

Une exploration numérique des performances de l'échantillonnage compressé

Charles Dossal, Gabriel Peyré, Jalal M. Fadili

► **To cite this version:**

Charles Dossal, Gabriel Peyré, Jalal M. Fadili. Une exploration numérique des performances de l'échantillonnage compressé. Grets'i'09, Sep 2009, Dijon, France. 2009. <hal-00385018>

HAL Id: hal-00385018

<https://hal.archives-ouvertes.fr/hal-00385018>

Submitted on 18 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une exploration numérique des performances de l'échantillonnage compressé

Charles DOSSAL¹, Gabriel PEYRE², Jalal FADILI³

¹LaBAG / Université Bordeaux 1

²CEREMADE / Université Paris Dauphine

³GREYC / ENSICAEN

charles.dossal@math.u-bordeaux1.fr, gabriel.peyre@ceremade.dauphine.fr,
jalal.fadili@greyc.ensicaen.fr

Résumé – Cet article explore numériquement l'efficacité de la minimisation ℓ^1 pour la restauration de signaux parcimonieux depuis des mesures compressibles, dans le cas sans bruit. Nous proposons un algorithme glouton qui calcule des vecteurs parcimonieux difficile à retrouver par minimisation ℓ^1 . Cet algorithme est inspiré par des critères topologiques d'identifiabilité ℓ^1 . Nous évaluons numériquement l'analyse théorique sans avoir à utiliser un échantillonnage de Monte-Carlo, qui tend à éviter les cas pathologiques. Ceci permet de mettre à l'épreuve les critères d'identifiabilité exploitant des projections de polytopes et des propriétés d'isométrie restreinte.

Abstract – This paper explores numerically the efficiency of ℓ^1 minimization for the recovery of sparse signals from compressed sampling measurements in the noiseless case. Inspired by topological criteria for ℓ^1 -identifiability, a greedy algorithm computes sparse vectors that are difficult to recover by ℓ^1 -minimization. We evaluate numerically the theoretical analysis without resorting to Monte-Carlo sampling, which tends to avoid worst case scenarios. This allows one to challenge sparse recovery conditions based on polytope projection and on the restricted isometry property.

1 Échantillonnage compressé

L'échantillonnage compressé correspond à l'acquisition d'un petit nombre de mesures linéaires $y = Ax$, où $x \in \mathbb{R}^N$ est le signal haute résolution que l'on souhaite retrouver et $y \in \mathbb{R}^P$ sont les $P \ll N$ mesures.

La résolution du problème inverse mal posé $y = Ax$ est stabilisé en considérant une matrice $A = (a_i)_{i=0}^{N-1} \in \mathbb{R}^{P \times N}$ tirée aléatoirement suivant une loi adéquate. Cet article considère, par simplicité, le cas où les coefficients de A sont tirés indépendamment selon une loi Gaussienne centrée de variance $1/P$.

Pour des mesures sans bruit $y = Ax$, un vecteur x est récupéré par minimisation convexe

$$x^* = \operatorname{argmin}_{\tilde{x} \in \mathbb{R}^N} \|\tilde{x}\|_1 \quad \text{sous la contrainte} \quad A\tilde{x} = y, \quad (1)$$

où $\|\tilde{x}\|_1 = \sum_i |\tilde{x}_i|$. Le vecteur x est dit identifiable si $x^* = x$ pour $y = Ax$.

Avec forte probabilité sur la matrice d'échantillonnage A , la théorie de l'échantillonnage compressé [3, 2] montre que n'importe quel vecteur x satisfaisant

$$\|x\|_0 = \#\{i \mid x_i \neq 0\} \leq \rho(P/N)P \quad (2)$$

est identifiable pour $\rho(\eta) > 0$ fonction croissance de η .

2 Critère parcimonieux d'identifiabilité

Critères topologiques d'identifiabilité. Des critères d'identifiabilité précis sont obtenus en considérant les positions et les signes des coefficients non nuls de x , qui sont indexés par le support de x

$$I = I(x) = \{i \mid x_i \neq 0\}.$$

Ceci permet d'utiliser les interactions entre les colonnes de $A = (a_i)_i$ indexées par I , notées $A_I = (a_i)_{i \in I}$ et les autres $(a_i)_{i \notin I}$. Fuchs [9] montre que x est ℓ^1 -identifiable si

$$F(x) = \max_{i \notin I} |\langle a_i, d(x) \rangle| < 1, \quad (3)$$

$$\text{où} \quad d(x) = A_I(A_I^*A_I)^{-1} \operatorname{sign}(x_I) \quad (4)$$

Critères d'identifiabilité topologiques. Le polytope centrosymétrique $A(B_1)$ où $B_1 = \{\tilde{x} \mid \|\tilde{x}\|_1 \leq 1\}$ est l'image de la boule ℓ^1 . C'est aussi l'enveloppe convexe de $\{\operatorname{sign}(x_i)a_i\}_i$. La $\|x\|_0$ -facette $f_x \subset A(B_1)$ sélectionnée par x est l'enveloppe convexe de $\{\pm a_i\}_{i \in I}$. Donoho [4] montre que

$$x \text{ est } \ell^1\text{-identifiable} \iff f_x \in \partial A(B_1) \quad (5)$$

où $\partial A(B_1)$ est la frontière du polytope $A(B_1)$. Dossal [5] montre que cette condition topologique est équivalente à ce que x soit la limite de vecteurs x_n tels que $F(x_n) < 1$.

En utilisant (5), Donoho [4] détermine, dans le cas sans bruit $y = Ax$, une valeur précise pour $\rho(\eta)$ dans (2). On a par exemple $\rho(1/2) \approx 0.089$ et $\rho(1/4) \approx 0.065$.

Critères d'isométrie restreinte. Les travaux originaux de Donoho [3], Candès, Romberg et Tao [2] se concentrent sur la stabilité de l'échantillonnage compressé. Dans ce but, ces auteurs introduisent la propriété d'isométrie restreinte (RIP), qui impose l'existence de constantes $0 < \delta_s^{\min} \leq \delta_s^{\max} < 1$ telles que pour tout $x \in \mathbb{R}^N$ avec $\|x\|_0 \leq s$,

$$(1 - \delta_s^{\min})\|x\|^2 \leq \|Ax\|^2 \leq (1 + \delta_s^{\max})\|x\|^2. \quad (6)$$

Soit $A_I = (a_i)_{i \in I}$ une matrice de $s = |I|$ colonnes extraite de A . Alors (6) équivaut à dire que pour tout I such that $|I| = s$, les plus petite et plus grande valeurs propres $\lambda^{\min}(A_I)$ et $\lambda^{\max}(A_I)$ de la matrice de Gram $A_I^* A_I$ sont bornées respectivement inférieurement et supérieurement par $1 - \delta_s^{\min}$ and $1 + \delta_s^{\max}$. Ainsi, les constantes RIP sont définies de façon équivalente comme

$$\delta_s^{\min} = \max_{|I|=s} \delta^{\min}(A_I) \quad \text{and} \quad \delta_s^{\max} = \max_{|I|=s} \delta^{\max}(A_I)$$

where $\begin{cases} \delta^{\min}(A_I) = 1 - \lambda^{\min}(A_I), \\ \delta^{\max}(A_I) = \lambda^{\max}(A_I) - 1 \end{cases}$.

Dans les travaux originaux de Candès et al., le RIP est symétrique, et des constantes égales sont utilisées, $\delta_s^{\min} = \delta_s^{\max} = \delta_s$. Ces auteurs montrent qu'une valeur assez petite de δ_{2s} assure l'identifiabilité de tous les vecteur s -parcimonieux. Ceci est obtenu avec grande probabilité sur la matrice A si $s \leq CP/\log(N/P)$, ce qui correspond à la condition (2) avec $\rho(\eta) \leq C/\log(\eta^{-1})$.

On peut montrer que les plus grande et plus petite valeurs propres de la matrice de Gram $A_I^* A_I$ ne dévient pas à la même vitesse de 1. En utilisant des constantes RIP asymétriques, Foucart and Lai [8] montrent que

$$(4\sqrt{2} - 3)\delta_{2s}^{\min} + \delta_{2s}^{\max} < 4(\sqrt{2} - 1) \quad (7)$$

assure l'identifiabilité de tout vecteur s -parcimonieux. Blanchard et al. [1] déterminent ρ_0 tel qu'avec forte probabilité sur la matrice A ,

$$\|x\|_0 \leq \rho_0(P/N)P \quad (8)$$

assure que la condition (7) est satisfaite. On a nécessairement $\rho^0(\eta) \leq \rho(\eta)$ puisque la condition (8) garantie l'identifiabilité, mais elle assure également une forte robustesse à des mesures bruitées. Les constantes ρ_0 ainsi obtenue sont assez petites, et on a par exemple $\rho_0(1/2) = 0.003$ et $\rho_0(1/4) = 0.0027$.

3 Facettes intérieures et vecteurs non-identifiables

Une heuristique pour l'identifiabilité exploitant $1/\|d(x)\|$. On déduit de (5) qu'un vecteur x non identifiable correspond à une facette f_x appartenant à l'intérieur du polytope $A(B_1)$. La proposition suivante, démontrée dans [7], permet de calculer la distance de f_x au centre du polytope.

Proposition 1. Pour tout vecteur x tel que $\text{rank}(A_I) = |I|$, la distance de la facette f_x à 0 est $\frac{1}{\|d(x)\|}$, où $d(x)$ est défini en (4).

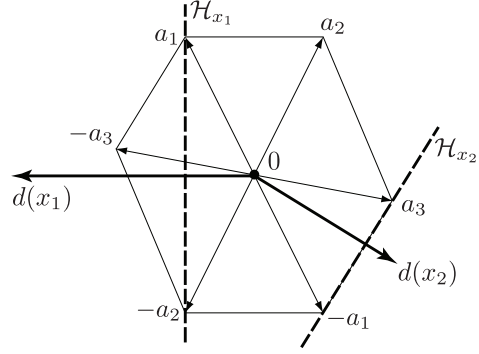


FIG. 1 – Géométrie de l'identifiabilité ℓ^1 , pour $N = 3$ and $P = 2$. Le vecteur $x_1 = (2, -3, 0)$ n'est pas identifiable car f_{x_1} est à l'intérieur du polytope $A(B_1)$, et a donc une grande valeur de $\|d(x_1)\|$. Au contraire, $x_2 = (-5, 0, 3)$ est identifiable car $f_{x_2} \in \partial A(B_1)$, et a une petite valeur de $\|d(x_2)\|$.

La figure 1 illustre cette proposition en dimension $P = 2$. Cette propriété, ainsi que la condition (5), suggère qu'un vecteur x ayant une petite valeur de $1/\|d(x)\|$ a plus de chance d'être non identifiable.

La figure 2 estime, par échantillonnage de Monte-Carlo, le ratio de vecteurs identifiables, en fonction de la parcimonie $\|x\|_0$ (courbe de gauche) et d'une valeur quantifiée de $\|d(x)\|$ (courbe de droite). La courbe paramétrée par $\|d(x)\|$ présente une transition de phase plus abrupte que la courbe paramétrée par la parcimonie (chaque point sur les courbes prend en compte 1000 réalisations aléatoires).

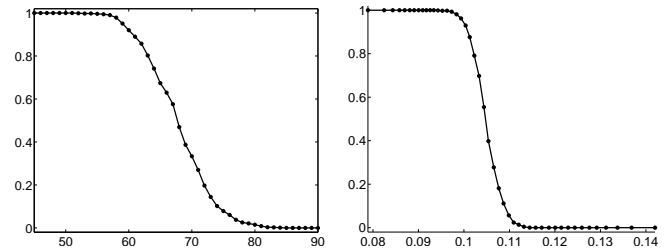


FIG. 2 – Gauche : ratio de vecteurs x identifiables en fonction de $\|x\|_0$, pour $(P, N) = (250, 1000)$. Droite : ratio de vecteurs x identifiables en fonction de $\|d(x)\|$.

La figure 2 suggère donc que des vecteurs non-identifiables peuvent être trouvés non-seulement en augmentant la parcimonie $\|x\|_0$ d'un vecteur x , mais aussi en décroissant la valeur de $1/\|d(x)\|$.

Une heuristique pour le conditionnement des sous matrice utilisant $1/\|d(x)\|$. Étant donné un vecteur $d \in \mathbb{R}^P$, on définit deux régions de \mathbb{R}^P

$$\begin{cases} C_d = \{v \in \mathbb{R}^P \mid |\langle d, v \rangle| \geq 1\}, \\ C_d^c = \{v \in \mathbb{R}^P \mid |\langle d, v \rangle| \leq 1\}. \end{cases}$$

dans le cas normalisé où les colonnes de A sont de norme unité, \mathcal{C}_d définit une double calotte sphérique, dont le rayon r vérifie $r^2 = 1 - 1/\|d\|^2$.

La proposition suivante, démontrée dans [6], montre que le regroupement des vecteurs $\{a_i\}_{i \in I}$ dans ces régions permet de définir des bornes inférieures des constantes RIP de A_I .

Proposition 2. *Si $\{a_i\}_{i \in I} \subset \mathcal{C}_d$, alors*

$$\delta_s^{\max} \geq s/\|d\|^2 - 1. \quad (9)$$

Si $\{a_i\}_{i \in I} \subset \mathcal{C}_d^c$ et si $d \in \text{Span}(a_i)_{i \in I}$ alors

$$\delta_s^{\min} \geq 1 - s/\|d\|^2. \quad (10)$$

Étant donnée une sous-matrice A_I , une estimée précise de $\delta^{\max}(A_I)$ est obtenue en maximisant le membre de droite de (9). Ceci est obtenu en identifiant la région \mathcal{C}_d contenant les colonnes de A_I et qui correspond à la plus petite valeur possible pour $\|d\|$. Le calcul de ce groupement optimal est cependant difficile en grande dimension. Il est donc nécessaire de calculer un groupement approché en utilisant un vecteur d bien choisi afin de définir \mathcal{C}_d . La proposition suivante, démontrée dans [6], montre que le vecteur $d(x)$ fournit une telle approximation.

Proposition 3. *Pour n'importe quel vecteur x tel que A_I soit de rang pleins, avec $I = I(x)$, le vecteur $d(x)$ est tel que $\langle \text{sign}(x_i)a_i, d(x) \rangle = 1$. N'importe quelle région \mathcal{C}_d satisfaisant cette propriété donne un plus mauvaise borne sur δ_s^{\max} .*

Ainsi, la région $\mathcal{C}_{d(x)}$ est un choix optimal pour estimer la valeur de $\delta^{\max}(A_I)$ à l'aide de (9) si on se restreint à des régions qui passent par les vecteurs $\{\text{sign}(x_i)a_i\}_{i \in I}$. De meilleures estimations sont possibles en utilisant une autre région \mathcal{C}_d passant seulement par un sous ensemble de ces vecteurs où en utilisant un autre jeu de signes, mais il n'est pas évident de les calculer efficacement. Lorsque les signes $\text{sign}(x_i)$ sont fixés, le vecteur $d(x)$ est rapide à calculer car il ne nécessite que l'inversion d'un système linéaire sur-déterminé.

4 Extension Parcimonieuses

Afin de construire un vecteur x non-identifiable, ou pour trouver une sous matrice A_I mal conditionnée avec $I = I(x)$, on fait croître progressivement la parcimonie $\|x\|_0$. On considère une extension signée \tilde{x} de x , écrite comme $\tilde{x} = x + \sigma \Delta_i$, où $\sigma \in \{+1, -1\}$, $i \notin I(x)$ et Δ_i est un Dirac. Une telle extension accroît la parcimonie $\|x\|_0$, et nous sélectionnons de façon optimale i et σ pour maximiser ou minimiser la variation de $1/\|d(x)\|$, où $d(x)$ est défini en (4).

La proposition suivante, démontrée dans [6, 7], nous donne la clef pour trouver une extension optimale.

Proposition 4. *Soit $\tilde{a}_i \in \text{Span}(a_j, j \in \tilde{I})$ le vecteur dual tel que*

$$\forall j \in I, \langle \tilde{a}_i, a_j \rangle = 0 \quad \text{et} \quad \langle \tilde{a}_i, a_i \rangle = 1.$$

Alors

$$\|d(\tilde{x})\|^2 = \|d(x)\|^2 + \|\tilde{a}_i\| |\langle d(x), a_i \rangle - \sigma|.$$

Le calcul de $\|\tilde{a}_j\|$ pour toutes les valeurs de $j \notin I(x)$ est coûteux car ceci demande la résolution d'un système linéaire sur-déterminé pour chaque j . Nous calculons donc une extension optimale approchée en maximisant ou minimisant $|\langle d(x), a_j \rangle - \sigma|$ au lieu de $\|\tilde{a}_j\| |\langle d(x), a_j \rangle - \sigma|$. Une telle maximisation (resp. minimisation) de $1/\|d(x)\|$ est ainsi obtenue à l'aide des extensions

$$\mathcal{E}^+(x) = x + \sigma^+ \Delta_{i^+} \quad \text{et} \quad \mathcal{E}^-(x) = x + \sigma^- \Delta_{i^-}$$

$$\text{où} \quad \begin{cases} i^+ = \underset{j \notin I(x)}{\text{argmin}} |1 - |\langle d(x), a_j \rangle|| \\ i^- = \underset{j \notin I(x)}{\text{argmax}} |\langle d(x), a_j \rangle| \end{cases} \quad (11)$$

$$\text{et} \quad \begin{cases} \sigma^+ = \text{sign}(\langle d(x), a_{i^+} \rangle), \\ \sigma^- = -\text{sign}(\langle d(x), a_{i^-} \rangle). \end{cases} \quad (12)$$

Pour chaque $j \in \{0, \dots, N-1\}$, en commençant par un vecteur 1-parcimonieux initial $x_{j,0}^\pm = \Delta_j$, nous calculons itérativement deux extensions s -parcimonieuses maximale et minimale comme

$$x_{j,s}^+ = \mathcal{E}^+(x_{j,s-1}^+), \quad \text{et} \quad x_{j,s}^- = \mathcal{E}^-(x_{j,s-1}^-). \quad (13)$$

5 Recherche de vecteurs non-identifiables

La proposition 1 suggère que l'extension gloutonne $x_{j,s}^-$ pour différents j définie en (13) a de forte chance d'être difficile à identifier.

Étant donné $\eta = P/N \leq 1$, nous utilisons une recherche dichotomique selon s pour calculer

$$s^*(\eta, P) = \min \{s \setminus \exists j, x_{j,s}^- \text{ n'est pas identifiable} \}. \quad (14)$$

Cette valeur fournit une borne supérieure empirique sur la parcimonie maximale garantissant l'identifiabilité.

La table ci-dessous détaille nos résultats numériques pour $\eta = 1/4$, et compare ces valeurs numériques avec la borne théorique de Donoho [4] $\rho(1/4) \sim 0.065$.

P	125	250	500	1000
$s^*(1/4, P)$	10	20	42	79
$\lceil \rho(1/4)P \rceil$	9	17	33	65

Par exemple, pour $N = 1000$ et $P = 250$, nous sommes capables de trouver un vecteur de parcimonie 20 qui n'est pas identifiable. Ceci est à comparer avec un échantillonnage de Monte Carlo, qui ne trouve aucun vecteur non identifiable de parcimonie $s < 54$, avec 1000 réalisations aléatoires pour chaque valeur de s , comme montré à la figure 2.

6 Recherche gloutonne de sous-matrices mal conditionnées

Bornes empiriques d'isométrie restreinte. L'extension $x_{j,s}^-$ définie en (13) est un vecteur s -parcimonieux vector avec une

petite valeur de $1/\|d(x)\|$. La proposition 2 suggère que son support $I = I(x_{j,s}^-)$ sélectionne une matrice de Gram $A_I^* A_I$ avec dont la plus petite valeur propre $\lambda_{\min}(A_I)$ est petite. De façon similaire, $I = I(x_{j,s}^+)$ peut être utilisé pour trouver une sous-matrice $A_I^* A_I$ dont la plus grande valeur propre $\lambda_{\max}(A_I)$ est grande.

On définit ainsi des borne inférieures empiriques des constantes d'isométrie restreintes $\tilde{\delta}_s^{\min} = \min_j 1 - \lambda_{\min}(A_{I(x_{j,s}^-)})$ et $\tilde{\delta}_s^{\max} = \max_j \lambda_{\max}(A_{I(x_{j,s}^+)}) - 1$.

La figure 3, gauche et droite, montre les valeurs numériques obtenues pour $\tilde{\delta}_s^{\min}$ et $\tilde{\delta}_s^{\max}$, et les compare avec des bornes plus simples, obtenues comme suit :

- *Échantillonnage aléatoire* : nous utilisons 10^4 sous-matrices de s colonnes extraites aléatoirement de A .
- *Échantillonnage conique* : nous sélectionnons pour chaque j les $s - 1$ colonnes $\{a_i\}_{i \in I \setminus j}$ qui sont les plus corrélées avec a_j .

Ceci montre que notre méthode gloutonne est capable de trouver des sous-matrices beaucoup plus mal conditionnées que des méthodes de recherches plus simples.

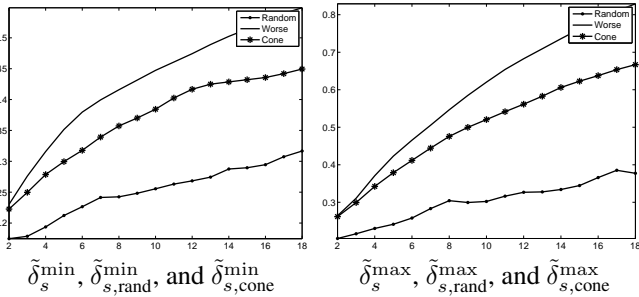


FIG. 3 – Constante empirique d'isométrie restreintes, pour $(N, P) = (4000, 1000)$. Les courbes en pointillés montrent $1 - s/\|d(x_s^-)\|^2$ sur la gauche, et $s/\|d(x_s^+)\|^2 - 1$ sur la droite.

On note x_s^- le vecteur atteignant les bornes empiriques $\tilde{\delta}_s^{\min} = 1 - \lambda_{\min}(A_{I(x_s^-)})$. La figure 4, gauche montre que les valeurs de $1 - s/\|d(x_s^-)\|^2$ sont proche des bornes empiriques d'isométrie restreintes $\tilde{\delta}_s^{\min}$. La même chose est vraie pour l'estimation de $\tilde{\delta}_s^{\max}$ à l'aide de $s/\|d(x_s^+)\|^2 - 1$. Ceci montre numériquement que notre heuristique est précis en pratique.

Bornes empiriques de parcimonie pour les conditions d'isométrie restreintes. Étant donné $\eta = P/N$, nous définissons $s^*(\eta, P)$ la parcimonie s minimum qui invalide la borne (7), c'est-à-dire

$$(4\sqrt{2} - 3)\tilde{\delta}_{2s}^{\min} + \tilde{\delta}_{2s}^{\max} \geq 4(\sqrt{2} - 1).$$

La figure 4, droite, montre notre estimation numérique de la borne (7) pour un s variable. La table ci-dessous détaille nos résultats numériques pour $\eta = 1/4$, et compare ces valeurs numériques avec la borne théorique de Blanchard et al. [1], $\rho(1/4) \approx 0.027$.

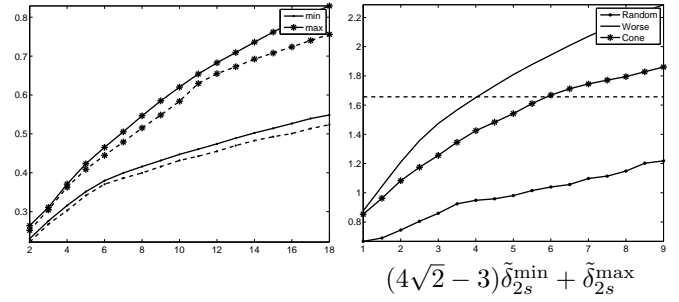


FIG. 4 – Gauche : courbes pleines : valeur de $\tilde{\delta}_s^{\min}$ et $\tilde{\delta}_s^{\max}$ en fonction de s ; courbes en pointillés : valeur de $1 - s/\|d(x_s^-)\|^2$ et $s/\|d(x_s^+)\|^2 - 1$. Droite : borne inférieure sur la condition (7), pour $(N, P) = (4000, 1000)$, la courbe en pointillés correspond à $y = 4(\sqrt{2} - 1)$.

P	250	500	1000	2000
$s_0^*(1/4, P)$	1	2	3	6
$\lceil \rho(1/4)P \rceil$	2	3	5	8

Références

- [1] J.D. Blanchard, C. Cartis, and J. Tanner. The restricted isometry property and ℓ^q -regularization : Phase transitions for sparse approximation. *Preprint*, 2009.
- [2] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Info. Theory*, 52(2) :489–509, 2006.
- [3] D. Donoho. Compressed sensing. *IEEE Trans. Info. Theory*, 52(4) :1289–1306, 2006.
- [4] D. L. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete & Computational Geometry*, 35(4) :617–652, 2006.
- [5] C. Dossal. A necessary and sufficient condition for exact recovery by ℓ_1 minimization. *Preprint*, 2007.
- [6] C. Dossal, G. Peyré, and J. Fadili. Challenging restricted isometry constants with greedy pursuit. *Preprint Hal-00373450*, 2009.
- [7] C. Dossal, G. Peyré, and J. Fadili. A numerical exploration of compressed sampling recovery. *Proceedings of SPARS'09*, 2009.
- [8] S. Foucart and M.-J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q -minimization for $0 < q \leq 1$. *to appear in Applied and Computational Harmonic Analysis*, 2009.
- [9] J.-J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Info. Theory*, 50(6) :1341–1344, 2004.
- [10] J. A. Tropp. Just relax : convex programming methods for identifying sparse signals in noise. *IEEE Trans. Info. Theory*, 52(3) :1030–1051, 2006.