

# **Echantillonnage continu, discret et approche de Candès et Tao**

Cuvillier Thomas

Sous la direction de Yves Meyer

Stage de Licence 3  
Ecole Normale Supérieure de Cachan  
94235 Cachan - FRANCE

24 Janvier-24 Juin

## **Introduction.**

Je me suis intéressé, au cours de ce stage, à divers textes tous liés les uns aux autres, en rapport avec le même sujet, mais interférant avec d'autres théories. Ce rapport aura pour but d'en montrer une partie, de rendre clair les liens et problématiques qui les lient.

Le sujet de base est l'échantillonnage de fonctions, continues ou discrètes, dont la transformée de Fourier est de support un ensemble de mesure finie ( compact dans le cas continu, fini dans le cas discret). Après avoir explicité la solution apportée au problème continu ( le quasicristal ), et ses avantages, on verra pourquoi il n'est pas possible d'en faire autant pour le cas discret.

On s'intéressera alors à l'approche de Candès et Tao du cas discret, et des théories qu'elle recoupe, notamment les épaisseurs de Gelfand, ainsi que les inégalités de " grandes déviations" . On explicitera deux de leurs théorèmes, dont le premier sera l'objet d'une analyse particulière, et d'un rapprochement (limité) avec un article de Bass et Gröchenig, concernant l'échantillonnage aléatoire dans le cas continu. Une dernière partie sera ensuite consacrée à une application en imagerie ( particulièrement médicale) de leur résultat, la tomographie.

# Table des matières

1	Définitions et propriétés de base	3
2	Résolution et analyse critique du cas continu	4
3	Cas discret	6
4	L'approche de Candès et Tao	8
5	Les épaisseurs Gerlfrand	9
6	Le second théorème de Candès et Tao, application avec Romberg	11
7	Premier théorème de Candès et Tao, les deux hypothèses	13
8	Idées de preuve de l'ERP et inégalités de grandes déviations	14
9	Idées de preuve de l'UUP, entropie, et ouverture	15
10	Application : La tomographie	17

# 1 Définitions et propriétés de base

## 1) Cas continu

On s'intéresse aux fonctions de carré intégrable, à support fréquentiel borné.

$$f \in L^2, \exists K \text{ compact, tel que } \text{supp}(\hat{f}) \subset K.$$

où  $\text{supp}$  représente le support

et où  $\hat{f}$  est la transformée de Fourier de  $f$ .

C'est pourquoi on appelle  $E_K$  l'ensemble des fonctions intégrables dont la transformée de Fourier est à support dans  $K$ .

DÉFINITION 1 (ECHANTILLONAGE STABLE)

On dit d'un ensemble  $\Lambda$  qu'il est d'échantillonnage stable ( stable sampling ) pour  $E_K$  si il existe  $A > 0, B > 0$ , tel que  $\forall f \in E_K$

$$A \sum_{i=1}^{\infty} |f(\lambda_i)|^2 \leq \int |f|^2 \leq B \sum_{i=1}^{\infty} |f(\lambda_i)|^2$$

Il se trouve qu'on peut trouver un ensemble d'échantillonnage qui soit valable pour tous les  $E_K$ ,  $K$  compact  $|K| < M$ .

## 2) Cas discret

On s'intéresse aux fonctions

$$P = \sum_{k \in F} c_k e^{2\pi i k x}$$

tq  $|F| \leq q$ , où  $|F| = \#F$

C'est l'équivalent discret du cas ci-dessus, on limite le domaine fréquentiel à  $q$  fréquences distinctes. Etant donné qu'on a trouvé un ensemble d'échantillonnage qui marchait pour les  $E_K$ , en ne prenant en compte que la mesure de  $K$ , notre première approche va être équivalente : on regarde s'il pourrait exister un ensemble d'échantillonnage stable pour les  $P_q$ , ensemble des fonctions qui sont somme de  $q$  fréquences.

On prouvera par la suite qu'il n'existe aucun ensemble vérifiant une telle propriété, et donc on s'intéressera plutôt à l'ensemble des

$$A_N = [f \in \mathbb{R}^{\mathbb{R}}, f = \sum_{i \in [1, \dots, N]} c_i \exp^{2\pi i x}].$$

C'est l'ensemble adapté pour modéliser une information numérique de type image, signal fini dans le temps (...). On a alors à faire à un espace vectoriel.

DÉFINITION 2 (ECHANTILLONAGE STABLE DISCRET)

Une partie finie  $F \subset [0, 1]$  est dite d'échantillonnage stable pour  $A_N$  si  $\forall f \in A_N$  on a

$$1/2 \sum_1^N |c_k|^2 \leq \sum_{x \in F} |P(x)|^2 \leq 3/2 \sum_1^N |c_k|^2$$

Les constantes  $1/2$  et  $3/2$  ont été choisies arbitrairement, et peuvent être remplacées par d'autres constantes positives.

Toutefois, on les gardera par la suite, car elles sont des valeurs de références dans les textes de recherche, celles-ci permettant un encadrement relativement proche.

Les ensembles d'échantillonnage sont importants du fait des deux propriétés suivantes.

#### THÉORÈME 1

*Une fonction est parfaitement définie par les valeurs qu'elle prend sur son ensemble d'échantillonnage.*

#### THÉORÈME 2 (COROLLAIRE)

*On peut reconstruire parfaitement la fonction à partir de sa donnée sur l'ensemble d'échantillonnage approprié.*

## 2 Résolution et analyse critique du cas continu

Dans le cas continu, on se rend compte qu'on peut trouver une solution qui ne dépend QUE de la mesure de  $K$ . C'est à dire, on peut trouver un ensemble  $\Lambda$ , tel que  $\forall K$  compact, tel que  $|K| < M$ ,  $\Lambda$  est un ensemble d'échantillonnage stable pour  $E_K$ . Ce  $\Lambda$  est un quasi-cristal. On a alors une non-uniformité entre la répartition de  $K$ , et donc des fréquences de  $f$ , et l'ensemble d'échantillonnage.

Ce manque d'uniformité se voit dès le théorème de Shannon généralisé, qui résoud un cas particulier.

### 1) Théorème de Shannon généralisé

Soit  $A$  une matrice inversible.

Soit  $\Lambda = A(\mathbb{Z}^n) \subset \mathbb{R}^n$  un réseau.

On appelle le réseau dual  $\Lambda^* = \{y; \exp(ix.y)=1, x \in \Lambda\}$

En fait on a  $\Lambda^* = 2\pi(A^*)^{-1}\mathbb{Z}^n$

#### THÉORÈME 3 (THÉORÈME DE SHANNON GÉNÉRALISÉ)

*Soit  $f \in L^2$  une fonction dont la transformée de Fourier est de support un ensemble  $E$ , ( $E$  Borélien), tel que  $\forall y \in \Lambda^*, y \neq 0$  on a  $E \cap (E + y) = \emptyset$ .*

*Alors soit  $\phi$  la transformée de Fourier inverse de la fonction indicatrice de  $E$ , on a*

$$f(x) = c. \sum_{\gamma \in \Lambda} f(\gamma) \phi(x - \gamma) \text{ où } c = |A([0, 1]^n)|$$

Ce théorème montre que  $\Lambda$  est un ensemble d'échantillonnage stable pour tous les  $E_K$ , tel que  $K \cap (K + y) = \emptyset$ , où  $y \in \Lambda^*, y \neq 0$ .

En particulier, soit  $K$  vérifiant cette propriété, alors  $\forall y \in \Lambda^*$ , on a  $K + y$  qui vérifie aussi la propriété.

On a donc

$$\sup[f \in K, \Lambda \text{ est un ensemble d'échantillonnage stable pour } K] = \infty$$

Ce théorème a l'avantage d'avoir une démonstration relativement simple, ce qui nous permet de la montrer ici.

### Démonstration

On note  $F$  la transformée de Fourier de  $f$ .

soit  $G = \sum_{\gamma \in \Lambda} F(x + \gamma)$

Cette somme existe bien car les  $F(x + \gamma)$  ont des supports distincts.

On a  $F = 1_E \cdot G$

On fait la transformation inverse pour avoir  $f$ . On note  $\phi_E$  la fonction transformée inverse de la fonction indicatrice de  $E$ , notée  $1_E$ .

$$f(x) = \sum_{\gamma \in \Lambda} \phi(x - \gamma) f(\gamma)$$

On remarque alors que c'est la fonction  $\phi$  qui va s'adapter à l'ensemble.

Si il y avait un  $y_0$  tel que  $E \cap (E + y_0) \neq \emptyset$  alors, en notant  $\Phi$  la fonction transformée inverse de  $1_{E \cap E + y_0}$ , en prenant

$$f = (e^{-i \cdot x \cdot y_0} - 1) \cdot \Phi(x)$$

on a la transformée de Fourier de  $f$  qui est supportée par  $E$ , et  $f$  s'annule en  $\Lambda$ , mais  $f$  différent de 0.

On ne peut donc pas reconstruire  $f$  à partir de sa donnée sur  $\Lambda$ .

## 2) Définitions et théorèmes finaux

### DÉFINITION 3 (DENSITÉ)

$\Lambda$  ensemble dénombrable de points tel que  $\exists \alpha > 0$

$$\|\lambda_i - \lambda_j\| > \alpha, \forall i \neq j$$

On définit  $\underline{\text{dens}}(\Lambda)$  (densité inférieure) par

$$\underline{\text{dens}}(\Lambda) = \lim_{R \rightarrow \infty} \inf_{x \in \mathbb{R}} \text{card}\{\Lambda \cap [x, x + R]\} / R$$

### THÉORÈME 4 (THÉORÈME DE LANDAU)

Si  $\Lambda$  est un ensemble d'échantillonnage stable pour  $E_K$  alors

$$\underline{\text{dens}}(\Lambda) \geq (2 \cdot \pi)^n \cdot |K|$$

### DÉFINITION 4 (ENSEMBLE DE DELONE)

Un ensemble  $\Lambda$  de  $\mathbb{R}^n$  est un ensemble de Delone si

$$\exists R_1 > 0, \exists R_2 > 0, \forall x \in \mathbb{R}, \text{Card}(B(x, R_1) \cap \Lambda) \geq 1 \text{ et } \text{Card}(B(x, R_2) \cap \Lambda) \leq 1$$

### DÉFINITION 5 (QUASICRISTAL)

Un quasicristal est un ensemble  $\Lambda$  de  $\mathbb{R}^n$  tel que

1)  $\Lambda$  est un ensemble de Delone

2)  $\exists F$  ensemble fini  $\subset \mathbb{R}^n$  tel que  $\Lambda - \Lambda \subset \Lambda + F$

Notre but n'est pas d'étudier les quasicristaux mais ceux-ci sont importants dans le cadre de cette théorie.

### THÉORÈME 5 (ENSEMBLE D'ÉCHANTILLONNAGE STABLE)

Soit  $\Lambda$ ,  $\Lambda$  quasicristal tel que

$$\text{dens}(\Lambda) \geq (2.\pi)^n.M$$

Alors  $\forall f$  tel que  $f \in E_K$ ,  $K$  compact,  $|K| < M$ ,  $\Lambda$  ensemble d'échantillonnage stable pour  $f$ .

### THÉORÈME 6 (COROLLAIRE)

Soit  $K = F + B$ ,

$B$  boule de centre 0

$F$  ensemble discret fini.

$$|K| = |F|.c_n r^n < M$$

Les fonctions dont le spectre est inclus dans  $K$  peuvent être échantillonnées sur  $\Lambda$ .

$\Rightarrow$  Il n'y a pas d'uniformité dans le cas continu.

## 3 Cas discret

C'est ce manque d'uniformité qui sera à l'origine de l'impossibilité de passer directement du cas continu au cas discret.

### 1) Cas discret général

Soit  $P_q$  l'ensemble des polynômes trigonométriques à  $q$  fréquences

### DÉFINITION 6 ( $P_q$ )

On note  $P_q$  l'ensemble des  $f$  tel que  $f \in P_q$ ,

$$f(x) = \sum_{k=1}^q c_k \exp(2.\pi.n_k.x)$$

où

$$0 < n_1 < \dots < n_q$$

On cherche une partie finie  $F \subset [0, 1]$  telle que

$$1/2 \sum_1^q |c_k|^2 \leq \sum_{x \in F} |P(x)|^2 \leq 3/2 \sum_1^q |c_k|^2$$

Si, comme dans le cas continu, on impose aucune limite à la fréquence maximale, on obtient une impossibilité. En effet, en utilisant le principe des " tiroirs de Dirichlet", on montre que :

$$\begin{aligned} \forall F \subset [0, 1], F \text{ fini}, \exists (u_n)_{n \in \mathbb{N}} \text{ tel que} \\ \forall f \in F, \lim_{n \rightarrow \infty} |1 - \exp^{2\pi \cdot f \cdot u_n}| = 0 \\ \text{Or pour } P_n(x) = 1 - \exp^{2\pi \cdot x \cdot u_n} \\ \text{on a } \sum_{k=1}^q |c_k|^2 = 2. \end{aligned}$$

On a donc

$\forall F \subset [0, 1], F \text{ fini}, F \text{ n'est pas un ensemble d'échantillonnage stable pour } P_q.$

Contrairement au cas continu, on a uniformité par rapport à la fréquence. Plus la fréquence maximale est élevée, plus il faut une résolution petite pour la détecter. Comme dans le cas du théorème de Shannon-Nyquist, qui, s'applique, à la base, au cas continu.

THÉORÈME 7 (SHANNON-NYQUIST)

Soit  $K$  un compact de  $\mathbb{R}^n$

Soit  $f \in E_K$

Soit  $M$  majorant de  $K$ .

Alors  $\mathbb{Z}^n \pi / M$  est un ensemble d'échantillonnage stable.

## 2) transition continu $\Rightarrow$ discret

Soit  $\phi$  la fonction transformée Fourier inverse de  $B(0, a)$ , où  $a < r$ . Alors on a un ensemble d'échantillonnage de

$$f = \sum_{k=1}^q (c_k \exp^{2i\pi \cdot n k x}) \phi(x)$$

en prenant un quasicristal  $\Lambda$  de densité

$$\underline{\text{dens}}(\Lambda) > (2\pi) \cdot q \pi \cdot r^2$$

Et ce,  $\forall a > 0, a < r$

Comme le quasicristal qui établit l'inégalité d'échantillonnage est indépendant de la taille de  $a$  ( pour  $a$  petit ), on pourrait penser qu'on puisse passer au cas  $a = 0$ , où on aurait alors

$$\text{supp}(\hat{f}) = [n_k, k = 1 \dots q]$$

Une tentative a été faite de ce côté, la voici.

## 3) Tentative d'adaptation du quasicristal

Il se trouve que, comme nous le verrons avec le premier théorème de Candès et Tao, le nombre de mesures nécessaires pour échantillonner  $f$ , est de l'ordre de  $C \cdot (\log N)^6$  (  $N$  étant la fréquence maximale ), ou, si il y a un faible nombre de fréquences  $q$ , on peut se limiter à  $C \cdot q \cdot \log(N)$ ,  $C$  étant une constante.

Soit  $F = [\sqrt{2}, 2\sqrt{2}, \dots, m\sqrt{2}]$  où  $m$  de l'ordre de  $C \cdot \log(N)^6$   
On l'adapte à  $[0, 1]$  en prenant :  
 $F' = [\sqrt{2} - 1, 2\sqrt{2} - E(2\sqrt{2}), \dots, m\sqrt{2} - E(m\sqrt{2})]$

où  $E$  est la fonction partie entière.

Echantillonner  $f$  sur  $F'$  revient à l'échantillonner sur le quasi cristal défini par  $\lambda_j = j + j\sqrt{2}$ ,  $j$  étant restreint à  $[1, \dots, m]$

Reste à voir si ce  $F'$  est bien un ensemble d'échantillonnage stable.

on utilise ce lemme :

Pour tout entier  $N$ , il existe un entier  $n_N$ , tel que  $1 \leq n_N \leq N$  et un entier  $u$  tels que

$$|n_N\sqrt{2} - u| \leq 1/N$$

Soit alors  $P(x) = \exp(2.\pi.n_N.x) - 1$ .

Alors on a  $\exp(2.\pi.n_N.j.\sqrt{2}) = 1 + \epsilon$ , où  $\epsilon \leq 1/N$ .

En l'élevant à la puissance  $j$  il vient :  $\exp(2.\pi.j.n_N.j.\sqrt{2}) = 1 + \epsilon_j$ , où  $\epsilon_j \leq j/N$ .

Et donc finalement

$$\sum i = 1^m |P(j.\sqrt{2})|^2 \leq m^3.N^{-2}$$

Ce qui tends vers 0 quand  $N$  tend vers l'infini car  $m^3 \leq C \log(N)^{18}$

#### 4) Problèmes

D'un point de vue du dénombrement, cela semble tout à fait normal. En effet, le quasicristal récupère une infinité (dénombrable) d'informations. Un ensemble  $F$  fini n'en récupérera que  $|F|$ .

D'autre part, dans le cas continu la fonction  $f \in L^2$ , lequel est séparable. L'utilisation du quasicristal n'est rien d'autre que l'exhibition d'une frame dans un espace séparable de dimension infinie. Dans le cas de  $P_q$  nous n'avons plus à faire à un espace vectoriel (nous perdons donc des propriétés), et nous ne cherchons qu'une somme finie, c'est à dire un nombre fini d'informations. La difficulté est donc accrue.

#### 5) Fréquences fixées

En effet, fixons les  $q$  fréquences. On a alors un espace vectoriel de dimension  $q$ . On voit facilement que en prenant  $q$  points quelconques de l'intervalle  $[0, 1]$ , distincts deux à deux, on peut retrouver les  $c_k, \forall k \in [1, \dots, q]$  (du moins mathématiquement ; l'algorithme utilisant une matrice avec des coefficients de type exponentiels, on peut vite se retrouver avec des problèmes de grandeurs de nombres dans les calculs informatiques).

## 4 L'approche de Candès et Tao

L'approche de Candès et Tao, dans le cadre de la théorie sur le compressed sensing, nécessite un certain nombre d'hypothèses en plus.

Premièrement, on ne considère plus les  $q$  fréquences réparties n'importe où dans  $[0, N]$ , mais ne peuvent prendre que des valeurs entières.



Si cette restriction peut sembler très forte, on remarquera qu'elle est tout de suite plus adaptée à la logique informatique, qui, sur un signal de longueur  $N$ , n'aura que des fréquences entières.

Deuxièmement, on peut aussi considérer (par simple rétraction fréquentiel), que cela équivaut à considérer les fréquences incluses dans  $[0, 1/N, 2/N, 3/N, \dots, 1]$ . Donc pour  $q'$  fréquences de  $[0, 1]$ , pour  $N$  grand (donc  $q \ll N$ ) on peut choisir  $q$  fréquences qui soit très proches des  $q'$  initiales.

Cette approche permet aussi d'exhiber un espace vectoriel. Celui des fonctions à  $N$  fréquences  $[0, 1, 2, \dots, N]$ .

Il y a deux approches différentes de Candès et Tao ensuite. La première consiste à considérer des fonctions à décroissance rapide.

En classant les  $c_k$ , les  $c_k$  étant les coefficients de  $\hat{f}$ , par ordre décroissant, on pose comme définition.

#### DÉFINITION 7

On dit que  $f \in \mathbb{R}^N$  appartient à la  $R - \text{weak} - l_p$  boule de rayon  $R$  si

$$c_k < R \cdot k^{-1/p}, \forall k \in [1, \dots, N]$$

Plus  $p$  est petit, plus la décroissance est rapide.

La seconde consiste à considérer que la fonction n'a que  $q$  fréquences,  $q \ll N$ . On en parlera dans la partie : le second théorème de Candès et Tao.

La décroissance rapide peut être vue comme une "extension" d'un nombre de coefficients  $q \ll N$ , car là, peu de coefficients auront une réelle importance.

En effet, elle implique

$$\# [n \in [1, \dots, N], c_n \geq \lambda] \leq (R/\lambda)^p$$

Candès et Tao recherchent des algorithmes permettant de retrouver  $f$  à partir d'un minimum d'informations, c'est à dire un échantillonnage minimal. Afin d'estimer ce qu'est le minimum requis, on utilise, dans le cas continu, le théorème de Landau, qui montre qu'un quasi-cristal est un ensemble minimal au sens de la densité. Dans le cas discret, on s'intéressera aux "Gelfand width", qui permettent d'estimer l'erreur minimale, de "combien" peut-on se rapprocher théoriquement de  $f$ , en prenant un nombre  $n$  d'échantillons bien choisis.

## 5 Les épaisseurs Gelfand

#### DÉFINITION 8 (EPAISSEUR DE GELFAND)

Soit  $X$  un espace de Banach

Soit  $K$  compact  $K \subset X$

On appelle et on note l'épaisseur de Gelfand

$$d^n(K)_X = \inf_{\text{codim}(Y)=n} \sup_{x \in K \cap Y} |x|_X$$

Les meilleurs  $Y$  dans cette définition sont ceux qui "traversent"  $K$  sur une "longueur" minimale.

Application aux espaces  $l_p$ .

Rappel : On est en dimension finie donc

$$||x||_{l_p} = (\sum_{j=1}^N |x_j|^p)^{1/p}$$

On note

$$U(L_p^N) = [x \in L_p^N, ||x||_{L_p} \leq 1] \text{ C'est la boule unit  dans } L_p^N$$

Alors on a ce r sultat

TH OR ME 8

$$C_0^{-1} \sqrt{(\log(N/n) + 1)/n} \leq d^n(U(L_1^N))_{l_2^N} \leq C_0 \sqrt{(\log(N/n) + 1)/n}$$

En termes non math matiques, cela donne.  $\Rightarrow$  l' paisseur minimale de la boule  $L_1$  en dimension  $N$ , par rapport   un sous espace de codimension  $n$  varie en  $\sqrt{\log(N/n)/n}$ . Les  paisseurs de Gelfrand permettent d'estimer les performances d'un algorithme, en les comparant aux performances th oriques maximales.

Ainsi, en algorithmique, on utilise une paire encodeur / d codeur.

Soit  $x$  un signal. Un encodeur  $\Phi$  est une matrice  $(N, n)$  qui va extraire des informations ( $n$  informations) de  $x$ .

$$y = \Phi x$$

Le d codeur est un algorithme  $\Delta$  de  $\mathbb{R}^n$  dans  $\mathbb{R}^N$  qui va chercher   "retrouver"  $x$ ,   partir de  $y$

c'est   dire on cherche  $\Delta$  tel que

$$||\Delta \Phi x - x||$$

soit minimal.

On mesure l'efficacit  de l'algorithme par la donn e suivante

D FINITION 9 (PERFORMANCE D'UN ALGORITHME)

Soit  $K$  un compact de  $X$ ,

$X$  espace de Banach

$$E(K, \Phi, \Delta)_X = \sup_{x \in K} ||\Delta \Phi x - x||_X$$

On introduit  $A_{n,N} = [(\Phi, \Delta); \Phi \text{ matrice } (n, N)]$ .

Le meilleur algorithme possible est

D FINITION 10 (ALGORITHME OPTIMAL)

L'algorithme optimal qui utilise  $n$  de  $N$  informations initiales est

$$E_{n,N}(K)_X = \inf_{(\Phi, \Delta) \in A_{n,N}} E(K, \Phi, \Delta)_X$$

TH OR ME 9 (ALGORITHME OPTIMAL ET  PAISSEURS DE GELFRAND)

Soit  $K$  compact,  $K \subset \mathbb{R}^N$  tq

$K = -K$  et

$\exists C$  tq  $K + K \subset C.K$  alors on a

$$d^n(K)_X \leq E_{n,N}(K)_X < C d^n(K)_X$$

C'est   dire, la meilleure performance algorithmique possible d pend lin airement des  paisseurs de Gelfand

Dans notre cas, si l'algorithme était optimal, il devrait vérifier (avec  $n$  le nombre de mesures).

$$\|\Delta(\Phi)(f) - f\|_{L^2} \leq C\sqrt{(\log(N/n) + 1)/n}$$

L'algorithme utilisé par Candès et Tao montre qu'avec une matrice  $\Phi$  prise aléatoirement (la manière exacte dont elle est choisie sera explicitée plus tard) et en utilisant comme  $\Delta$  la minimisation de la norme  $L_1$  de  $\hat{f}$ , on respecte presque cette inégalité.

$$\Delta(y) = \operatorname{argmin}_{x, \Phi(x)=y} \|\hat{x}\|_{L_1}$$

En effet leur théorème final est :

En notant  $F$  la matrice de mesure ( qui mesure aléatoirement certaines valeurs de  $\hat{f}$  ou non, avec un paramètre  $\tau$ , voir détails ci-dessous)

**THÉORÈME 10 (PREMIER THÉORÈME DE CANDÈS ET TAO)**

*Soit  $n > \log(N)^6$  le nombre de mesures*

*soit  $f$  un signal dans  $\mathbb{R}^N$  tel que  $\hat{f} \in R_{\text{weak}} - L_p$  boule, pour  $p \leq 1$ .*

*Alors  $\forall \alpha > 0$ ,  $\alpha$  petit*

*N'importe quel ( le minimiseur peut ne pas être unique)  $f^\#$  vérifiant  $f^\# = \Delta(\Phi(f))$ , vérifie avec probabilité  $1 - O(N^{-p/\alpha})$*

$$\|f - f^\#\|_{L^2} \leq C_{p,\alpha} R(n/(\log(N)^6)^{-r}$$

Donc, avec  $p = 1$ , on a

$$\|f - f^\#\|_{L^2} \leq C\|x\|_{L^1} (\log(N))^3 / \sqrt{n}$$

donc, à une constante multiplicative ( de l'ordre de  $\log(N)^3$  ) près, l'algorithme est maximal au sens des épaisseurs de Gelfrand, avec une décroissance en  $\sqrt{n}$ ,  $n$  étant le nombre de mesures.

## 6 Le second théorème de Candès et Tao, application avec Romberg

Candès et Tao ont montré que l'on pouvait avoir une reconstruction exacte (et non proche), si le signal était "sparse", c'est à dire qu'il ne possédait qu'un petit nombre de fréquences non nulles.

Dans l'article précédent ( qui prouve le Premier Théorème), une des propriétés élémentaires utilisée était que la matrice de mesure se devait de vérifier le UUP : Uniform Uncertainty Princip, qui s'énonce ainsi.

**DÉFINITION 11 (UUP)**

*Une matrice  $A$  de taille  $N \times K$  vérifie le UUP avec paramètre  $\lambda$  si*

*$\forall \alpha > 0$   $\alpha$  petit*

*$\forall T \subset \mathbb{R}^N$  tq  $T \leq \alpha K / \lambda$  on a*

*$\forall f$  de support  $T$*

$$1/2/(K/N)||f||_{l_2}^2 \leq ||Af||_{l_2}^2 \leq 3/2(K/N)||f||_{l_2}^2$$

avec probabilité  $\geq 1 - O(N^{-p/a})$

On vérifie que la mesure aléatoire des fréquences utilisées vérifie cette propriété avec  $\lambda = \log(n)^6$

Pour vérifier le théorème de Candès et Tao, on utilise cette propriété (l'UUP), ainsi qu'une autre, l'ERP. Voir la section sur ce théorème pour plus de précisions.

Cette définition, bien que différente est à mettre en rapport avec la définition ci-dessous qui sera utilisée par la suite.

#### DÉFINITION 12 (RIP)

On dit que  $A$  vérifie le RIP ( *restricted isometry property*) de paramètre  $\lambda$  si

$\exists a_\lambda > 0$  tel que

$\forall T \subset \mathbb{R}^N$  tq  $|T| \leq \lambda$

$\forall f$  de support  $T$

$$(1 - a_\lambda)||f||_{l_2}^2 \leq ||A_T f||_{l_2}^2 \leq (1 + a_\lambda)||f||_{l_2}^2$$

ou  $A_T$  de taille  $N \times T$  est la matrice  $A$  restreinte aux colonnes de  $T$ .

Dans un cas (uniforme), on regarde toute la matrice, tandis que dans le cas du RIP, on restreint la matrice aux colonnes de  $T$ .

#### THÉORÈME 11 (SECOND THÉORÈME DE CANDÈS ET TAO.)

Soit  $\Phi$  une matrice vérifiant le RIP.

On note  $a_\lambda$  le plus petit nombre tel que la propriété du RIP soit vraie.

alors si  $a_{2,\lambda} < \sqrt{2} - 1$ ,

Soit  $x^\sharp$  le minimiseur  $l_1$  de  $\Phi(x)$

Soit  $x_s$  le vecteur  $x$  restreint à ses  $s$  plus grandes composantes.

Alors on a :

$$||x^\sharp - x||_{l_1} \leq C_0 ||x^\sharp - x_s||_{l_1}$$

et

$$||x^\sharp - x||_{l_2} \leq C_0 s^{-1/2} ||x^\sharp - x_s||_{l_2}$$

Donc pour une matrice vérifiant le RIP, la reconstruction des signaux "sparse" est exacte !

J.Romberg Candès et Tao on en déduit que si le signal comportait qu'un nombre faible de fréquences, alors on pouvait le retrouver exactement à partir d'un faible nombre de mesures.

#### THÉORÈME 12 (SECOND THÉORÈME DE CANDÈS ET TAO)

$f$  signal,  $f \in \mathbb{R}^n$

Soit  $T$  le support de  $\hat{f}$ ,  $\hat{f}$  transformée de Fourier de  $f$ .

Soit  $K \subset [0, 1, \dots, n]$  l'ensemble des mesures.

Alors si  $K > C \cdot |T| \cdot \log(N)$  on peut retrouver parfaitement  $f$  à partir de la donnée de  $f(k)$ , avec probabilité  $\geq 1 - O(N^{-p \cdot C})$

La reconstruction se fait comme précédemment.  
On résoud  $\argmin_{x \in \mathbb{R}^N} (\|\hat{x}\|_{L_1}, \Phi x = \Phi f)$

## 7 Premier théorème de Candès et Tao, les deux hypothèses

### 1) la matrice aléatoire

On se place dans  $\mathbb{R}^N$   
Soit  $\tau$  tel que  $0 < \tau < 1$   
Soit  $F$  la matrice  $N \times N$  :  $F(k, t) = \exp(-2i.\pi.k.t/N) / \sqrt{N}$   
Soit  $I_k, k \in [1, \dots, N]$   $N$  variables aléatoires IID de Bernouilli, de paramètre  $\tau$ .  
Soit  $\Omega = \{k : I_k = 1\}$   
On prends comme matrice  $F_\Omega$  la matrice  $F$  restreintes au lignes de  $\Omega$ . La matrice  $F_\Omega$  appartient a  $M(N, \Omega)$ .  
On note  $n = E(|\Omega|) = N.\tau$

### 2) Théorème

Rappelons le théorème prouvé par Candès et Tao qui nous intéresse.

THÉORÈME 13 (PREMIER THÉORÈME DE CANDÈS ET TAO)

Soit  $n > \log(N)^6$  le nombre de mesures

soit  $f$  un signal dans  $\mathbb{R}^N$  tel que  $f \in R_{\text{weak}} - L_p$  boule, pour  $p \leq 1$ .

Alors  $\forall \alpha > 0$ ,  $\alpha$  petit

N'importe quel (le minimiseur peut ne pas être unique)  $f^\#$  vérifiant  $f^\# = \Delta(\Phi(f))$  (on minimise la norme  $l_1$  de  $f$ ), vérifie avec probabilité  $1 - O(N^{-p/a})$

$$\|f - f^\#\|_{L^2} \leq C_{p,a} R(n / (\log(N)^6))^{-r}$$

En fait, on prouvera que  $n$  est généralement proche  $N.\tau$

### Propriétés préliminaires

La démarche utilisée pour parvenir à ce résultat consiste à prouver deux hypothèses plus faibles, qui, réunies, entraînent le théorème.

La première de ces hypothèses a déjà été explicitée, il s'agit de l'UUP (Uniform Uncertainty Principle). On le montre ici dans une version légèrement modifiée adaptée à l'ensemble de mesure choisie (on supprime le paramètre  $\lambda$  présent dans la version présentée ci-dessus en le remplaçant directement par une valeur adaptée).

DÉFINITION 13 (UUP)

On dit qu'une matrice  $A$  vérifie le UUP si

$\forall \alpha > 0$   $\alpha$  petit

$\forall T \subset \mathbb{R}^N$  tq  $T \leq \alpha K / (\log(N)^6)$

$\forall f$  de support  $T$ , on a

$$1/2/K/N||f||_{l_2}^2 \leq ||Af||_{l_2}^2 \leq 3/2(K/N)||f||_{l_2}^2$$

avec probabilité  $\geq 1 - O(N^{-p/a})$

La seconde hypothèse que nous avons à vérifier est nommé ERP, pour Exact Reconstruction Principle. La voici.

#### DÉFINITION 14 (ERP)

On dit que une matrice  $A$  vérifie l'ERP si

$\forall \alpha > 0$ ,  $\alpha$  petit

$\forall T \subset \mathbb{R}^N$  tq  $T \leq \alpha K / \log(N)$

$\forall$  vecteur signe  $\sigma$  défini sur  $T$ , ( $|\sigma(t)| = 1$ ),

$\exists$  avec probabilité  $> (1 - O(N^{p/\alpha}))$   $P \in \mathbb{R}^N$  tel que

- 1)  $P(t) = \sigma(t), \forall t \in T$
- 2)  $\exists V \in |\Omega|$  tq  $P = F_\Omega V$
- 3)  $||P(t)||_{T^c}^\infty \leq 1/2$

La preuve exacte et détaillée que la matrice  $F_\Omega$  vérifie ces deux propriétés est très calculatoire et présente à première vue peu d'intérêt, De même, la démonstration que ces deux propriétés entraînent le théorème est basé sur des propriétés simples de minimisation, propriétés de la matrice adjointe, propriétés de norme.. et ne sera pas abordée. Je tacherai donc de tenter d'extraire les données et principes essentiels qui permettent de montrer que la matrice  $F_\Omega$  vérifie ces propriétés, en faisant le détour par de nouvelles théories.

## 8 Idées de preuve de l'ERP et inégalités de grandes déviations

### 1) les inégalités de "grandes déviations"

La vraie théorie des inégalités de grandes déviations n'est pas celle explicité ici. On parlera des inégalités qui permettent d'estimer la probabilité qu'une somme s'écarte de son espérance, qu'elle dévie.

C'est pourquoi nous emploieront le terme "grande déviation".

Deux de ces inégalités sont utilisées, et sont indispensables par la suite. on remarquera que les deux sont très proches et sont compatibles entre elles.

La première inégalité est celle de Bernstein.

#### THÉORÈME 14 (INÉGALITÉ DE BERNSTEIN)

Soit  $X_1 \dots X_n$  n IID VA tel que

$\forall i \in [1, \dots, n]$  on a  $E(X_i) = 0$  et  $|X_i| < c$

alors on a

$$P(\sum_{i=1}^N X_i > \lambda) \leq 2 \exp(-\lambda^2 / (2.\sigma^2 + 2.c\lambda/3))$$

où on pose  $\sigma^2 = \sum_{i=1}^n \text{Var}(X_i)$

La seconde en est, d'une certaine manière une conséquence. Bien que non utilisée par la suite dans le rapport, elle est utilisée par Candès et Tao dans leur article. Je la montre ici afin de montrer à quel point les théories probabilistes de compressed sensing reposent sur ces inégalités.

Il s'agit de l'inégalité de Hoeffding.

**THÉORÈME 15 (INÉGALITÉ DE Hoeffding)**

*Soit  $X_1 \dots X_n$   $n$  IID VA de Bernoulli de paramètre  $1/2$*

*En considérant  $S$  la somme  $S = \sum_{j=1}^N a_j X_j$*

*alors on a*

$$P(S > \lambda) \leq 2 \exp(-\lambda^2 / (2 \|a\|_2^2))$$

**2) Application.**  $|\Omega| \approx N\tau$

On observe, en utilisant l'inégalité de Bernstein que :

$$P(S = \sum_{k \in \mathbb{Z}^n} (I_k - \tau) > \lambda) < 4 \exp((- \sigma^2) / (4(\sigma^2 + \lambda / (3\sqrt{2}))))$$

Inégalité à partir de laquelle on peut simplement déduire que, pour  $K \geq C\alpha^{-1} \log(N)$

$$P(K/2 < |\Omega| < 2K) \geq 1 - O(N^{-p/\alpha})$$

C'est cette condition : " $K > \text{Constante} \cdot \log(N)$ "; qui va faire qu'on va voir réapparaître le  $\log(N)$  dans l'ERP.

## 9 Idées de preuve de l'UUP, entropie, et ouverture

### 1) Ouverture

Revenons au cas continu. On rappelle que  $E_K = \{f \in L^2(\mathbb{R}^n), \text{supp}(TF(f)) \in K\}$   
On appelle  $C_R$  le cube  $[-R, R]^d$

**DÉFINITION 15** ( $\Gamma(R, \delta)$ )

*Soit  $\delta \in \mathbb{R}$ , tq  $1 > \delta > 0$*

*On appelle*

$$\Gamma(R, \delta) = \{f \text{ tq } f \in E_{[-P_i, P_i]}, \int_{C_R} f^2 \geq (1 - \delta) \int f^2 \}$$

Encore une fois, le rapport avec le discret est évident, on s'intéresse à l'échantillonnage d'une fonction à fréquences limitées, le tout sur un compact, sur lequel la quasi totalité de l'énergie de  $f$  est contenue. Bass et Gröchenig ont montré le théorème suivant. Soit  $\{x_j, j \in \mathbb{N}\}$  une séquence de variables aléatoires indépendantes de loi uniforme sur  $C_R$

**THÉORÈME 16 (THÉORÈME DE BASS ET GRÖCHENIG)**

$\exists A, B > 0$  proches de 1

$\exists C, D > 0$  tq ( $D$  dépend de  $A$  et  $B$ )

tels que l'inégalité  $A.(r/R^d) \|f\|_2^2 \leq \sum_{j=1}^r |f(x_j)|^2 \leq B.(r/R^d)$

soit vérifiée avec une probabilité

$$P \geq 1 - C \exp^{-Dr/R^d}$$

On voit que, comme pour l'UUP, il s'agit de montrer une inégalité de type inégalité d'échantillonnage, le tout pour des variables aléatoires. On peut voir aussi que, comme pour l'UUP, les probabilités de réussite sont de type exponentiel. On retrouve dans les deux démonstrations de fortes similitudes, que nous considérerons comme étant le socle d'une telle inégalité.

## 2) entropie

La première consiste donc à utiliser l'entropie, les covering numbers, et à s'en servir pour construire une suite convergeant diadiquement vers le signal d'origine. On notera que, auparavant, Bass et Gröchenig ont montré que  $\Gamma(R, \delta)$  était un compact pour la norme infinie, élément indispensable pour la suite.

On restera dans le cadre discret par la suite, mais les résultats exhibés sont utilisés de façon équivalentes dans le cas continu.

DÉFINITION 16 (NORME X)

soit  $f \in \mathbb{R}^N$

On appelle X norme de f la valeur

$$||f||_X = \sqrt{N} ||f||_\infty$$

DÉFINITION 17 (ENTROPIE DE KOLMOGOROV)

$B_X = \{x \in X, ||x||_X \leq 1\}$

Soit  $U \subset X$  un sous ensemble borné de X. On appelle  $N(U, B_X, r) \in \mathbb{N}$  le plus petit entier tel que on puisse recouvrir U par  $N(U, B_X, r)$  boules  $rB_X$ .

L'entropie de Kolmogorov est définie par

$$\varepsilon(U, B_X, r) = \log_2(N(U, B_X, r))$$

On utilise les "covering number", pour construire des "maillages"  $(A_n)$  de  $B_X$ , tel que la finesse du "maillage"  $A_n$  soit de  $2^{-n}$ , tout en controlant le cardinal des  $\#A_n$ . Les  $A_n$  étant donc des sous ensembles discrets de  $B_X$ , tels que  $B_X \subset A_n + 2^{-n}B_X$ .

Dans le cas de Bass et Gröchenig on se contente de la  $||\cdot||_\infty$  car on mesure directement f et non pas  $\hat{f}$ , on a donc pas besoin de passer dans le domaine fréquentiel.

**remarque :** Dans les deux articles, on se ramène facilement a une inégalité centrée en 0. Dans le cas discret on montre qu'il suffit de montrer

$$\sup_f |\sum_{k \in [1, \dots, N]} (I_k - \tau) \hat{f}(k)^2| \leq (1/4)\tau$$

On trouve alors des  $f_n$  où  $f_n \in A_n$  tel que  $||f - f_n|| \leq 2^{-n}$ .

On ne va pas jusqu'à l'infini, la précision devenant suffisamment bonne au bout d'un certain rang. On a alors  $f = f_0 + \sum_{i=1..N} f_{j+1} - f_j$ , où  $||f_{j+1} - f_j|| \leq 2^{j+1}$

On a alors une dissension des cas.

1) Dans le cas discret, Candès et Tao utilisent la norme  $l_2$ , et passent donc au carré. Cela s'explique par le fait qu'ils travaillent sur  $\hat{f}$  alors qu'ils cherchent une inégalité sur f, ils ont donc besoin de la propriété d'isométrie de la transformée de Fourier. Le passage au carré induit une complication dans les calculs, qu'on ne cherchera pas a mettre en valeur.

2) Bass et Gröchenig travaillent déjà sur f, et ont donc une approche plus simpliste. On voit toutefois qu'au stade final, les deux approches arrivent à un type d'inégalité



relativement proche. et utilisent le même argument pour conclure.

### 3) Inégalités et arguments finaux

Dans le cas discret, malgré le fait qu'il soit scindé, les deux voies arrivent à une inégalité du même type. Candès et Tao utilisent les propriétés métriques qui lient les normes  $1, 2, \infty$  et la norme  $X$ , pour arriver à se rapprocher de résultats obtenus beaucoup plus simplement par Bass et Gröchenig.

Dans le cas discret on arrive à une inégalité du type

$$P(|\sum(I_k - \tau a(k))| \geq \gamma)$$

la somme étant sur un ensemble  $E_1$  dans le premier cas, et sur un ensemble  $E_2$  dans le second cas.

On utilise alors les inégalités de grandes deviations pour conclure. On peut notamment extraire de l'inégalité de Bernstein explicitée ci-dessus une inégalité pour ce cas particulier.

Pour Bass et Gröchenig, on arrive de même a une inégalité de ce type. Explicitement, on calcule

$$P(\sum_{j=1}^r Y_j f_n - Y_j f_{n-1} \geq \lambda / (2.n^2))$$

Où  $Y_j$  est une variable aléatoire de  $[0, 1]$ ,  $f_n$  étant la fonction de  $A_n$  tel que  $\|f - f_n\| \leq 2^{-n}$ , et où  $Y_j f = f(Y_j)$ .

De même, on utilise l'inégalité de Bernstein, car on a des résultats sur la variance et la moyenne de  $Y_j f$ .

### Sur l'article de Bass et Gröchenig

L'article de Bass et Gröchenig et l'inégalité qu'ils prouvent ne sont d'aucune utilité pratique, et peuvent être considérés comme simplement intéressant mathématiquement. En effet, leur approche se situe a l'opposé de Candès et Tao, et peut être comparée à une méthode de MonteCarlo. Cependant, les résultats prouvés ont l'avantage d'indiquer la vitesse de convergence. De ce fait, l'exhibition de cette vitesse permet de rejoindre l'approche de Candès et Tao, car elle montre "quel est le nombre minimal de mesures a faire pour estimer  $f$ ".

## 10 Application : La tomographie

L'une des applications les plus répandues des théorèmes de Candès et Tao est la tomographie, technique d'imagerie particulièrement utilisée en imagerie médicale, mais aussi en géophysique ainsi qu'en astrophysique.

Il s'agit de retrouver une image alors qu'on ne possède des échantillons de  $\hat{f}$  que sur des tranches ; tranches que l'on peut observer sur le schéma (b) ci dessous.

Ici, les images sont extraites d'un article de Candès et Tao ( voir [9] ), mettant en valeur leur second théorème.

En (a), on part d'une image initiale.

On mesure ses coefficients de fourier sur les tranches visibles en blanc sur l'image (b).

L'image (c) résulte de la minimisation de la norme  $l_2$ , le résultat est de qualité médiocre. L'image (d) résulte de la minimisation de la norme  $l_1$ , il s'agit d'une reconstruction exacte, l'image (d) est indentique à (a).

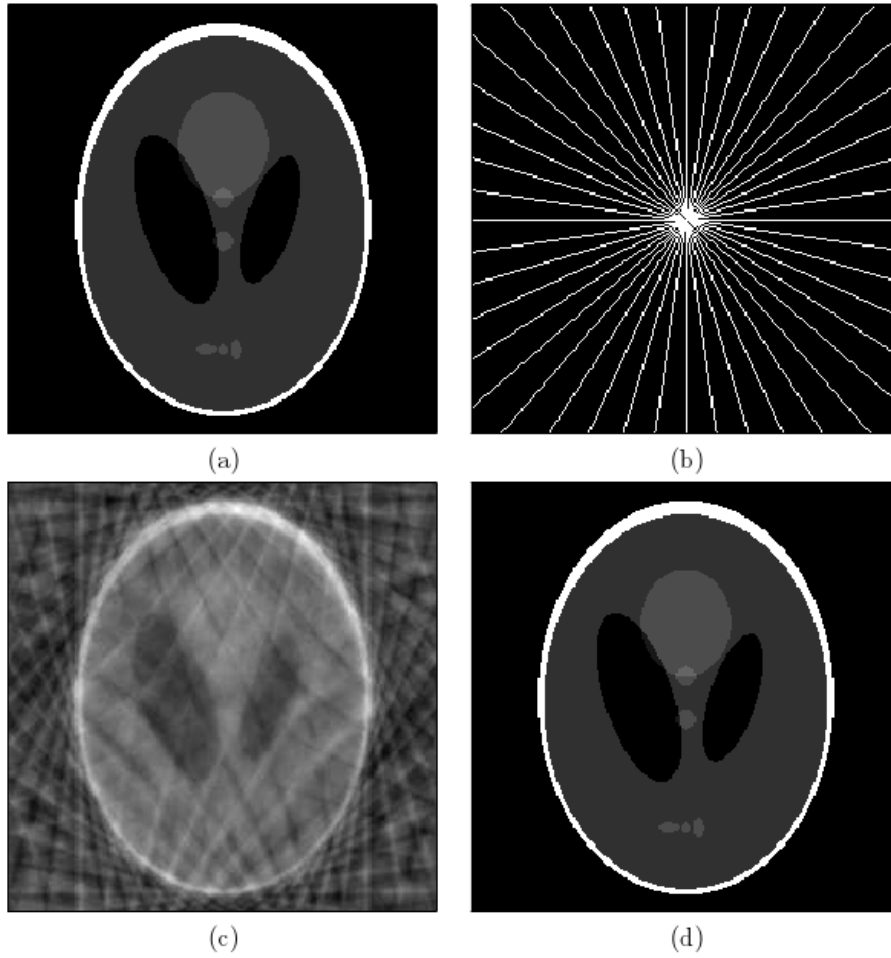


FIG. 1 – Exemples de reconstruction exacte d'image.

## Références

- [1] Yves Meyer, *Un contre exemple*.
- [2] F. Bass and Grochenig, *Random sampling for bandlimited functions* arXiv :0706.3818v2, 16 Juin 2008.
- [3] Yves Meyer, *Beyond the Shannon-Nyquist rate*
- [4] Yves Meyer, *Le compressed sensing* 2010
- [5] Candès, *The Restricted Isometry Property and Its Implications for Compressed Sensing*
- [6] E. Candès et T.Tao, *Near Optimal Signal Recovery From Random Projections Universal Encoding Strategies ?*. October 2004, Revised March 2006
- [7] R. Baraniuk, M. Davenport, R. DeVore et M. Wakin, *The Johnson-Lindenstrauss Lemma meets compressed sensing*. Rice University 2006
- [9] E. Candès et T. Tao, *Robust Uncertainty Principles : Exact Signal Reconstruction from Highly Incomplete Frequency Information*. 2004