



## Theory of Signals

A necessary and sufficient condition for exact sparse recovery by  $\ell_1$  minimization*Une condition nécessaire et suffisante d'identifiabilité parcimonieuse par minimisation  $\ell_1$* 

Charles Dossal

IMB, Université Bordeaux 1, 351, Cours de la Libération, 33405 Talence cedex, France

## ARTICLE INFO

## Article history:

Received 18 November 2011

Accepted after revision 21 December 2011

Available online 9 January 2012

Presented by Yves Meyer

## ABSTRACT

In this Note, a new sharp sufficient condition for exact sparse recovery by  $\ell_1$ -penalized minimization from linear measurements is proposed. The main contribution of this paper is to show that, for most matrices, this condition is also necessary. Moreover, when the  $\ell_1$  minimizer is unique, we investigate its sensitivity to the measurements and we establish that the application associating the measurements to this minimizer is Lipschitz-continuous.

© 2012 Published by Elsevier Masson SAS on behalf of Académie des sciences.

## R É S U M É

Dans cette Note, une nouvelle condition suffisante pour l'identifiabilité parcimonieuse par minimisation  $\ell_1$  pénalisée à partir de mesures linéaires est proposée. La contribution majeure de ce travail est de prouver que pour la plupart des matrices, cette condition est aussi nécessaire. Par ailleurs, lorsque le minimiseur du problème  $\ell_1$  est unique, sa sensibilité aux mesures est étudiée et il est montré que l'application qui envoie les mesures sur ce minimiseur est Lipschitz-continue.

© 2012 Published by Elsevier Masson SAS on behalf of Académie des sciences.

## 1. Introduction

Let  $x^0 \in \mathbb{R}^N$  be a vector and  $A$  be a real matrix with  $n$  rows and  $N$  columns. Let  $y^0$  be  $n$  linear measurements of  $x^0$ , i.e.  $y^0 = Ax^0 \in \mathbb{R}^n$ , where typically  $n < N$ . In this paper, we propose a *necessary and sufficient* condition ensuring that  $x^0$  can be recovered from  $y^0$  by solving the following optimization problem  $P_1(y^0)$

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{such that } y = Ax. \quad P_1(y)$$

There is of course a huge literature on the subject, and covering it fairly is beyond the scope of this paper. We restrict our overview to those works pertaining to ours. For instance, our new condition can be seen as an extension of [6] and [7]. The geometric interpretations of results are related to the geometry of polytopes previously used by Donoho [2] and Donoho and Tanner [3] to study  $\ell_1$  minimization. Indeed, in [6], the following optimization problem (the so-called Lasso problem)

$$\min_{x \in \mathbb{R}^N} \frac{1}{2} \|y - Ax\|_2^2 + \gamma \|x\|_1 \quad P_1(y, \gamma)$$

E-mail address: charles.dossal@math.u-bordeaux1.fr.

is studied. It is proved that any  $x^0$  belonging to the set  $\mathcal{F}$  is the unique solution of  $P_1(Ax^0)$ , where

$$\mathcal{F} = \{x \text{ such that } \text{rank}(A_I) = |I| \text{ and } \forall j \notin I, |(a_j, A_I(A_I^t A_I)^{-1} \text{sign}(x_I))| < 1\},$$

where  $I$  is the support of  $x$ ,  $A_I$  the active matrix associated to  $x$ , whose columns are those of  $A$  indexed by  $I$ ,  $x_I$  are the non-zero components of  $x$  and  $a_j$  is the column of  $A$  indexed by  $j$ . In the sequel,  $\text{Span}(a_i)_{i \in I}$  will be denoted  $V_I$ . The sufficient condition developed in [6] plays a pivotal role in several papers which investigate support recovery in presence of noise by solving  $P_1(y, \gamma)$ , e.g. [1,8,5] random vectors [1] or random measurements [5].

In [7], the Exact Recovery Condition (ERC) is defined. This condition does not depend on the sign but only on the support of  $x^0$  and provides results about the recovery of  $x^0$  and stability to noise.

The goal of this Note is to show that the set of vectors that can be recovered by  $\ell_1$  minimization is exactly, for most matrices,  $\overline{\mathcal{F}}$  the closure of  $\mathcal{F}$ . This result highlights the fact that

$$\max_{j \notin I} |(a_j, A_I(A_I^t A_I)^{-1} \text{sign}(x_I))| \quad \text{and} \quad \|A_I(A_I^t A_I)^{-1} \text{sign}(x_I)\|_2$$

are good indicators of the identifiability, and brings arguments justifying the success of algorithms developed in [4] to find very sparse but non-identifiable vectors.

Before proceeding, let us fix some terminology and definitions. A vector  $x^0$  is said *identifiable* if and only if it is the unique solution of  $P_1(Ax^0)$ .

**Definition 1.1.** Let  $(x^i)_{i \leq N}$  be  $N$  points of  $\mathbb{R}^n$ . These points  $(x^i)_{i \leq N}$  are said in *general position (GP)* if all affine subspaces of  $\mathbb{R}^n$  which dimension  $k < n$  contain at most  $k + 1$  points  $x^i$ . A matrix  $A$  satisfies condition (GP) if for all sign vector  $S \in \{-1, 1\}^N$ , points  $(S[i]a_i)_{i \leq N}$  are in general position.

It can be noticed that for any matrix  $A$ , the matrix  $A + E$  satisfies condition (GP) with probability 1 if  $E$  is any random perturbation with an absolutely continuous density with respect to the Lebesgue measure: with this definition most matrices satisfy condition (GP).

## 2. Contributions

The contributions of this paper are summarized as follows:

**Theorem 2.1.** *If  $x^0 \in \overline{\mathcal{F}}$ , then  $x^0$  is identifiable.*

**Theorem 2.2.**

- (i) *If  $x^0$  is identifiable, and if for all  $y$  in a neighborhood of  $y^0 = Ax^0$   $P_1(y)$  has only one solution, then  $x^0 \in \overline{\mathcal{F}}$ .*
- (ii) *If  $A$  satisfies (GP), for all  $y \in \text{Im}(A)$ ,  $P_1(y)$  has a unique solution and  $x$  is identifiable if and only if  $x \in \overline{\mathcal{F}}$ .*

**Theorem 2.3.** *If for all  $y \in \text{Im}(A)$  the solution of  $P_1(y)$  is unique, the application  $\phi$  associating  $y$  to this solution is Lipschitz.*

## 3. Preliminary lemmas

The two following lemmas can be found in [6].

**Lemma 3.1.** *A vector  $x^*$  is a solution of  $P_1(y, \gamma)$  if and only if*

$$A_I^t(y - Ax^*) = \gamma \text{sign}(x_I^*) \quad \text{and} \quad \forall j \notin I \quad |(a_j, y - Ax^*)| \leq 1$$

where the support of  $x^*$  is denoted by  $I$ .

**Lemma 3.2.** *If for a vector  $x^*$  the matrix  $A_I$  satisfies  $\text{rank}(A_I) = |I|$  and*

$$A_I^t(y - Ax^*) = \gamma \text{sign}(x_I^*) \quad \text{and} \quad \forall j \notin I, |(a_j, A_I(A_I^t A_I)^{-1} \text{sign}(x_I^*))| < 1$$

then  $x^*$  is the unique minimizer of  $P_1(y, \gamma)$ .

**Lemma 3.3.** *If  $x^0$  is the unique solution of  $P_1(Ax^0)$  then  $\text{rank}(A_I) = |I|$ .*

**Proof.** If there exists  $h \in \text{Ker}(A)$  that is supported on  $I = I(x^0)$ , one has for all  $t \in \mathbb{R}$ ,  $A(x^0 + th) = Ax^0$ . Consequently the application  $t \mapsto \|x^0 + th\|_1$  is locally affine in a neighborhood of 0. It follows that there exists  $t \neq 0$  such that  $\|x^0 + th\|_1 \leq \|x^0\|_1$ .  $\square$

**Lemma 3.4.** For all  $y \in \text{Im}(A)$  there exists a solution  $x$  to  $P_1(y)$  such that  $\text{rank}(I) = |I|$ .

**Proof.** Let  $x^0$  be a solution of  $P_1(y)$ . If  $\text{rank}(I) < |I|$ , there exists  $h \in \text{Ker}(A)$  that is supported on  $I$ . For a suitable  $t$ ,  $I(x^0 + th) \subsetneq I(x^0)$  and  $x^0 + th$  is a solution of  $P_1(y)$ .  $\square$

**Lemma 3.5.** If  $x^0 \in \mathcal{F}$ , then  $x^0$  is the unique solution of  $P_1(Ax^0)$ .

**Proof.** Lemma 3.2 shows that  $x^0$  is the unique solution of  $P_1(A_I(x_I^0 + \gamma(A_I^t A_I)^{-1} \text{sign}(x_I^0)), \gamma)$  if  $\gamma > 0$  is small enough.  $\square$

**Lemma 3.6.** Let  $y \in \text{Im}(A)$  and  $(\gamma_n)_{n \in \mathbb{N}}$  a sequence of positive real numbers tending to 0. If  $(x^n)_{n \in \mathbb{N}}$  are some solutions of  $P_1(y, \gamma_n)$  such that  $\lim_{n \rightarrow \infty} x^n = x^0$  one gets that  $x^0$  is a solution of  $P_1(y)$ .

## 4. Proofs

### 4.1. Proof of Theorem 2.1

From Lemma 3.5 one has that vectors in  $\mathcal{F}$  are identifiable. Since  $A_I(A_I^t A_I)^{-1} \text{sign}(x_I)$  only depends on the sign and the support of  $x$ ,  $\mathcal{F}$  is a union of cones of various dimensions. It follows that the closure  $\overline{\mathcal{F}}$  of  $\mathcal{F}$  is exactly equal to the set of vectors  $x^0$  that can be extended into a vector of  $\mathcal{F}$ :

$$\overline{\mathcal{F}} = \{x^0 \text{ such that } \exists x^1 \text{ such that } I(x^0) \cap I(x^1) = \emptyset \text{ and } x^0 + x^1 = x^2 \in \mathcal{F}\}.$$

Let us now suppose that  $x^0 \in \overline{\mathcal{F}}$ , and that there exists  $x^1$  such that  $I(x^0) \cap I(x^1) = \emptyset$  and  $x^2 = x^0 + x^1$  belongs to  $\mathcal{F}$ .

Choose  $x^3 \in \mathbb{R}^N$  such that  $Ax^0 = Ax^3$  and define  $x^4 := x^3 + x^1$ . We have  $Ax^4 = Ax^2$  and since  $x^2 \in \mathcal{F}$ ,  $x^2$  is the unique solution of  $P_1(Ax^2)$  which implies that  $\|x^2\|_1 < \|x^4\|_1$ . It follows that

$$\|x^2\|_1 = \|x^0\|_1 + \|x^1\|_1 < \|x^4\|_1 \leq \|x^3\|_1 + \|x^1\|_1, \quad (1)$$

which implies  $\|x^0\|_1 < \|x^3\|_1$ . That is,  $x^0$  is the unique solution of  $P_1(Ax^0)$ .

### 4.2. Proof of Theorem 2.2

**Proof.** Let  $y \in \text{Im}(A)$  and  $(\gamma_n)_{n \in \mathbb{N}}$  a sequence of positive real numbers tending to 0. From Lemma 3.4 we can always choose a solution  $x(\gamma_n)$  of  $P_1(y, \gamma)$  such that the associated active matrix  $A_{I_n}$  has a full rank. Since the number of possible supports and signs is finite, up to an extraction of a subsequence, we can suppose that all  $x(\gamma_n)$  share the same support  $I$  and the same sign  $S$ . From Lemma 3.1, we know that  $x(\gamma_n)$  satisfies

$$x(\gamma_n)_I = x_I^0 - \gamma_n (A_I^t A_I)^{-1} S \quad \text{and} \quad \forall j \notin I \quad |(a_j, y - Ax(\gamma_n))| \leq \gamma_n \quad (2)$$

where  $x^0$  is the vector supported on  $I$  such that  $x_I^0 = (A_I^t A_I)^{-1} A_I^t y$ . Using Lemma 3.6,  $x^0$  is a solution of  $P_1(y)$ . From (2) we deduce that  $I(x^0) \subset I$  for  $\gamma_n$  small enough. It follows that  $y - Ax(\gamma_n) = \gamma_n A_I (A_I^t A_I)^{-1} S = \gamma_n d_{I,S}$  and thus  $\forall j \notin I$ ,  $|(a_j, d_{I,S})| \leq 1$ .

Let us define  $J$  to be the set of all indices such that  $|(a_j, d_{I,S})| = 1$ . One can first notice that  $I \subset J$ .

We now show that

- (i) either  $(a_j)_{j \in J}$  are linearly dependent and one can build vectors  $x^1$  close to  $x^0$  such that  $P_1(x^1)$  has several solutions, in which case condition (GP) cannot be satisfied;
- (ii) or  $(a_j)_{j \in J}$  are linearly independent and  $x(\gamma_n) \in \overline{\mathcal{F}}$  and  $x^0 \in \overline{\mathcal{F}}$  and  $x^0$  is identifiable.

(i) Suppose that  $(a_j)_{j \in J}$  are linearly dependent and let  $\varepsilon > 0$ . Define  $K \subsetneq J$  such that  $I \subset K$  and  $(a_k)_{k \in K}$  is a basis of  $V_J = \text{Span}(a_j)_{j \in J}$  and  $j_0 \in J \cap K^c$ . Define  $S_K \in \mathbb{R}^{|K|}$  by  $S_K = A_K^t d_{I,S}$ . Then,  $S_K$  is a sign vector and  $x^1$  the vector supported on  $K$  defined by

$$x_K^1 = x_K^0 + \frac{\varepsilon}{\|S_K\|_2} S_K.$$

Now choose  $\varepsilon > 0$  small enough to ensure that  $\text{sign}(x_{I(x_0)}^1) = \text{sign}(x_{I(x_0)}^0)$ . From definition of  $x^1$ , it follows that  $\|x^0 - x^1\|_2 = \varepsilon$ . For  $\gamma$  small enough the support and the sign of  $x^1$  and the vector  $x^1(\gamma)$  supported in  $K$  and defined by  $x^1(\gamma)_K = x_K^1 - \gamma (A_K^t A_K)^{-1} S_K$  are identical.

We shall prove that  $x^1(\gamma)$  is a solution of  $P_1(Ax^1, \gamma)$  using Lemma 3.1. Indeed  $A_K^t (Ax^1 - Ax^1(\gamma)) = \gamma S_K$ . The vector  $S = A_I^t d_{I,S}$  is the common sign of vectors  $(x(\gamma_n))_{n \in \mathbb{N}}$ . Since  $\lim_{n \rightarrow \infty} x(\gamma_n) = x^0$  it follows that for all  $i \in I(x^0) \subset I$ ,  $\text{sign}(x^0[i]) =$

$S[i]$ . Since for all  $i \in I$ ,  $S[i] = \langle a_i, d_{i,S} \rangle = S_K[i]$  it follows that  $S_K = \text{sign}(x_K^1)$ . Moreover  $\text{sign}(x_K^1) = \text{sign}(x^1(\gamma)_K)$  which yields that  $A_K^t(Ax^1 - Ax^1(\gamma)) = \gamma \text{sign}(x^1(\gamma)_K)$ . Since  $d_{i,S} \in V_I \subset V_K$ , one has  $d_{i,S} = A_K(A_K^t A_K)^{-1} A_K^t d_{i,S} = A_K(A_K^t A_K)^{-1} S_K$ . If  $j \notin K$ ,

$$|\langle a_j, Ax^1 - Ax^1(\gamma) \rangle| = \gamma |\langle a_j, A_K(A_K^t A_K)^{-1} S_K \rangle| = \gamma |\langle a_j, d_{i,S} \rangle| < \gamma,$$

and by Lemma 3.1 it follows that  $x^1(\gamma)$  is a solution of  $P_1(Ax^1, \gamma)$ . Now Lemma 3.6 shows that  $x^1$  is a solution of  $P_1(Ax^1)$ . Choose some  $h \in \text{Ker}(A)$  that is supported on  $K \cup \{j_0\}$  and such that  $h[j_0] = S[j_0] = a_{j_0}^t d_{i,S}$ . Denote  $h_K$  the restriction of  $h$  on indices  $K$ . Since  $0 = Ah = S[j_0]a_{j_0} + A_K h_K$  and  $S_K = A_K^t d_{i,S}$  one has  $\langle \text{sign}(x_K^1), h_K \rangle = S_K^t h_K = d_{i,S}^t A_K h_K = -S[j_0] a_{j_0}^t = -1$ . Moreover, for all  $t \in \mathbb{R}$ ,  $A(x^1 + th) = Ax^1$  and for small non-negative  $t$

$$\|x^1 + th\|_1 = \|x^1\|_1 + t + t \langle \text{sign}(x^1), h_K \rangle = \|x^1\|_1,$$

which shows that  $x^1$  is not the unique minimizer of  $P_1(Ax^1)$ .

Notice that for each  $l \in K \cup \{j_0\}$ ,  $S[l]a_l \in V_K$  and  $S[l]a_l$  belongs to the affine hyperplan  $\mathcal{H}_{d_{i,S}} = \{u, \text{ such that } \langle u, d_{i,S} \rangle = 1\}$ . Hence at least  $|K| + 1$  points  $S[l]a_l$  belong to an affine subspace  $V_K \cap \mathcal{H}_{d_{i,S}}$  which dimension equals to  $|K| - 1$ . Therefore, (GP) is not satisfied.

(ii) Let us now suppose that the  $(a_j)_{j \in J}$  are linearly independent. Define  $S_J = A_J^t d_{i,S}$ ,  $\forall i \in I$ ,  $S[i] = S_J[i]$ . Let  $x^1$  be the vector supported on  $J$  such that for all  $i \in I(x^0)$ ,  $x^1[i] = x^0[i]$  and for all  $j \in J \cap I(x^0)^c$ ,  $x^1[j] = S_J[j]$ . One has  $\text{sign}(x^1) = S_J$  and since  $d_{i,S} \in V_I \subset V_J$ , one has  $d_{i,S} = A_J(A_J^t A_J)^{-1} A_J^t d_{i,S} = A_J(A_J^t A_J)^{-1} S_J$ . It follows that for all  $l \notin J$ ,

$$|\langle a_l, A_J(A_J^t A_J)^{-1} S_J \rangle| = |\langle a_l, d_{i,S} \rangle| < 1$$

implying that  $x^1 \in \mathcal{F}$  and hence  $x^0 \in \overline{\mathcal{F}}$ .  $\square$

### 4.3. Proof of Theorem 2.3

The proof relies on the following lemma.

**Lemma 4.1.** *For all  $y^0 \in \text{Im}(A)$ , there exists  $\varepsilon_0$  such that for all  $y \in \text{Im}(A) \cap \mathcal{B}(y^0, \varepsilon_0)$ , one has  $I(\phi(y^0)) \subset I(\phi(y))$ .*

**Proof.** To prove this lemma we will prove that for any  $y^0 \in \text{Im}(A)$  and any sequence  $(y^n)_{n \in \mathbb{N}}$  tending to  $y^0$ , any subsequence  $(y^{u_n})_{n \in \mathbb{N}}$  such that  $I(\phi(y^{u_n})) = J_{u_n} = J$  is constant, the support  $J$  satisfies  $J \supset I = I(\phi(y^0))$ .

Denote by  $x^0 = \phi(y^0)$  and  $I = I(x^0)$ . One has  $x^0(I) = (A_I^t A_I)^{-1} A_I^t y^0 = A_I^+ y^0$ . Let  $(y^n)_{n \in \mathbb{N}}$  be a sequence of elements of  $\text{Im}(A)$  tending to  $y^0$  and  $x^n = \phi(y^n)$ . Up to an extraction of a subsequence, one can suppose that for all  $n \in \mathbb{N}$ ,  $I(x^n)$  is constant. Denote by  $J$  this common support. Then  $x_J^n = (A_J^t A_J)^{-1} A_J^t y^n = A_J^+ y^n$ . Let  $x^\infty = \lim_{n \rightarrow \infty} x^n$  and  $x_J^\infty = A_J^+ y^0$ . Let  $z^n$  be the sequence of vectors supported on  $I$  defined by  $z_I^n = A_I^+ y^n$ . By the definition of  $z^n$ ,  $Az^n = P_{V_I} y^n$ . Let  $K$  be a set such that  $(a_k)_{k \in K}$  is a basis of  $\text{Im}(A)$ , and  $v^n$  the vector supported on  $K$  such that  $v_K^n = A_K^+(P_{V_I^\perp}(y^n))$ . One has  $A(z^n + v^n) = y^n$  and  $\|z^n + v^n\|_1 \leq \|z^n\|_1 + \|v^n\|_1$ . Noticing that

$$\|v^n\|_1 = \|A_K^+(P_{V_I^\perp}(y^n - y^0))\|_1 \xrightarrow{n \rightarrow \infty} 0$$

one deduces that  $\lim_{n \rightarrow \infty} \|z^n + v^n\|_1 = \|x^0\|_1$ . Moreover  $\lim_{n \rightarrow \infty} \|x^n\|_1 = \|x^\infty\|_1$  and since  $\|x^n\|_1 \leq \|z^n + v^n\|_1$  we deduce that  $\|x^\infty\|_1 \leq \|x^0\|_1$  and finally  $x^0 = x^\infty$  since  $x^0 = \phi(y^0)$ . Thus one has  $I = I(x^0) = I(x^\infty) \subset J$  which concludes the proof of the lemma.  $\square$

Let  $y^0 \in \text{Im}(A)$ . There exists  $\varepsilon_0 > 0$  such that for all  $y \in \text{Im}(A) \cap \mathcal{B}(y^0, \varepsilon_0)$ , there exists  $J$  satisfying  $I \subset J$ . Moreover  $x = \phi(y)$  is supported in  $J$  and  $x_J = A_J^+ y$ . Since  $I \subset J$ , one also has  $x_I^0 = A_I^+ y^0$  and thus  $\|x - x^0\|_2 = \|A_J^+(y - y^0)\|_2$ . One deduces that  $\forall y \in \text{Im}(A) \cap \mathcal{B}(y^0, \varepsilon_0)$ ,  $\|\phi(y^0) - \phi(y)\|_2 \leq C \|y^0 - y\|_2$  with  $C = \max_{J, \text{rank}(A_J)=|J|} \lambda_{\max}(A_J^+)$  which concludes the proof of the theorem.  $\square$

### References

- [1] E.J. Candès, Y. Plan, Near ideal model selection by  $\ell_1$  minimization, *Ann. Statist.* 37 (2009) 2145–2177.
- [2] D. Donoho, Neighborly polytopes and sparse solution of underdetermined linear equations, Preprint, 2004.
- [3] D. Donoho, J. Tanner, Neighborliness of randomly-projected simplices in high dimensions, Preprint, 2005.
- [4] Ch. Dossal, G. Peyré, J. Fadili, A numerical exploration of compressed sampling recovery, *Linear Algebra Appl.* 432 (2010) 1663–1679.
- [5] Ch. Dossal, M.L. Chabanol, G. Peyré, J. Fadili, Sharp support recovery from noisy random measurements by  $\ell_1$  minimization, *Appl. Comput. Harmon. Anal.* (2011), doi:10.1016/j.acha.2011.09.003.
- [6] J.J. Fuchs, On sparse representations in arbitrary redundant bases, *IEEE Trans. Inform. Theory* 50 (2004) 1341–1344.
- [7] J.A. Tropp, Just relax: convex programming methods for identifying sparse signals in noise, *IEEE Trans. Inform. Theory* 52 (2006) 1030–1051.
- [8] Wainwright, Sharp thresholds for high-dimensional and noisy sparsity recovery using  $\ell_1$ -constrained quadratic programming (lasso), *IEEE Trans. Inform. Theory* 55 (5) (2009) 2183–2202.