



HAL
open science

Critère d'identifiabilité pour la minimisation L1

Charles H Dossal, Gabriel Peyré

► **To cite this version:**

Charles H Dossal, Gabriel Peyré. Critère d'identifiabilité pour la minimisation L1. GRETSI'07, Sep 2007, Troye, France. pp.993-996. hal-00365959

HAL Id: hal-00365959

<https://hal.science/hal-00365959>

Submitted on 5 Mar 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.

Critère d'identifiabilité pour la minimisation ℓ^1

Charles DOSSAL¹, Gabriel PEYRÉ²

¹LaBAG / Université Bordeaux 1

¹CEREMADE / Université Paris Dauphine

charles.dossal@math.u-bordeaux1.fr, gabriel.peyre@ceremade.dauphine.fr

Résumé — Cet article propose une condition nécessaire et suffisante pour garantir que la solution d'un système linéaire sous-déterminé de norme ℓ^1 la plus faible est aussi celle de norme ℓ^0 la plus faible. La condition proposée caractérise pour la première fois de façon générique l'identifiabilité d'un vecteur pour un système linéaire donné. Ce critère a des applications potentielles pour la résolution de problèmes inverses en traitement d'images ou en compressed sensing. Outre le résultat théorique d'identifiabilité, cet article propose un algorithme glouton pour estimer de façon rapide le critère. Des résultats numériques montrent que ce critère est estimable de façon prescrite parfaite.

Abstract — This article propose a necessary and sufficient condition to ensure that the minimum ℓ^1 solution to an under-determined linear system is also the sparsest one. The proposed condition gives for the first time a generic condition for the identifiability of a vector for a given linear system. This criterion can be applied to various inverse problem in image processing and shed light on the practical efficiency of compressed sensing. This article also explains a greedy algorithm to estimate in practice this criterion. Numerical results show that this greedy procedure is indeed able to predict recovery exactly.

1 Minimisation ℓ^1

La minimisation sous contrainte ℓ^1 est un problème d'optimisation désormais classique en traitement du signal, voir par exemple [CDS98, Fuc04, Tro06, GN03] pour des travaux récents. Cette contrainte de parcimonie est bien adaptée à la description des sons et des images naturelles qui ont une représentation creuse dans des bases de cosinus locaux ou d'ondelettes, voir [Mal98]. La résolution de problèmes inverses tels que la déconvolution, la tomographie ou encore le compressed sensing [Don06, CRT04] nécessite ainsi l'optimisation conjointe d'une attache aux données quadratique et d'une pénalisation non-lisse ℓ^1 . Cet article étudie les minimiseurs de ces fonctionnelles en détaillant une condition nécessaire et suffisante pour prédire si le vecteur cherché est retrouvé par l'optimisation.

1.1 Notations

Dans la suite, on note $I(v) \stackrel{\text{def.}}{=} \{i \mid v_i \neq 0\}$ le support d'un vecteur v . La restriction d'un vecteur à son support est $\bar{v} = (v_i)_{i \in I(v)}$. On note $A_I = (a_j)_{j \in I}$ une sous matrice extraite de A . Lorsqu'il n'y a pas de confusion possible, on note $A_I \stackrel{\text{def.}}{=} A_{I(x_0)}$ la matrice correspondant aux colonnes actives de x_0 . La pseudo inverse B^\dagger d'une matrice B est $B^\dagger \stackrel{\text{def.}}{=} (B^T B)^{-1} B^T$.

1.2 Système linéaire sous contrainte ℓ^1

Étant donnée une matrice A de taille $n \times p$ avec $n \ll p$ dont les colonnes sont notées $\{a_j\}_{j=1}^p$, on souhaite retrouver un vecteur x_0 à partir de mesures $y = Ax_0$. Une hy-

pothèse classique consiste à supposer que x_0 est *creux*, c'est-à-dire qu'il a un petit nombre s de composantes non nulles, ce que l'on note $\|x_0\|_{\ell^0} \leq s$. On est ainsi amené à résoudre

$$\min_x \|x\|_{\ell^0} \quad \text{sous la contrainte} \quad Ax = y. \quad (1)$$

Ce problème combinatoire est algorithmiquement insoluble et on considère à la place la solution de

$$\min_x \|x\|_{\ell^1} \quad \text{sous la contrainte} \quad Ax = y, \quad (P_1(y))$$

où $\|v\|_{\ell^1} \stackrel{\text{def.}}{=} \sum_i |v_i|$.

Définition 1. *Le vecteur x_0 est dit identifiable si il est l'unique solution de $P_1(Ax_0)$.*

Pour prendre en compte un bruit additionnel lors des mesures $y = Ax_0 + w$, avec $\|w\|$ de l'ordre de λ , on préfère en général résoudre une version relaxée de $P_1(y)$

$$\min_x \|y - Ax\|^2 + \lambda \|x\|_{\ell^1}. \quad (P_1(y, \lambda))$$

Ces problèmes d'optimisation $P_1(y)$ et $P_1(y, \varepsilon)$ peuvent se résoudre de façon rapide à l'aide d'algorithmes de programmation convexe [CDS98].

On remarque que la solution $x(\lambda)$ de $P_1(Ax_0, \lambda)$ converge vers la solution de $P_1(Ax_0)$ lorsque $\lambda \rightarrow 0$, quand les solutions de ces problèmes sont uniques, de sorte que l'identifiabilité correspond à la convergence de $x(\lambda)$ vers x_0

1.3 Conditions d'identifiabilité

Cohérence. Sous des conditions plus ou moins restrictives sur A et sur x_0 , de nombreux travaux permettent

d'assurer l'identifiabilité de x . La condition la plus simple requiert que les paires de vecteurs a_i, a_j soient les plus orthogonaux possibles [Tro06, CDS98]

$$C_A(x_0) \stackrel{\text{def.}}{=} \frac{1}{2} \left(1 + \frac{1}{\mu(A)} \right) - \|x_0\| > 0 \quad (\text{C})$$

implique que x_0 est identifiable, où la cohérence est définie par $\mu(A) \stackrel{\text{def.}}{=} \max_{i \neq j} \left| \frac{\langle a_i, a_j \rangle}{\|a_i\|_{\ell_2} \|a_j\|_{\ell_2}} \right|$. Le critère (C) est cependant grossier car il ne prend en compte que le cardinal du support $I(x_0)$ de x_0 .

Critère de support. Une condition plus fine que (C) est proposée par Tropp [Tro06], voir également [GN03, Fuc04]. Sous l'hypothèse que A_I est de rang maximal, on a

$$E_A(x_0) \stackrel{\text{def.}}{=} 1 - \max_{j \notin I(x_0)} \|A_I^\dagger a_j\|_{\ell^1} > 0 \quad (\text{E})$$

implique que x_0 est identifiable. La condition (E) ne dépend que du support de x_0 et assure de plus la convergence des supports : $I(x(\lambda)) \rightarrow I(x_0)$.

Critère de support signé. Afin de prendre en compte le signe des composantes de x_0 , Fuchs [Fuc04] propose, sous l'hypothèse que A_I est de rang maximal, le critère suivant

$$F_A(x_0) \stackrel{\text{def.}}{=} 1 - \max_{j \notin I(x_0)} |\langle A_I^\dagger a_j, \text{sign}(x_0) \rangle| > 0 \quad (\text{F})$$

implique que x_0 est identifiable. Dans cet article, nous montrons comment on peut relâcher ce critère de convergence des supports pour obtenir une condition générique d'identifiabilité.

Compressed sensing. La théorie du compressed sensing [Don06, CRT04] propose une autre gamme de critères d'identifiabilité sous des hypothèses contraignantes sur A . Sous ces hypothèses, on peut montrer que x_0 est identifiable avec grande probabilité dès lors que $\|x_0\|_{\ell^0} \leq Cn/\log(p)$. Cette borne est nettement meilleure que celle obtenue par la cohérence (C), mais dans la pratique, seules des matrices générées aléatoirement vérifient les hypothèses du compressed sensing.

2 Condition nécessaire et suffisante d'identifiabilité

La convergence du support $I(x(\lambda)) \rightarrow I(x_0)$ impliquée par les critères C, E et F est une contrainte forte car l'identifiabilité ne nécessite qu'une convergence $x(\lambda) \rightarrow x_0$. Le critère d'identifiabilité développé dans cet article étend le critère \mathcal{F} de l'équation (F).

Définition 2. On note $\mathcal{K} = \bar{\mathcal{F}}$ l'adhérence des vecteurs satisfaisant le critère de \mathcal{F}

$$\mathcal{F} \stackrel{\text{def.}}{=} \{x_0 \mid F(x_0) > 0\}. \quad (2)$$

Cet ensemble \mathcal{K} est formé d'un ensemble de cônes sur lesquels le support et le signe des vecteurs restent constants. L'ensemble \mathcal{K} est ainsi formé des vecteurs x_0 qui peuvent

être étendus en un vecteur de \mathcal{F} , c'est à dire, $x_0 \in \mathcal{K}$ si et seulement s'il existe x_2 avec $I(x_2) \cup I(x_0) = \emptyset$ tel que $x_0 + x_2 \in \mathcal{F}$.

Cet article étudie en détail le résultat suivant [Dos07].

Theorème 1. Si $x \in \mathcal{K}$ alors x est identifiable.

La condition réciproque du théorème 1 nécessite une condition additionnelle de non-dégénérescence.

Définition 3. A satisfait la condition (UC) si pour tout $I \subset \{1, \dots, p\}$, tel que les vecteurs $(a_i)_{i \in I}$ sont linéairement indépendants, pour tout indice $j \notin I$, pour tout vecteur $S \in \{-1, 1\}^{|I|}$,

$$|\langle a_j, (A_I^\dagger)^T S \rangle| \neq 1 \quad (\text{UC})$$

Cette condition n'est pas très contraignante et est particulièrement vérifiée presque sûrement pour des matrices A tirées aléatoirement selon une loi à densité sur l'ensemble des matrices. Elle permet d'énoncer le théorème suivant [Dos07].

Theorème 2. Si A vérifie la condition (UC), alors x_0 identifiable implique que $x \in \mathcal{K}$.

On peut également montrer que sous ces conditions, l'application qui à $y \in \text{Im}(A)$ associe l'unique minimiseur de $P_1(y)$ est uniformément Lipschitz, voir [Dos07] pour une preuve complète.

2.1 Démonstration du théorème 1

Pour démontrer le théorème 1 on montre d'abord que tout vecteur x de \mathcal{F} est identifiable, c'est à dire est l'unique minimiseur de $P_1(Ax_0)$. Pour cela on montre qu'il existe $y_1 \in \text{Im}(A)$ et $\varepsilon > 0$ tels que x_0 est l'unique solution de $P_1(y_1, \varepsilon)$. On en déduit alors que x_0 est le vecteur de norme ℓ_1 minimum parmi les vecteurs vérifiant $\|y - Ax\|_2 \leq \|y - Ax_0\|_2$ et donc parmi les éléments vérifiant $Ax = Ax_0$.

Soit $x_0 \in \mathcal{F}$ et $\varepsilon > 0$ tel que

$$\text{sign}(\bar{x}_0) = \text{sign}(\bar{x}_0 + \varepsilon(A_I^T A_I)^{-1} \text{sign}(\bar{x}_0)).$$

On définit x_1 comme étant le vecteur ayant même support que x_0 tel que

$$\bar{x}_1 = \bar{x}_0 + \varepsilon(A_I^T A_I)^{-1} \text{sign}(\bar{x}_0) \quad \text{et} \quad y_1 = Ax_1.$$

On utilise ensuite le lemme d'optimisation suivant, voir Fuchs [Fuc04]

Lemme 1. Soit x_0 un vecteur et I son support. Si x_0 vérifie les trois conditions suivantes

- La matrice A_I associée à x_0 est de rang maximum.
- $A_I^T(y_1 - Ax_0) = \varepsilon \text{sign}(\bar{x}_0)$
- $\forall j \notin I, |\langle a_j, y_1 - Ax_0 \rangle| < \varepsilon$.

alors x_0 est l'unique solution de $P_1(y_1, \varepsilon)$.

Comme les deux premières conditions du lemme 1 sont vérifiées par définition de x_0 et que $\forall j \notin I, |\langle a_j, y_1 - Ax_0 \rangle| \leq \varepsilon F(x_0) < \varepsilon$, on déduit que x_0 est l'unique minimiseur de $P_1(y_1, \varepsilon)$.

Supposons maintenant que $x_0 \in \mathcal{K}$, il existe x_2 dont le support est disjoint de x_0 est tel que $x_0 + x_2 = x_1 \in \mathcal{F}$.

Soit x_3 tel que $Ax_3 = Ax_0$ et $\|x_3\|_{\ell^1} \leq \|x_0\|_{\ell^1}$, on pose $x_4 = x_3 + x_2$. On a ainsi $Ax_4 = Ax_1$ et

$$\|x_4\|_{\ell^1} \leq \|x_3\|_{\ell^1} + \|x_2\|_{\ell^1} \leq \|x_0\|_{\ell^1} + \|x_2\|_{\ell^1} = \|x_1\|_{\ell^1}.$$

Comme $x_1 \in \mathcal{F}$, x_1 est l'unique solution de $P_1(Ax_1)$ et donc $x_4 = x_1$ et ainsi $x_3 = x_0$ et x_0 est l'unique solution de $P_1(Ax_0)$.

2.2 Démonstration du théorème 2

Nous proposons ici une synthèse de la démonstration dont une version complète figure dans [Dos07]. Soit $y \in \text{Im}(A)$, nous allons montrer qu'il existe un antécédent x_0 de y par A dans \mathcal{K} . On utilise ensuite le théorème 1 pour conclure.

Soit (γ_n) une suite de réels strictement positifs tendant vers 0 et $x(\gamma_n)$ une suite de solutions de $P_1(y, \gamma_n)$. Cette suite n'est a priori pas unique mais on peut en choisir une. Quitte à extraire une sous suite, on peut supposer que la suite $x(\gamma_n)$ tend vers un vecteur x_0 . Comme $x(\gamma_n)$ est une solution de $P_1(y, \gamma_n)$, on déduit que $y = Ax_0$. Soit n_0 tel que $\forall n \geq n_0$ on ait $I(x_0) \subset I(x(\gamma_n))$. Soit $n \geq n_0$, on note $I = I(x(\gamma_n))$.

On rappelle ici un second lemme d'optimisation qui permet de caractériser une solution de $P_1(y, \gamma)$, pour y et γ quelconque, voir Fuchs [Fuc04].

Lemme 2. *Si $x(\gamma)$ est une solution du problème $P_1(y, \gamma)$ alors $x(\gamma)$ vérifie*

$$A_I^T(y - Ax(\gamma)) = \gamma \text{sign}(\bar{x}(\gamma)) \quad (3)$$

$$\forall j \notin I, \quad |\langle a_j, y - Ax(\gamma) \rangle| \leq \gamma. \quad (4)$$

Ce lemme s'applique donc à $x(\gamma_n)$. On utilise le fait que A vérifie la condition (UC) pour montrer que la matrice active A_I associée à $x(\gamma_n)$ est de rang maximum. Pour ce faire on suppose que la matrice A_I n'est pas de rang plein et on considère $J \subset I$ et $k \in I \setminus J$ tel que $|J| = \text{rang}(A_J) = \text{rang}(A_I)$. On a ainsi $a_k \in \text{Vect}(a_j)_{j \in J}$ et donc $a_k = A_J A_J^\dagger a_k$. De plus (3) implique que

$$A_J^T(y - Ax(\gamma_n)) = \gamma_n \text{sign}(\bar{x}_J(\gamma_n)).$$

où $\bar{x}_J(\gamma_n)$ est le vecteur extrait de $x(\gamma_n)$ formé des composantes indicées par J . De (3) on déduit également que

$$\begin{aligned} \gamma_n &= |\langle a_k, y - Ax(\gamma_n) \rangle| \\ &= |\langle A_J A_J^\dagger a_k, y - Ax(\gamma_n) \rangle| \\ &= |\langle a_k, A_J^T A_J^\dagger (y - Ax(\gamma_n)) \rangle| \\ &= |\gamma_n \langle a_k, A_J^T \text{sign}(\bar{x}_J(\gamma_n)) \rangle| \end{aligned}$$

et donc que $|\langle a_k, A_J^T \text{sign}(\bar{x}_J(\gamma_n)) \rangle| = 1$ ce qui est impossible si A vérifie la condition (UC). Donc A_I est de rang maximum.

On déduit de (3) que

$$\bar{x}(\gamma_n) = A_I^\dagger y - \gamma_n (A_I^T A_I)^{-1} \text{sign}(\bar{x}(\gamma_n)).$$

On a alors que pour tout $j \notin I$,

$$\begin{aligned} \langle a_j, y - Ax(\gamma_n) \rangle &= \\ \langle a_j, y - A_I A_I^\dagger y - \gamma_n A_I (A_I^T A_I)^{-1} \text{sign}(\bar{x}(\gamma_n)) \rangle &= \end{aligned}$$

Comme $I(x_0) \subset I(x(\gamma_n))$, on a $\bar{x}_0 = A_I^\dagger A_I \bar{x}_0$ et donc $A_I A_I^\dagger y = A_I A_I^\dagger A_I \bar{x}_0 = A_I \bar{x}_0 = y$ et ainsi

$$\langle a_j, y - Ax(\gamma_n) \rangle = -\gamma_n \langle a_j, A_I (A_I^T A_I)^{-1} \text{sign}(\bar{x}(\gamma_n)) \rangle.$$

De (4) on déduit que

$$|\langle a_j, A_I (A_I^T A_I)^{-1} \text{sign}(\bar{x}(\gamma_n)) \rangle| \leq 1.$$

Comme A vérifie (UC),

$$|\langle a_j, A_I (A_I^T A_I)^{-1} \text{sign}(\bar{x}(\gamma_n)) \rangle| \neq 1$$

et donc

$$|\langle a_j, A_I (A_I^T A_I)^{-1} \text{sign}(\bar{x}(\gamma_n)) \rangle| < 1$$

et on déduit du lemme 1 que $x(\gamma_n)$ est l'unique solution de $P_1(y, \gamma_n)$ et surtout que $x(\gamma_n) \in \mathcal{F}$. Ainsi $\forall n \geq n_0$, $x(\gamma_n) \in \mathcal{F}$ et donc $x_0 \in \mathcal{K}$.

3 Critère d'identifiabilité

3.1 Critère générique d'identifiabilité

Les théorèmes 1 et 2 montrent que, sous la condition (UC) sur la matrice A , le critère suivant

$$F_A^+(x_0) \stackrel{\text{def}}{=} \max_{x \in I(x_0) \subset I(x)} F_A(x) > 0 \quad (\text{F}^+)$$

est équivalent à l'identifiabilité de x_0 .

3.2 Algorithme d'estimation du critère

Le calcul de $F_A^+(x_0)$ est algorithmiquement impossible, c'est pourquoi nous proposons un algorithme glouton pour trouver un vecteur x_1 tel que $I(x_0) \subset I(x_1)$ et $F_A(x_1) > 0$. Cet algorithme approché n'est pas assuré de toujours trouver un tel vecteur, mais les expériences numériques de la section 4 prouvent son efficacité en pratique. L'algorithme consiste à augmenter progressivement le support de x_0 en ajoutant des colonnes a_j inactives maximisant une corrélation signée.

(Initialisation) $x_1 \leftarrow x_0$, $I \leftarrow I(x_1)$. (Boucle) Tant que la matrice A_I n'est pas de rang maximum et que $F(x_1) > 1$,

– Calculer $j_0 = \arg \max_{j \notin I} |(A_I^\dagger a_j)^T \text{sign}(\bar{x}_1)|$ et mettre à jour

$$x_1(j_0) \leftarrow \text{sign} \left((A_I^\dagger a_{j_0})^T \text{sign}(\bar{x}_1) \right) \quad \text{et} \quad I \leftarrow I(x_1)$$

– Soit le vecteur

$$\bar{v} \stackrel{\text{def}}{=} (A_I^T A_I)^{-1} \text{sign}(\bar{x}_1).$$

Le vecteur x_0 est mis à jour

$$\forall k \in I \setminus I(x_0), \quad x_1(k) \leftarrow 0 \quad \text{si} \quad v(k)x_1(k) > 0.$$

Le support est mis à jour : $I \leftarrow I(x_1)$.

La valeur maximale de $F_A(x_1)$ calculée pendant l'algorithme est une bonne approximation de $F_A^+(x_0)$ et c'est cette quantité que nous avons utilisé dans les expériences numériques.

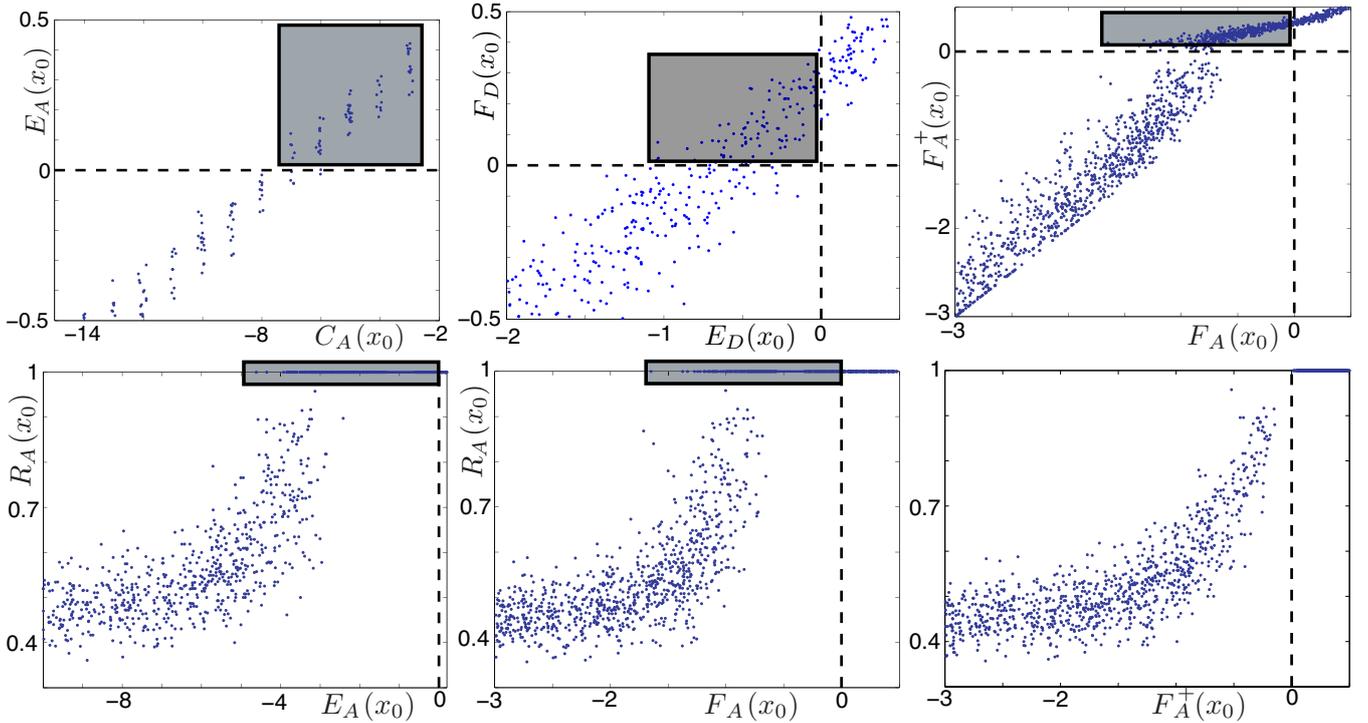


FIG. 1 – *Première ligne* : comparaison par paires des critères (C), (E), (F) et (F^+). La partie grisée représente les x_0 pour lesquelles le critère en abscisse n'arrive pas à prévoir l'identifiabilité mais le critère en ordonnée y arrive. *Deuxième ligne* : mesure de l'identifiabilité $R_A(x_0)$ pour les critères (E), (F) et (F^+). La partie grisée représente les x_0 pour lesquels le critère n'a pas été capable de caractériser l'identifiabilité.

4 Résultats numériques

Nous avons testé les différents critères d'identifiabilité énoncés, qui peuvent se classer par degré de finesse croissante $(C) \Rightarrow (E) \Rightarrow (F) \Rightarrow (F^+)$. Pour se faire, nous avons fixé une taille de problèmes ($n = 300, p = 1200$) et nous avons généré aléatoirement des matrices A dont les colonnes sont tirées uniformément sur la sphère de \mathbb{R}^n . Pour chaque matrice, nous avons tiré aléatoirement un entier $1 \leq s \leq 150$ puis un vecteur x_0 de signes aléatoires tel que $\#I(x_0) = s$.

Pour une matrice A et un vecteur x_0 , on note x^* la solution de $P_1(y)$. On mesure l'identifiabilité de x_0 par

$$R_A(x_0) \stackrel{\text{def.}}{=} 1 - \frac{1}{2s} \|\text{sign}(\bar{x}^*) - \text{sign}(x_0)\|_{\ell_0}$$

où \bar{x} est le vecteur issu de x en mettant à zéro les composantes n'étant pas parmi les s plus grandes en valeur absolue. On a $0 \leq R_A(x_0) \leq 1$ et $R_A(x_0) = 1$ correspond à un vecteur identifiable. La quantité $F_A^+(x_0)$ est estimée à l'aide de l'algorithme glouton approché.

Chaque point de la figure 1 correspond à un triplet (A, s, x_0) tiré aléatoirement. La figure (1), première ligne, compare les critères entre eux. Ceci montre bien que chaque critère mis en ordonné permet caractériser plus finement l'identifiabilité que le critère en abscisse. En particulier, on a systématiquement $F_A^+(x_0) > F_A(x_0)$, ce qui montre que ce nouveau critère (F^+) est nettement plus fin que (F). La figure 1, deuxième ligne, confronte les critères (E), (F) et (F^+) à $R_A(x_0)$. On constate en particulier que seul le critère (F^+) caractérise de façon parfaite l'identifiabilité $R_A(x_0)$.

Références

- [CDS98] S. CHEN, D. DONOHO et M. SAUNDERS : Atomic decomposition by basis pursuit. *SIAM J. Sci. Comp.*, 20(1):33–61, 1998.
- [CRT04] E. CANDÈS, J. ROMBERG et T. TAO : Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 2004. Submitted.
- [Don06] D. DONOHO : Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.
- [Dos07] C. DOSSAL : A necessary and sufficient condition for exact recovery by ℓ^1 minimization. *preprint*, 2007.
- [Fuc04] J.-J. FUCHS : On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344, 2004.
- [GN03] R. GRIBONVAL et M. NIELSEN : Sparse decompositions in unions of bases. *IEEE Transactions on Information Theory*, 49(12):3320–3325, 2003.
- [Mal98] S. MALLAT : *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 1998.
- [Tro06] J. A. TROPP : Just relax : convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006.