. VETTERLI et J. KOVACEVIC : *Wavelets and subband coding*. Prentice Hall, Englewood Cliffs, NJ, USA, 1995.
- [22] M. VETTERLI, P. MARZILIANO et T. BLU : Sampling signals with finite rate of innovation. *IEEE Transactions on Signal Processing*, 50(6):1417–1428, 2002.
- [23] M. V. WICKERHAUSER : *Adapted Wavelet Analysis from Theory to Software*. AK Peters, Boston, MA, USA, 1994.

# Index

- Acquisition, 7, 41
- Application coefficient, 11
- Approximation adaptative, 17
- Approximation linéaire, 18
- Approximation non-linéaire, 18
  
- Base Hilbertienne, 11
- Base canonique, 10
- Base de Karhunen-Loève, 14
- Base de Kronecker, 10
- Bases d'ondelettes, 17
- Bases MDCT, 12
- Basis pursuit, 23
- Basis pursuit denoising, 27
  
- Codage part transformation, 12
- Codeur par transformation, 11
- Coefficients de Fourier, 8
- Compressed sensing, 41
- Conversion analogique-numérique, 7
  
- Décomposition biorthogonale, 14
- Dictionnaire, 22
  
- Echantillonnage, 7
- Echantillonnage à taux d'innovation fini, 49
- Echantillonnage compressif, 41
- Echantillonnage irrégulier, 49
- Espace de Paley-Wiener, 9
  
- Fenêtre, 12
- Filtrage passe-bas, 10
- Fonction de transfert, 10
- Fonction semi-définie positive, 13
- Formule de Parseval, 11
- Fréquence d'échantillonnage, 7
  
- Imagette, 15
- Inégalité de Bessel, 11
- Inégalité d'Elad-Bruckstein, 33
- Inégalité d'Heisenberg, 28
- Inégalité de concentration, 44
- Inégalité de Donoho-Elad, 35
- Inégalité de Donoho-Huo, 31
  
- Inégalité de Donoho-Stark, 29
- Instance optimalité, 38
- Itération de Landweber, 28
  
- JPEG, 15
  
- LASSO, 27
- Lemme de Johnson-Lindenstrauss, 44
  
- Méthode des repères, 23
- Matching pursuit, 24
- Matching pursuit orthogonal, 26
- Matrice de Gram, 26
- MP3, 12, 15
  
- Noyau de Dirichlet, 8
- NSPq, 36
  
- Ondelettes, 17
- Opérateur d'analyse, 22
- Opérateur de covariance, 13
- Opérateur de repère, 22
- Opérateur de synthèse, 22
  
- Parcimonie, 17, 18
- Pas d'échantillonnage, 7
- Principe d'incertitude, 28
- Propriété d'isométrie restreinte, 42
- Propriété du noyau, 36
  
- Quantification, 13
- Quasi-norme, 5
  
- Rang de Kruskal, 35
- Repère, 21
- Repliement de spectre, 10
- RIP, 42
  
- Séquence d'Alltop, 31
- Sensing, 41
- Seuillage dur, 27
- Seuillage doux, 27
- Signal compressible, 18
- Signal aléatoire centré, 13

## *Index*

- Signal aléatoire continu en moyenne quadratique, 13
- Signal aléatoire cycliquement stationnaire, 15
- Signal aléatoire du second ordre, 13
- Signal clôture, 31
- Signaux à bande limitée, 9
- SNSP $q;p$ , 42
- Sous-échantillonnage, 10
- Spark, 35
- Structure oblique, 21
- Système orthonormal, 11
  
- Taux, 13
- TFD, 5
- TFF, 5
- Théorème de Riesz-Fischer, 11
- Théorème d'Eckart-Young, 15
- Théorème de Mercer, 14
- Trame, 21
- Transformation de Fourier discrète, 5
- Transformation de Fourier finie, 5
- Transformation intégrale de Fourier, 6
  
- Variable aléatoire sous-Gaussienne, 48