A necessary and sufficient condition for exact recovery by l1 minimization.

Charles H Dossal

A necessary and sufficient condition for exact recovery by ℓ_1 minimization

Charles Dossal
LaBAG, Université Bordeaux 1,
351, cours de la Libération,
F-33405 Talence cedex (FRANCE)
charles.dossal@math.u-bordeaux1.fr

July 23, 2007

Abstract

The minimum ℓ_1-norm solution to an underdetermined system of linear equations $y = Ax$, is often, remarkably, also the sparsest solution to that system. Since the seminal work of Donoho and co-workers, we have witnessed a flurry of research activity which has focused on sufficient conditions ensuring a unique sparsest solution, in both noiseless and noisy settings. This sparsity-seeking property is of interest in many practical areas such as image and signal processing, communication and information theory, etc. However, most of these sufficient conditions are either too pessimistic although easily computable (e.g. bounds with mutual coherence), or sharp but difficult to check in practice.

In this paper, we provide a *necessary* and *sufficient* condition for x to be identifiable for a large set of matrices A; that is to be the unique sparsest solution to the ℓ_1-norm minimization problem. Furthermore, we prove that this sparsest solution is stable under a reasonable perturbation of the observations y. We also propose an efficient semi-greedy algorithm to check our condition for any vector x. We present numerical experiments showing that our condition is able to predict almost perfectly all identifiable solutions x, whereas other previously proposed criteria are too pessimistic and fail to identify properly some identifiable vectors x. Beside the theoretical proof, this provides empirical evidence to support the sharpness of our condition.

Keywords: Sparse representations, underdetermined linear systems, ℓ_1-minimization, identifiable vectors.

1 Introduction

1.1 Sparse Recovery

Let A be a matrix whose columns vectors $(a_i)_{i\leqslant p}$ are p vectors of $\mathbb{R}^n$, where $n \ll p$ (dimension of observations n is much smaller than that of data p). Let $x_0 \in \mathbb{R}^p$ and $y = Ax_0$. x_0 can be seen as a data vector and y as observations of these data.

One wants to recover data x_0 from the observations y. However, because the underlying linear system is underdetermined, recovery of the overcomplete representation vector x_0 from y faces an apparent obstacle, based on elementary linear algebra. Nevertheless, although the problem of recovering of x_0 is admittedly ill-posed in general, introducing the hypothesis that x_0 has a simple structure can radically change the situation. In this case, one can hope to properly recover x_0 from y under appropriate conditions.

The sparsity assumption is to suppose that x_0 has few non-zero components. Such a hypothesis on the structure of x_0 can often be rephrased as : the expansion of x_0 in a familly of vectors is essentially supported on a few of them.

As a measure of sparsity of a vector x, we may take the ℓ_0 (quasi)-norm $\|x\|_0$, which is the number of non-zero components of x. Hence, if x_0 has few non-zero components, one can hope

that x_0 is the unique minimizer of

$$\min_x \|x\|_0 \text{ under the constraint } Ax = y \qquad \mathrm{P}_0(y)$$

with $y = Ax_0$. We will refer to this problem by $\mathrm{P}_0(y)$ to specify the second member.
This is an NP-hard problem since there is no way to solve it except testing all possible k-collections of columns of A with $k = 1, \ldots, p$, and looking for the smallest k-collection that synthesize y. That is why [12] proposed to substitute this highly non-convex problem with the following convex ℓ_1 minimization problem

$$\min_x \|x\|_1 \text{ under the constraint } Ax = y \qquad \mathrm{P}_1(y)$$

with $y = Ax_0$. We will refer to this problem by $\mathrm{P}_1(y)$ to specify the second member.

It is well known that under appropriate conditions, both problems $\mathrm{P}_0(y)$ and $\mathrm{P}_1(y)$ share the same solutions; see [10, 12, 16, 23, 20] to cite only a few. Furthermore, several conditions on x_0 have been proposed in the literature to guaranty the uniqueness of the solution x_0 to the problem $\mathrm{P}_1(Ax_0)$. Before proceeding, we will need some terminology that will be used in the remainder of the paper.

Definition 1 *A vector x_0 is said to be identifiable if it is the unique solution to $P_1(Ax_0)$.*

Notations The support of x_0 and its cardinal are defined by

$$\mathrm{supp}\,(x_0) = I = \{i | \, x_0(i) \neq 0\} \subset \{1, \ldots, p\} \qquad \text{and} \qquad |I| = |\mathrm{supp}\,(x_0)\,| = \|x_0\|_0 .$$

The vector $\bar{x}$ is obtained by selecting from x components with indices in its support. The matrix $A_I = (a_i)_{i \in I}$ is obtained by selecting the columns of A indexed by I. This matrix is called the active matrix associated to the set I. The active matrix A_I associated to a vector x_0 is the active matrix associated to the support I of x_0. The columns $(a_i)_{i \in I}$ are called active columns or vectors and $(a_j)_{j \notin I}$ are called inactive columns or vectors. The pseudo-inverse A_I^+ of A_I is defined as

$$A_I^+ = (A_I^t A_I)^{-1} A_I^t.$$

A vector x_0 is said to be included in a vector x_1 if $\mathrm{supp}\,(x_0) \subset \mathrm{supp}\,(x_1)$ and $\mathrm{sign}\,(x_0) = \mathrm{sign}\,(x_1)$ on their common support. The vector x_1 is an extension of the vector x_0 if the latter is included in the former.

The vector $\delta_k \in \mathbb{R}^p$ is such that $\|\delta_k\|_0 = 1$ and $\delta_k(i) = 1$ and 0 otherwise.

1.2 State of affairs

1.2.1 The coherence

The most popular sufficient condition that guaranties the identifiablility of x_0 relates its support $\mathrm{supp}\,(x_0)$ to the coherence of A :

$$C(A) = \max_{i \neq j} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \, \|a_j\|_2}. \tag{1}$$

If

$$|I| = \|x_0\|_0 < \frac{1}{2}(1 + \frac{1}{C}) \tag{2}$$

then x_0 is identifiable, moreover x_0 is also the solution of (1.1). This bound has appeared in many papers, e.g. [12, 10, 16, 23, 20].

This bound on the cardinal of $\mathrm{supp}\,(x_0)$ is optimal if one does not have any additionnal information on A. Donoho and Elad [10] proposed to improve this bound using the Spark of A. $\mathrm{Spark}\,(A)$ is defined as the minimal number of linearly dependent vectors a_i. Estimating $\mathrm{Spark}\,(A)$ is computationly prohibitive. Moreover, it is unstable under a small perturbation of A. That is, $\mathrm{Spark}\,(A)$ may change dramatically with A.

Although it appears as a simple and computationally tractable a priori test of identifiability, the coherence-based bound $\frac{1}{2}(1 + \frac{1}{C})$ is pessimistic and can be improved in many situations.

1.2.2 Compressed Sensing

In a series of papers Candès, Tao and Romberg [5, 2, 7, 6, 3, 4] studied different optimization problems including $\mathrm{P}_1(y)$ in [7]. They especially investigated [4] a problem where data y may be corrupted by gaussian noise w whose ℓ_2 norm is bounded by ε, i.e. $y = Ax_0 + w$. The authors suggested to solve the following minimization problem

$$\min_x \|x\|_1 \qquad \text{under the constraint} \qquad \|y - Ax\|_2 \leqslant \varepsilon \tag{3}$$

where ε is the size of the error term.

The authors established conditions on vectors x_0 ensuring identifiability, equivalence between solutions of $\mathrm{P}_0(Ax_0)$ and $\mathrm{P}_1(Ax_0)$, as well as stability to noise. These conditions are uniform on the cardinal of the support. To state their conditions, those authors have introduced the so-called Restricted Isometry Hypothesis (RIH) on A. RIH assumes that any subset I, $|I| \leqslant S$ of vectors $(a_i)_{i\in I}$ defining the matrix A_I, is a Riesz basis with the associated constants uniformely controled by a function δ_S. For all set of indices I such that $|I| \leqslant S$,

$$\forall x \in \mathbb{R}^{|I|}, \qquad (1 - \delta_S)\|x\|_2^2 \leqslant \|A_I x\|_2^2 \leqslant (1 + \delta_S)\|x\|_2^2 \tag{4}$$

The RIH requirement states that for all sets I whose cardinal is smaller than S, the mapping A_I approximately acts like an isometry on $|I|$-sparse vectors.

One of the RIH-based indentifiability conditions requires that

$$\delta_S + \delta_{2S} + \delta_{3S} < 1 \tag{5}$$

see [4], the reader may find others conditions in [8]. These differents RIH lead to differents results of recovery and robustness to noise.

Compressed Sensing (CS) theory shows that all vectors such that $\|x\|_0 \leqslant \frac{Cn}{\log(p/n)}$, are identifiable, with a probability close to 1 when A satifies a RIH. The CS recovery results extend also to vectors x_0 which are nearly sparse [6]; vectors whose ℓ_p norm is concentrated on a sparse set. This point is important in mathematical image or signal processing applications, where x_0 is not exactly sparse but compressible in some transform domain e.g. wavelets [21], curvelets [1]. However, these results do not apply to any matrix. Furthermore, for a deterministic A, there is no simple way to check the RIH and compute the constants δ_S, except testing all possible $|I|$-collections of columns of A which is a combinatorial process. Moreover, it is hard to build matrices satifying RIH for large S. Consequently, if a matrix A is given and not built to satisfy RIH, we cannot staightforwardly use CS bounds to ensure that a vector x_0 is identifiable. To date, the only deterministic contruction of matrices obeing RIH for large values of S has been proposed in 2007 by R. De Vore [24].

1.2.3 Conditions on support and sign for any matrix A

Although the previous bound (2) is optimal for certain dictionaries (matrices), there are many vectors violating the coherence-based sparsity bound that are still identifiable. Gribonval and Nielsen [20] proposed an identifiability criterion which does not depend on the cardinal of $\mathrm{supp}\,(x_0)$, but rather on the support itself. Their criterion can be verified by sparse vectors in the sense of the bound (2), as well as by other vectors that are less sparse. Unfortunately there is no simple way to compute their criterion. As previously pointed out for the Spark, their criterion depends directly on the null space of A and is unstable under small variation of A.

Tropp [23] and Fuchs [18, 19] proposed criteria that apply to any matrix and depends on the structure of the vector x, not only on $|I|$. Both authors study solutions of the relaxed problem

$$\min_x \frac{1}{2}\|y - Ax\|_2^2 + \gamma\|x\|_1 \tag{6}$$

which is equivalent to the previous minimisation (3), i.e. for all $\varepsilon > 0$ there exists a bijection $\gamma(\varepsilon)$ such that problems (3) and (6) share the same solutions. The limit of the solution of (6) when it is unique, tends to the solution of $\mathrm{P}_1(y)$ when γ tends to 0. Tropp defined the Exact Recovery Coefficient (ERC) as follows,

$$\mathrm{ERC}(I) = 1 - \sup_{j\notin I} \left\|A_I^+ a_j\right\|_1$$

The ERC is then defined only when A_I has full rank. Tropp [23] proved that if $\mathrm{ERC}(I) > 0$, any vector x_0 supported in I is identifiable. He also showed that $\mathrm{ERC}(I) > 0$ guaranties stability when obsvervations are corrupted by an additive noise with bounded variance.

From a vector x_0 Fuchs [18] introduces the following vector

$$d_0 = A_I^{+t}\mathrm{sign}\,(\bar{x}_0) \in \mathbb{R}^n \tag{7}$$

which plays a major role to state recovery conditions for x_0. This vectors is spanned by the active vectors $(a_i)_{i\in I}$. Section 2 presents geometrical interpretations of d_0 when n equals 2 and 3. Fuchs proposed sharper sufficiency results by defining a criterion depending on the support and on the sign of x_0 through d_0.

Definition 2 *$\mathcal{F}$ is the set of vectors x_0 such that the active matrix A_I associated to x_0 has full rank and*

$$F(x_0) = \max_{j\notin I} |\langle a_j, A_I^{+t}\mathrm{sign}\,(\bar{x}_0)\rangle| = \max_{j\notin I} |\langle a_j, d_0\rangle| < 1.$$

Fuchs [18] proved that for all vectors $x_0 \in \mathcal{F}$, the minimizer $x(\gamma)$ of (6) is unique for γ small enough, tends to x_0 when γ tends to 0, and then that x_0 is a minimizer of $\mathrm{P}_1(Ax_0)$. Note that if x_1 and x_2 have the same support and the same sign, $x_1 \in \mathcal{F}$ implies $x_2 \in \mathcal{F}$. Hence $\mathcal{F}$ is the union of cones of various dimensions. The condition $F(x_0) < 1$ asserts that the correlation between d_0 and all inactive vectors $(a_j)_{j\notin I}$ is stricly smaller than 1.

Donoho in [9] proposed a necessary and sufficient condition to ensure that a vector x is a minimizer of $\mathrm{P}_1(Ax)$. The author considered the image of the unit ℓ_1-ball by A, which is the polytope whose vertices are $(\pm a_j)_{j\leqslant p}$, and associated to each vector x the corresponding facet $\mathcal{H}(x)$ of the ℓ_1 ball. Hence, this facet $\mathcal{H}(x)$ depends only on the sign and support of x. Donoho proved that x is a minimizer of $\mathrm{P}_1(Ax)$ if and only if the image of the facet $\mathcal{H}(x)$ belongs to the convex hull of the image of the unit ℓ_1-ball by A. This geometrical and topological condition was subsequently used by Donoho and Tanner [14, 13] to estimate the number of vectors x minimizing $\mathrm{P}_1(Ax)$ for a given sparsity. The authors propose sharp results for random projectors and show that the bounds derived from *ERC* or Compressed Sensing are often pessimistic.

The new condition proposed in this paper is strongly linked with both conditions proposed by Fuchs [18] on the one hand, and by Donoho [9] on the other hand.

1.3 Contributions

If no hypothesis are made on A, it may happen that for some y, $\mathrm{P}_1(y)$ has several solutions, for example if two vectors a_i and a_j coincide. The question of the identifiability of a vector x_0 is then ill-posed. The following condition coined (UC), which stands for Unicity Condition, guarantees the unicity of the minimizer of $\mathrm{P}_1(y)$ for any $y \in \mathrm{Im}\,(A)$.

Definition 3 *A satisfies condition* (UC) *if, for all subsets $I \subset \{1,\dots,p\}$, such that $(a_i)_{i\in I}$ are linearly independent, for all indices $j \notin I$ and all vectors $S \in \{-1,1\}^{|I|}$,*

$$|\langle a_j, A_I^{+t}S\rangle| \neq 1 \tag{UC}$$

ie for all x_0 such that A_I has full rank and for all $j \notin I$

$$|\langle a_j, d_0\rangle| \neq 1 \tag{8}$$

where d_0 is the vector defined in section 1.2.3 by (7)*. In words, the correlation between any sign vector and the projection of a_j on the subspace spanned by $(a_i)_{i\in I}$ is never exactely 1 or -1.*

This condition implies that there is no vector x_0 such that $F(x_0) = 1$ and no vectors such that $\mathrm{ERC}(I(x_0)) = 0$.

The main contribution of this paper is Theorem 1 which provides a necessary and sufficient condition for x_0 to be identifiable for the set of matrices A satisfying condition (UC).

Theorem 1 *Suppose that A satisfies condition* (UC)*. Then x_0 is identifiable if and only if $x_0 \in \mathcal{K}$, where $\mathcal{K}$ is the closure of $\mathcal{F}$.*

More precisely, the condition $x_0 \in \mathcal{K}$ is always a sufficient condition for x_0, to be identifibale. If A satisfies the (UC) condition, it becomes also a necessary condition.

Theorem 1 allows then to define the mapping φ

$$\varphi \quad \left\{ \begin{array}{ccc} \mathrm{Im}\,(A) & \longrightarrow & \mathcal{K} \\ y & \longmapsto & x \quad \text{solution of problem } \mathrm{P}_1(y) \end{array} \right. \tag{9}$$

The mapping φ is a non-linear inverse of the linear map A. The following theorem controls the continuity of φ.

Theorem 2 *If A satisfies condition* (UC)*, then φ is uniformly Lipschitz, hence continuous.*

The set $\mathcal{K}$ is composed of any vector that can be extented to yield a vector belonging to $\mathcal{F}$. This means that for any vector $x_0 \in \mathcal{K}$ there is a vector x_1 whose support is disjoint from the one of x_0 such that $x_0 + x_1 = x_2 \in \mathcal{F}$.

Condition $x_0 \in \mathcal{K}$ cannot be easily verified. To circumvent this difficulty section 4 proposes a semi-greedy algorithm termed *SupportExtension*, which exploits the above characterization of $\mathcal{K}$ to recognize those vectors which are in $\mathcal{K}$. More precisely, the success of this algorithm guarantees that $x_0 \in \mathcal{K}$, but its failure does not ensure that $x_0 \notin \mathcal{K}$. Actually, simulations did not provide any identifiable x_0 such that the algorithm fails, but they may exist.

The study of the relaxed formulation (6) associated to problem $\mathrm{P}_1(y)$ is at the heart of the proof of Theorem 1 and 2. These two optimization problems are closely linked and Theorem 1 leads to the following result.

Theorem 3 *If A satisfies condition* (UC)*, for any $y \in \mathbb{R}^n$ and $\gamma > 0$, the minimizer $x(\gamma)$ of*

$$\frac{1}{2} \|y - Ax\|_2^2 + \gamma \|x\|_1$$

is unique and belongs to $\mathcal{K}$.

Actually, Theorem 1 leads to a more general result than Theorem 3. Indeed, with a similar proof, it can be shown that if A satifies (UC), for any closed set $\mathcal{D}$, the set of solutions of

$$\min_x \|x\|_1 \text{ under the constraint } Ax \in \mathcal{D} \qquad \mathrm{P}_1(\mathcal{D})$$

is included in $\mathcal{K}$. Moreover, if $\mathcal{D}$ is striclty convex, then the solution of $\mathrm{P}_1(\mathcal{D})$ is unique and belongs to $\mathcal{K}$. It follows that, if A satifies (UC), the solutions of the following minimization problem, see [6]

$$\min_x \|x\|_1 \text{ under the constraint } \left\|A^t(y - Ax)\right\|_\infty \leqslant \gamma$$

belong to $\mathcal{K}$.

1.4 Relation to prior work

One of the first approaches dealing with this identifiability problem proposes a bound on $\|x_0\|_0$ based on the coherence of matrix A, see section 1.2.1. This bound (2) is often pessimistic. Two approaches to improve over this result can be distinguished. The first one adds hypotheses on the matrix A. Hence, if A obeys some restrictive conditions, bound (2) may be improved. It is the point of view of compressed sensing, see section 1.2.2. Such an approach, provides bounds on $|\mathrm{supp}\,(x_0)|$ to guarantee identifiability, and also ensures the equivalence between $\mathrm{P}_0(y)$ and $\mathrm{P}_1(y)$. It also has the drawback to give pessimistic bounds in many situations, see [7, Theorem 1.6]. Compressed Sensing provides good asymptotic bounds which can be worse than (2).

The second approach abandons the idea of a uniform bound on $|\mathrm{supp}\,(x_0)|$ and uses $\mathrm{supp}\,(x_0)$ itself and sometimes even $\mathrm{sign}\,(x_0)$. These approaches followed by Gribonval and Nielsen [20], Fuchs [18] and Tropp [23] may explain why many vectors that are not so sparse are identifiable, but do not ensure the equivalence between problems $\mathrm{P}_0(y)$ and $\mathrm{P}_1(y)$. This paper improves over the results of Tropp and Fuchs by providing a necessary and sufficient condition for identifiability for a large set of matrices A.

This work may be viewed a complementary approach to Compressed Sensing : whereas Compressed Sensing needs a strong hyopthesis on A and gives strong results of recovery, stability and equivalence between $\mathrm{P}_0(y)$ and $\mathrm{P}_1(y)$, this work proposes near minimal hypotheses on A and gives tight conditions for the ℓ_1 recovery, with minimal stability results and no equivalence between $\mathrm{P}_0(y)$ and $\mathrm{P}_1(y)$.

Following Fuchs [18, 19] and Tropp [23], this papers focuses on the relaxed and convex problem

$$\min_x \frac{1}{2} \|y - Ax\|_2^2 + \gamma \|x\|_1 \qquad \mathrm{P}_1(y,\gamma)$$

More precisely, it investigates the properties of the solutions of $\mathrm{P}_1(y,\gamma)$ for small γ. This problem is refered to as $\mathrm{P}_1(y,\gamma)$ to specify the second member y and the parameter γ. A solution of $\mathrm{P}_1(y,\gamma)$ will be denoted $x(\gamma)$.This relaxed formulation $\mathrm{P}_1(y,\gamma)$ is particularly well adapted to deal with observations corrupted by an additional noise $y = Ax_0 + w$, but can also give information about the solution of $\mathrm{P}_1(Ax_0)$.

As previously mentionned, this formulation is equivalent, under appropriate correspondence of parameters γ and ε, to (3) used by Candès et al. to develop some Compressed Sensing results [4] and also by Donoho, Elad Temlyakov [11] and others.

Even if the final condition $x \in \mathcal{K}$ is derived from algebraic relationships satisfied by the solutions of $\mathrm{P}_1(Ax,\gamma)$, it is clearly related to the topological properties of the set $\mathcal{F}$. It turns out that it is also naturally linked to the topological condition proposed by Donoho [9]. Hence $\forall x$, $\langle u, A_I^{+t}\mathrm{sign}\,(\bar{x})\rangle = 1$ is the equation of a hyperplane P containing all signed active vectors associated to x, i.e. if $i \in \mathrm{supp}\,(x)$, $\mathrm{sign}\,(x(i))\,a_i \in P$. Condition $F(x) < 1$ ensures that all inactive vectors $(a_j)_{j \notin I(x)}$ belong to the same half-space "below" the hyperplane P. P is then one of the hyperplanes defining the convex hull of the polytope formed by the vectors $(\pm a_j)_{j \leqslant p}$. Moreover if x can be extented into a vector $x_1 \in \mathcal{F}$, one can define a hypperplane P containing all signed and active vectors associated to x_1 such that all $(\pm a_j)_{j \notin I(x_1)}$ belong to the same half-space defined by P, i.e. P is one of the hyperplanes defining the convex hull of $(\pm a_j)_{j \leqslant p}$.

Hence if A satisfies condition (UC), both our characterization of identifiability and the one proposed by Donoho [9] are equivalent. In fact, this paper sheds light on the relation between the algebraic and the analytical point of view on this characterization. It also provides a condition ensuring unicity and a fast algorithm to check the identifiability of a vector x.

Section 4 devoted to the numerical experiments shows that there are many identifiable x_0 that do not satisfy any of the previous conditions reviewed above. The different bounds derived from the mutual coherence (2), the ERC or the CS theory are actually too pessimistic. This pessimism is necessary to get these bounds to be uniform over the support or, even worse, over the cardinal of the support. We will see that for a given sparsity or a given support, most of vectors may be identifiable and only a small fraction of them may not.

Hence, this new approach sheds light on those identifiable vectors x_0 that are not very sparse. In particular, it gives clues to understand why CS can be used with a good probability of success beyond theoretical bounds.

All results presented here hold true provided that A satisfies condition (UC). Indeed, as previously said, the unicity of the solution is important to define the identifiability, and condition (UC) ensures this uniqueness. It turns out that this condition is not really restrictive. In particular if the vectors a_i are independent and randomly generated according to a probability law with a density, the probability that A satisfies (UC) is exactly 1.

2 Geometric insight

The analytical details look more complicated than the simple underlying geometry. Hence, before giving a proof of Theorem 1, some insight may be gleaned by considering the geometry underlying the set $\mathcal{K}$ and condition (UC) for $n = 2$ and $n = 3$. In this section A is supposed to satisfy condition (UC), $\mathrm{Im}\,(A) = \mathbb{R}^n$ and the vectors $(a_i)_{j \leqslant p}$ belong to the unit sphere $\mathcal{S}$ of $\mathbb{R}^2$ or $\mathbb{R}^3$. We denote $\mathcal{B} = (b_j)_{j \leqslant 2p} = (\pm a_i)_{i \leqslant p}$ and $b_j = \sigma_j a_{\psi(j)}$, where $\sigma_j \in \{-1, 1\}$ and $\psi(j) \leqslant p$.

By definition, $\mathcal{F}$ is a finite union of half-cones of various dimensions, each half-cone being defined by positions and signs of non-zero components. Consequently $\mathcal{K}$ is also a union of closed

half-cones. This section exemplifies these half-cones when $n = 2$ and $n = 3$ and gives a geometrical interpretation of the function φ as a bijection between two sets of cones.

For a given set $J \subset \{1, \cdots, 2p\}$, a half-cone C_J in $\mathbb{R}^n$ by

$$C_J = \left\{ \sum_{j \in J} \lambda_j b_j \quad \text{with} \quad \lambda_j > 0, \forall j \in J \right\} \tag{10}$$

and a half-cone K_J in $\mathbb{R}^p$

$$K_J = \left\{ \sum_{j \in J} \sigma_j \lambda_j \delta_{\psi(j)} \quad \text{with} \quad \lambda_j > 0, \forall j \in J \right\}. \tag{11}$$

Hence C_J is the image of K_J by A, $\forall J \subset \{1, \cdots, 2p\}$.

The following theorem explains how application φ induces a tiling of sets $\mathrm{Im}\,(A)$ and $\mathcal{K}$.

Theorem 4 *For $n = 2$ and $n = 3$, there is a set $\mathcal{P}$ such that*

$$\mathbb{R}^n = \cup_{J \in \mathcal{P}} \bar{C_J} \quad \textit{and} \quad \mathcal{K} = \cup_{J \in \mathcal{P}} \bar{K_J} \tag{12}$$

where $\bar{C_J}$ and $\bar{K_J}$ are the closure sets of C_J and K_J. Moreover, for all $J \in \mathcal{P}$, φ acts as a linear bijection from C_J to K_J.

The purpose of the two following subsections is to describe the set $\mathcal{P}$ for $n = 2$ and $n = 3$. More precisely, it is shown that $\forall J \in \mathcal{P}$ $K_J \subset \mathcal{F}$, from which it is deduced that K_J is the image of C_J by φ using Theorem 1.

2.1 Example in two dimensions $n = 2$

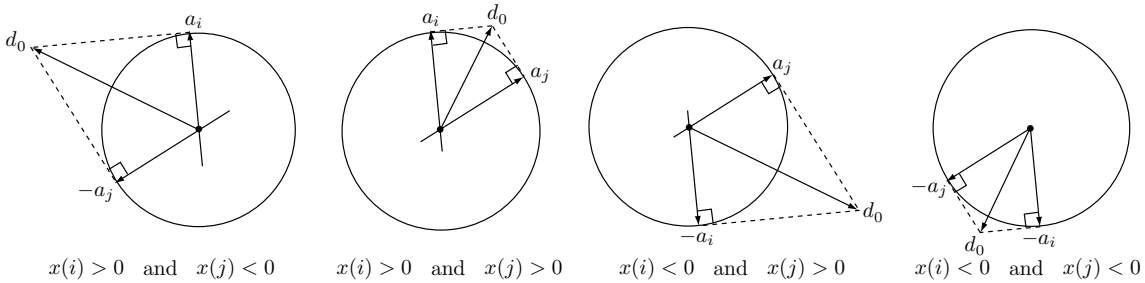
First, we note that condition (UC) implies that $(b_j)_{j \leqslant 2p}$ are all disctinct. Let's define the set $\mathcal{P} = \{1, \cdots, 2p\}^2$ by $\mathcal{P} = \{J = (j, l) \text{ such that } \mathcal{B} \cap C_J = \emptyset\}$.
One can notice that $(j, l) \in \mathcal{P}$ if and only if b_j and b_l correspond to two conscutive points of the set $\mathcal{B}$ on the unit circle $\mathcal{S}$. Hence one gets

$$\mathbb{R}^2 = \cup_{J \in \mathcal{P}} \bar{C_J}. \tag{13}$$

We now prove that for any $I \in \mathcal{P}$, $K_I \in \mathcal{F}$. Suppose $J = \{j, l\} \in \mathcal{P}$ and $x_0 \in K_J$. Let's denote $I = \{\psi(j), \psi(l)\}$ the support of x_0. Since $(b_j)_{j \leqslant 2p}$ are disctinct, $\mathrm{rank}(A_I) = 2$ and is then maximal. The vector d_0 is defined by $d_0 = A_I^{+t} \mathrm{sign}\,(\sigma_j, \sigma_l)$, it is spanned by b_j and b_l and satisfies

$$\begin{aligned} \langle d_0, a_{\psi(j)} \rangle &= \mathrm{sign}\,(\sigma_j) \text{ i.e. } \langle d_0, b_j \rangle = 1 \\ \langle d_0, a_{\psi(l)} \rangle &= \mathrm{sign}\,(\sigma_l) \text{ i.e. } \langle d_0, b_l \rangle = 1 \end{aligned}$$

This vector d_0 belongs to the bisector of b_j and b_l.



$x(i) > 0$ and $x(j) < 0$ $\quad$ $x(i) > 0$ and $x(j) > 0$ $\quad$ $x(i) < 0$ and $x(j) > 0$ $\quad$ $x(i) < 0$ and $x(j) < 0$

Figure 1: *Four different configurations of coefficients $(x_0(i), x_0(j))$ and the corresponding d_0.*

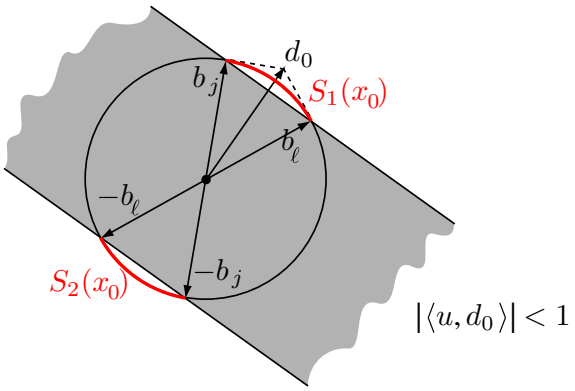


Figure 2: *Example of vector d_0 and caps $S_1(x_0)$ and $S_2(x_0)$ for a vector $x_0 \in K_{j,l}$.*

The set of vectors u such that $\langle u, d_0\rangle = 1$ is the chord (b_j, b_l). Hence, from the definition of d_0, one has

$$S_1(x_0) = \mathcal{S} \cap \{u \text{ such that } \langle u, d_0\rangle > 1\} = \mathcal{S} \cap C_J \text{ and } S_2(x_0) = \mathcal{S} \cap \{u \text{ such that } \langle u, d_0\rangle < -1\} = -S_1(x_0)$$

S_1 and S_2 correspond to the arcs shown with the red lines in Figure 2.

Thus, a vector x_0 belongs to $\mathcal{F}$ if and only if

$$F(x_0) = \max_{k \notin I} |\langle a_k, d_0\rangle| < 1$$

Hence

$$\begin{aligned} x_0 \in \mathcal{F} \quad &\iff \quad (\mathcal{B} \cap S_1(x_0) = \emptyset \quad \text{and} \quad \mathcal{B} \cap S_2(x_0) = \emptyset) \\ &\iff \quad \mathcal{B} \cap C_J = \emptyset, \end{aligned}$$

because the set $\mathcal{B}$ is anti-symmetric, i.e. $\mathcal{B} = -\mathcal{B}$. Hence, $x_0 \in \mathcal{F}$ if and only if $J \in \mathcal{P}$, and then

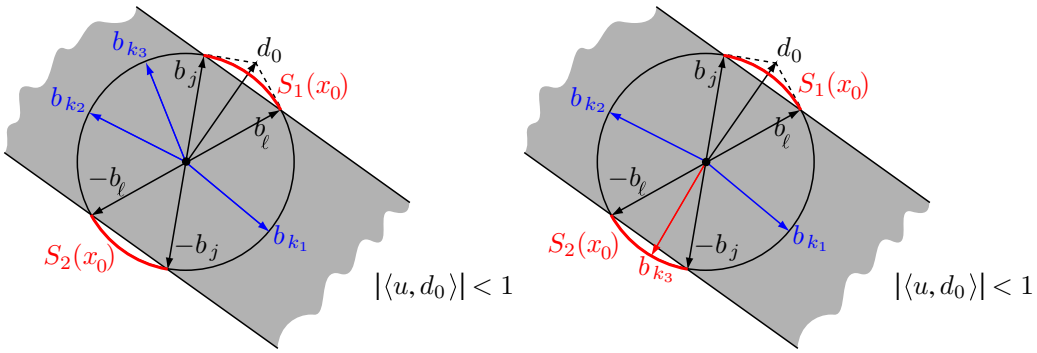


Figure 3: *On the left, $(S_1(x_0) \cup S_2(x_0)) \cap \mathcal{B} = \emptyset$ and then $x_0 \in \mathcal{F}$, on the right $S_2(x_0) \cap \mathcal{B} \neq \emptyset$ and then $x_0 \notin \mathcal{F}$.*

$K_J \subset \mathcal{F}$ if and only if $J \in \mathcal{P}$. Thus $\cup_{J\in\mathcal{P}} K_J \subset \mathcal{F}$ implying that $\cup_{J\in\mathcal{P}} \bar{K}_J \subset \mathcal{K}$.

For all $y \in \mathbb{R}^2$, from (13) there exists some $J \in \mathcal{P}$ such that $y \in \bar{C}_J$ and then there is $x \in \bar{K}_J$ such that $y = Ax$. Since $x \in \mathcal{K}$, from Theorem 1, it is concluded that $\varphi(y) = x$. Hence, $\forall x \in \mathbb{R}^p$, $\varphi(Ax) \in \cup_{J\in\mathcal{P}} K_J$ and then $\mathcal{K} \subset \cup_{J\in\mathcal{P}} K_J$, that is $\mathcal{K} = \cup_{J\in\mathcal{P}} K_J$ and $\mathcal{F} = \mathcal{K}$. Moreover $\forall x \in K_J$, $J = \{j, l\}$, $Ax = A_I \bar{x}$ where $I = \{\psi(j), \psi(l)\}$ we have $\bar{x} = A_I^+ A_I \bar{x}$. Hence $\forall y \in C_J$, $\varphi(y) = A_I^+ y$ and then φ is linear from C_J to K_J. It follows that $\forall (y_1, y_2) \in (\mathbb{R}^2)^2$,

$$\|\varphi(y_1) - \varphi(y_2)\|_2 \leqslant \max_{J\in\mathcal{P}} \left\|A_I^+\right\|_{2,2} \|y_1 - y_2\|_2 \quad \text{with} \quad \left\|A_I^+\right\|_{2,2} = \max_{x\in\mathcal{S}} \left\|A_I^+ x\right\|_2 \tag{14}$$

Thus, $\max_{J\in\mathcal{P}} \left\|A_I^+\right\|_{2,2}$ is the best Lipschitz constant associated to the function φ.

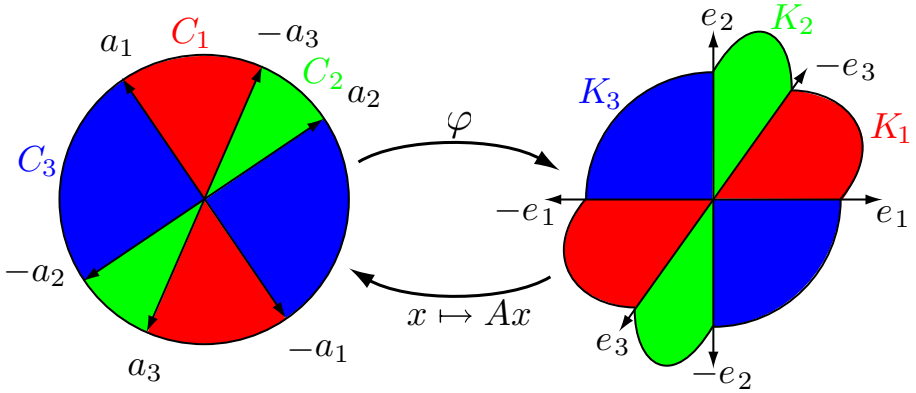


Figure 4: *When $n = 2$ and $p = 3$, the map φ acts as a bijection cone by cone and sends the unit disk onto a manifold of $\mathbb{R}^3$.*

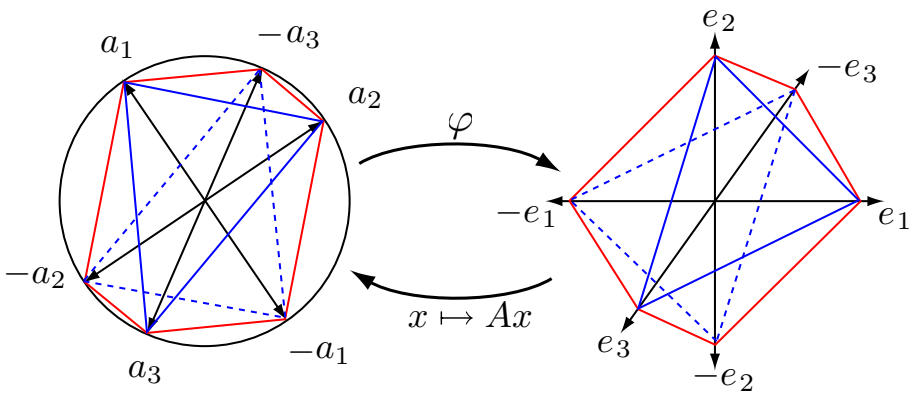


Figure 5: *Right: The cones $K_J \subset \mathcal{F}$ correspond to the edges, here in red, of the unit ℓ_1-ball. Left: The images by A of these edges are the (red) edges of the convex hull of the polytope $(\pm a_j)_{j \leqslant p}$.*

2.2 Example in three dimensions $n = 3$

We now investigate the case $n = 3$ where $(a_i)_{i \leqslant p}$ and $\|a_j\|_2 = 1, \forall j \leqslant p$. To give a geometric intuition of what happens in dimension 3, some properties of spherical triangulation are recalled in the following. To begin, definitions of facets and spherical caps are given.

Definition 4 *Let $(x_l)_{l \leqslant n} \in \mathcal{S}$ a set of vectors on the unit sphere and $J \subset \{1, \cdots, n\}$, such that points $(x_j)_{j \in J}$ are coplanar and such that $\dim(\mathrm{Span}(x_j)_{j \in J}) = 3$. The set $(x_j)_{j \in J}$ is called a facet of the set $(x_l)_{l \leqslant n}$. There is a vector x such that for all $j \in J$, $\langle x_j, x \rangle = 1$. The spherical cap S_J associated to the facet $(x_j)_{j \in J}$ is defined by*

$$S_J = \{u \text{ such that } \langle u, x \rangle > 1\} \cap \mathcal{S} \tag{15}$$

Then one defines a general triangulation on the sphere $\mathcal{S}$.

Definition 5 *A triangulation $\mathcal{T}$ of $(x_i)_{i \leqslant n} \in \mathbb{R}^3$ is a set of triplets (i, j, k) with an adjacence relationship. If $(i, j, k) \in \mathcal{T}$, the segments (i, j), (j, k) and (k, i) belong to two triangles.*

A spherical Delaunay triangulation is defined by

Definition 6 *A spherical Delaunay triangulation of a set $(x_i)_{i \leqslant n} \in \mathcal{S}$ is a triangulation $\mathcal{T}$ such that for any $J = (i, j, k) \in \mathcal{T}$, no vectors x_l, for $l \notin J$ belongs to the cap S_J, $S_J \cap (x_l)_{l \leqslant n} = \emptyset$.*

This definition is an extension of the definition of a Delaunay triangulation in the plane, where interiors of circumcircles of triangles of the triangulation for points $(x_i)_{i\leqslant n}$ do not intersect the set $(x_i)_{i\leqslant n}$.

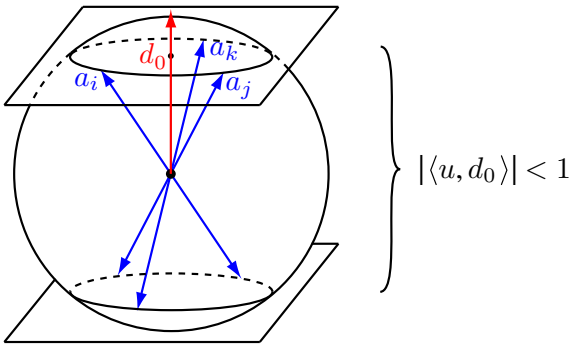


Figure 6: *Example of 3D spherical caps associated to a vector x such that $\|x\|_0 = 3$ and $x(i) > 0$, $x(j) > 0$ and $x(k) > 0$.*

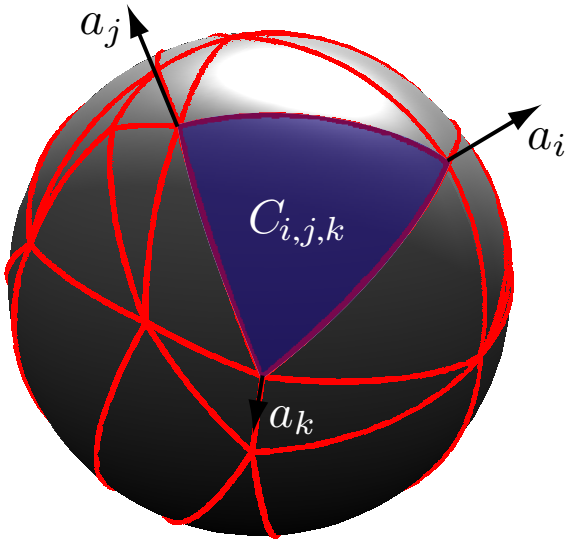


Figure 7: *Example of a cone $C_{i,j,k}$ belonging to the set $\mathcal{T}$.*

The following lemma is needed to ensure that, for a spherical Delaunay triangulation, the only points b_j on the border of a spherical cap S_J are b_j, for any $j \in J$. This lemma actually guaranties, under the hypothesis (UC), the unicity of the spherical Delaunay triangulation. The proof of the latter assertion is omitted here.

Lemma 1 *Assume that A satisfies* (UC) *and $\mathcal{T}$ is a spherical Delaunay triangulation of the set $\mathcal{B}$. For all $J = (i, j, k) \in \mathcal{T}$,*

$$\bar{S_J} \cap \mathcal{B} = \{b_i, b_j, b_k\} \tag{16}$$

that is if the cap S_J is defined by (15)*, $\forall m \notin J$, $\langle b_m, u\rangle < 1$ and then $\forall l \leqslant 2p$, $\langle b_l, u\rangle \leqslant 1$.*

Proof: Let's define $x_0 = \sigma_i\delta_{\psi(i)} + \sigma_j\delta_{\psi(j)} + \sigma_k\delta_{\psi(k)}$, and $I = (\psi(i), \psi(j), \psi(k))$ its support. A satisfies condition (UC), then $\forall m \notin I$, $|\langle a_m, d_0\rangle| \neq 1$, where $d_0 = A_I(A_I^tA_I)^{-1}\text{sign}\,(\sigma_i, \sigma_j, \sigma_k)$. From the definition of the spherical Delaunay triangulation, $S_J \cap \mathcal{B} = \emptyset$ and then, $\forall l \leqslant 2p$, $\langle d_0, b_l\rangle \leqslant 1$. Equation $\langle u, d_0\rangle = 1$ is that of the plane (b_i, b_j, b_k), since $\langle b_i, d_0\rangle = \langle b_j, d_0\rangle = \langle b_k, d_0\rangle = 1$. We then deduce that there are no other points b_m satisfying $\langle b_m, d_0\rangle \geqslant 1$, which concludes the proof. ■

For any set of points $(x_i)_{i\leqslant n}$ in $\mathbb{R}^3$, the triangulation of the convex hull is a spherical Delaunay triangulation and then there is always such a triangulation. Let $\mathcal{T}$ be the spherical Delaunay triangulation of $\mathcal{B}$. Since $\mathcal{T}$ is a triangulation of $\mathcal{B}$, $\forall y \in \mathbb{R}^3$, there is $J \in \mathcal{T}$ such that $y \in \bar{C_J}$, and then

$$\mathbb{R}^3 = \cup_{J\in\mathcal{T}} \bar{C_J} \tag{17}$$

We now prove that $\forall J \in \mathcal{T}$, $K_J \in \mathcal{F}$. Suppose that $J = (i,j,k) \in \mathcal{T}$ and $x_0 \in K_J$. The set $I = (\psi(i), \psi(j), \psi(k))$ is the support of x_0. One first notices that $\text{rank}(A_I) = 3$ and then is maximal. Equation $\langle u, d_0\rangle = 1$, where $d_0 = A_I(A_I^t A_I)^{-1}\text{sign}\,(\sigma_i, \sigma_j, \sigma_k)$, is then the equation of the plane defined by the points (b_i, b_j, b_k). Hence condition $F(x_0) = \max_{m\notin I} |\langle a_m, d_0\rangle| = \max_{l\notin J} |\langle b_l, d_0\rangle| < 1$ is equivalent to assert that $\bar{S_J} \cap (\mathcal{B}\backslash(b_i, b_j, b_k)) = \emptyset$. Since $x_0 \in K_J$, $J \in \mathcal{T}$ and $\mathcal{T}$ is a spherical Delaunay triangulation, from Lemma 1, it is straightward to seen that $\bar{S_J} \cap (\mathcal{B}\backslash(b_i, b_j, b_k)) = \emptyset$. As a consequence, $F(x_0) < 1$ i.e. $x_0 \in \mathcal{F}$. Hence $\cup_{J\in\mathcal{T}} K_J \subset \mathcal{F}$. Using the same arguments as the previous subsection, it is easy to prove that $\mathcal{K} = \cup_{J\in\mathcal{T}} \bar{K_J}$ and that φ is linear from C_J to K_J.

Here again, inequality (14) holds and $\max_{J\in\mathcal{P}} \left\|A_I^+\right\|_{2,2}$ is the best Lipschitz constant associated to φ. For any dimension n, the set $\mathcal{K}$ is a union of cones $\mathcal{K} = \cup_{J\in\mathcal{P}} \bar{K_J}$, where $K_J \subset \mathcal{F}$.

3 A sufficient and necessary condition of identifiability

In this section we givethe proofs of Theorems 1, 2 and 3. The proof of Theorem 1 is split into two propositions. The first one corresponding to a sufficient condition on x_0 to be identifiable:

Proposition 1 *If $x_0 \in \mathcal{K}$ then x_0 is identifiable, that is x_0 is the unique solution of $P_1(Ax_0)$.*

The second one corresponds to a necessary condition for x_0 to be identifiable:

Proposition 2 *Let A be a matrix satisfying* (UC)*, for any $y \in \text{Im}\,(A)$ there is a unique solution x_0 of $P_1(y)$, moreover $x_0 \in \mathcal{K}$.*

More precisely, one proves that if A satisfies condition (UC), for any $y \in \text{Im}\,(A)$, the solution of $\text{P}_1(y)$ is unique and is in $\mathcal{K}$. After developing the main key ideas giving a flavour of the proof, the proofs of Proposition 1, Proposition 2, Theorem 2 and Theorem 3 are detailed in four subsections. Some intermediate technical lemmas will be needed. For the sake of conciseness, their proofs are deferred to the appendix awaiting inspection by the genuinely interested reader.

3.1 Strategy of proof

As previously mentioned, this paper focuses on the properties of the minimizer of $\text{P}_1(y,\gamma)$ for a small γ. A key ingredient of this proof is to notice that if $x(\gamma)$ is the unique minimizer of $\text{P}_1(y,\gamma)$, then $x(\gamma)$ is the unique minimizer of $\text{P}_1(Ax(\gamma))$ that is $x(\gamma)$ is identifiable.

To prove Proposition 1, it is shown that any $x_0 \in \mathcal{F}$ is the unique solution of $\text{P}_1(y_1, \varepsilon)$ for a suitable y_1 and ε, and then that x_0 is identifiable. The rest of the proof relies on the fact that any vector $x_0 \in \mathcal{K}$ can be extended into a vector $x_1 \in \mathcal{F}$. To prove Proposition 2, it is argued that there is a sequence of $x(\gamma_n)$, solutions of $\text{P}_1(y,\gamma_n)$ belonging to $\mathcal{F}$ and tending to a vector x_0 such that $y = Ax_0$.

The proof of Theorem 2 uses the fact that $x(\gamma)$, solution of $\text{P}_1(y,\gamma)$ varies on a continuous piecewise linear curve when γ varies. As a byproduct, the proof of this theorem establishes the stability of $\text{P}_1(y,\gamma)$ to a small variation of y.

To show Theorem 3, it is first proved that all solutions of $\text{P}_1(y,\gamma)$ have the same image by A, using convexity. The unicity and the fact that this solution belongs to $\mathcal{K}$ is a consequence of Theorem 1.

3.2 If $x_0 \in \mathcal{K}$, then x_0 is identifiable

To prove that $x_0 \in \mathcal{K}$ is a sufficient condition for x_0 to be identifiable we do not require that A satisfies condition (UC). The following lemma establishes that $x_0 \in \mathcal{F}$ is a sufficient condition for x_0 to be identifiable.

Lemma 2 *If $x_0 \in \mathcal{F}$, x_0 is the unique minimizer of $P_1(Ax_0)$.*

Proof: The proof is started by appealing to the following classical optimization lemma, which gives sufficient conditions under which a vector x^* is the unique minimizer of $\mathrm{P}_1(y,\gamma)$, see [17, 18].

Lemma 3 *The three following conditions are sufficient for x^* to be the unique minimizer of $P_1(y,\gamma)$*

1. $A_I^t(y - Ax^*) = \gamma(\mathrm{sign}\,(\bar{x}^*))$,
2. $|\langle a_j, y - Ax^*\rangle| < \gamma$ *for any inactive vector a_j associated to x^*,*
3. A_I *is full rank.*

where I is the support of x^*.

Moreover x^* satisfies the following implicit relationship:

$$\bar{x}^* = A_I^+ y - \gamma(A_I^t A_I)^{-1}\mathrm{sign}\,(\bar{x}^*)\,. \tag{18}$$

Let $x_0 \in \mathcal{F}$, and A_I be the associated active matrix and $\varepsilon > 0$ such that

$$\mathrm{sign}\left(\bar{x}_0 + \varepsilon(A_I^t A_I)^{-1}\mathrm{sign}\,(\bar{x}_0)\right) = \mathrm{sign}\,(\bar{x}_0)\,. \tag{19}$$

If ε is small enough the previous relation (19) always holds.

Let x_1 be the vector satisfying $I(x_1) = I(x_0) = I$ and defined by $\bar{x}_1 = \bar{x}_0 + \varepsilon(A_I^t A_I)^{-1}\mathrm{sign}\,(\bar{x}_0)$ and $y_1 = Ax_1$. By construction $x_1 \in \mathcal{F}$ and $y_1 - Ax_0 = \varepsilon A_I(A_I^t A_I)^{-1}\mathrm{sign}\,(\bar{x}_0)$ and then

$$A_I^t(y_1 - Ax_0) = \varepsilon A_I^t A_I(A_I^t A_I)^{-1}\mathrm{sign}\,(\bar{x}_0) = \varepsilon\mathrm{sign}\,(\bar{x}_0)\,.$$

Moreover, for all inactive vector a_j,

$$|\langle a_j, y_1 - Ax_0\rangle| = \varepsilon|\langle a_j, A_I(A_I^t A_I)^{-1}\mathrm{sign}\,(\bar{x}_0)\rangle| \leqslant \varepsilon F(x_0) < \varepsilon.$$

Then Lemma 3 implies that x_0 is the unique minimizer of $\mathrm{P}_1(y_1,\varepsilon)$.

Then, for any $x_2 \in \mathbb{R}^p$,

$$\frac{1}{2}\left\|y_1 - Ax_2\right\|_2^2 + \varepsilon\left\|x_2\right\|_1 > \frac{1}{2}\left\|y_1 - Ax_0\right\|_2^2 + \varepsilon\left\|x_0\right\|_1. \tag{20}$$

In particular, if $Ax_2 = Ax_0$, the relation (20) implies that $\|x_2\|_1 > \|x_0\|_1$, i.e. x_0 is identifiable which concludes the proof of the lemma. ∎

Let $x_0 \in \mathcal{K}$, and let $x_2 \in \mathbb{R}^p$ such that $Ax_0 = Ax_2$ and $\|x_2\|_1 \leqslant \|x_0\|_1$. Since $x_0 \in \mathcal{K}$, there is a vector x_1 whose support is disjoint from that of x_0, such that $x_0 + x_1 = x_3 \in \mathcal{F}$. Let $x_4 = x_2 + x_1$, by definition $Ax_4 = Ax_3$ and

$$\left\|x_4\right\|_1 \leqslant \left\|x_2\right\|_1 + \left\|x_1\right\|_1 \leqslant \left\|x_0\right\|_1 + \left\|x_1\right\|_1 = \left\|x_3\right\|_1, \tag{21}$$

which implies, from lemma 2 that $x_4 = x_3$ and then $x_2 = x_0$. That is, x_0 is identifiable.

3.3 If x_0 is identifiable, $x_0 \in \mathcal{K}$

In this subsection, A is supposed to satisfy condition (UC). As mentionned in the strategy of the proof, subsection 3.1, we start by showing that under condition (UC), a solution $x(\gamma)$ of $\mathrm{P}_1(y,\gamma)$ is in $\mathcal{F}$, for small γ.

Let $y \in \mathrm{Im}\,(A)$, $\gamma_n > 0$ a sequence of real numbers decaying to zero, and $x(\gamma_n)$ a sequence of solutions of $\mathrm{P}_1(y,\gamma_n)$. Such a sequence does not need to be uniquely defined and an arbitrary solution is chosen for each γ_n. Up to the extraction of a sub-sequence, it is supposed that the sequence $x(\gamma_n)$ converges to some x_0. From the definition of $x(\gamma_n)$, $\|y - Ax(\gamma_n)\|_2^2 + \gamma_n\|x(\gamma_n)\|_1 \leqslant \gamma_n\|z\|_1$, where z is a vector such that $y = Az$ and then $\|y - Ax(\gamma_n)\|_2 \to 0$ when $\gamma_n \to 0$ and then $Ax_0 = y$. Let n_0 such that $\forall n \geqslant n_0$, $I(x_0) \subset I(x(\gamma_n))$. From now, it is assumed that $n \geqslant n_0$. We use the following optimization lemma (see e.g. Fuchs [18]) and condition (UC) to prove that the rank of the active matrix A_I associated to $x(\gamma_n)$ is maximum.

Lemma 4 *A necessary and sufficient condition for $x(\gamma)$ to be a minimizer of $P_1(y,\gamma)$ is that $x(\gamma)$ satisfies the two following conditions*

$$A_I^t(y - A_I\bar{x}(\gamma)) = \gamma \text{sign}\left(\bar{x}(\gamma)\right), \tag{22}$$

$$|\langle a_k, y - A_I\bar{x}(\gamma)\rangle| \leqslant \gamma \quad \textit{for all inactive vectors } (a_k)_{k\notin I}. \tag{23}$$

where $I = \text{supp}\left(x(\gamma)\right)$ and $\bar{x}(\gamma)$ is the vector obtained by keeping the non-zero components of $x(\gamma)$.

Let's suppose A_I does not have a full rank. There exists a set $J \subset I$ and an index $k \in I \setminus J$ such that $|J| = \text{rank}(A_J) = \text{rank}(A_I)$ and $a_k \in \text{Span}(a_j)_{j\in J}$, i.e. $a_k = A_I A_I^+ a_k$. Moreover, (22) implies that

$$A_J^t(y - Ax(\gamma_n)) = \gamma_n \text{sign}\left(\bar{x}_J(\gamma_n)\right)$$

where $\bar{x}_J(\gamma_n)$ is the vector extracted from $x(\gamma_n)$ whose components are indexed by J. From (22), it is also deduced that

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

and then $|\langle a_k, A_J^{+t}\text{sign}\left(\bar{x}_J(\gamma_n)\right)\rangle| = 1$, which is impossible since A satisfies condition (UC). Hence, the rank of A_I is maximum and $A_I^t A_I$ is non-singular.

From (22), it follows that

$$\bar{x}(\gamma_n) = A_I^+ y - \gamma_n (A_I^t A_I)^{-1}\text{sign}\left(\bar{x}(\gamma_n)\right).$$

Then for all $j \notin I$

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

Since $I(x_0) \subset I(x(\gamma_n))$, one has $\bar{x}_0 = A_I^+ A_I \bar{x}_0$ and then $A_I A_I^+ y = A_I A_I^+ A_I \bar{x}_0 = A_I \bar{x}_0 = y$ which gives

$$\langle a_j, y - Ax(\gamma_n)\rangle = -\gamma_n \langle a_j, A_I^{+t}\text{sign}\left(\bar{x}(\gamma_n)\right)\rangle.$$

Using (23),

$$|\langle a_j, A_I^{+t}\text{sign}\left(\bar{x}(\gamma_n)\right)\rangle| \leqslant 1.$$

Since A satisfies (UC),

$$|\langle a_j, A_I^{+t}\text{sign}\left(\bar{x}(\gamma_n)\right)\rangle| \neq 1$$

and then

$$|\langle a_j, A_I^{+t}\text{sign}\left(\bar{x}(\gamma_n)\right)\rangle| < 1$$

It follows from Lemma 3 that $x(\gamma_n)$ is the unique solution of $\text{P}_1(y,\gamma_n)$ and $x(\gamma_n) \in \mathcal{F}$. Hence, x_0 the limit of elements of $\mathcal{F}$, belongs to $\mathcal{K}$.

Using Proposition 1, x_0 is then the unique solution of $\text{P}_1(Ax_0)$ which concludes the proof of Proposition 2.

3.4 Proof of theorem 2

Let y_0 and y_1 be two elements of $\text{Im}\left(A\right)$. If y_0 and y_1 are close enough, the two associated minimizers $x_0 = \varphi(y_0)$ and $x_1 = \varphi(y_1)$ are also close. More precisely, it will be shortly shown that there is a constant C, independent of y_0 and y_1, such that

$$\|x_1 - x_0\|_2 \leqslant C \|y_1 - y_0\|_2 \tag{24}$$

owing to the properties of the minimizer of $\text{P}_1(y,\gamma)$.

Let $x_0(\gamma)$ (resp. $x_1(\gamma)$) denotes the minimizer of $\text{P}_1(y_0,\gamma)$ (resp $\text{P}_1(y_1,\gamma)$). For all $\gamma > 0$,

$$\|x_0 - x_1\|_2 \leqslant \|x_0 - x_0(\gamma)\|_2 + \|x_0(\gamma) - x_1(\gamma)\|_2 + \|x_1 - x_1(\gamma)\|_2$$

The following lemma bounds $\|x_0 - x_0(\gamma)\|_2$ and $\|x_1 - x_1(\gamma)\|_2$.

Lemma 5 *For all $y \in \mathbb{R}^n$, $x(\gamma)$, the minimizer $P_1(y,\gamma)$, is a continuous function of γ and lives on a polygonal path. Moreover $x(\gamma)$ is C_0-Lipschitz where C_0 does not depend on y.*

A proof of this lemma can be found in the appendix. This lemma is at the heart of the homotopy method, see for example [22, 15]

It follows from this lemma that for all $\gamma > 0$

$$\|x_0 - x_1\|_2 \leqslant 2C_0\gamma + \|x_0(\gamma) - x_1(\gamma)\|_2 .$$

To bound $\|x_0(\gamma) - x_1(\gamma)\|_2$, the stability of the minimization problem $\mathrm{P}_1(y,\gamma)$ to a small additive noise is exploited. This is formally summarised in the following lemma.

Lemma 6 *There exists two real positive numbers C_1 and C_2 such that $\forall y_0 \in \mathbb{R}^n$ if $\|y_1 - y_0\|_2 \leqslant \varepsilon \leqslant \varepsilon_0$ for a noise level $\varepsilon_0 > 0$, then,*

$$\|x_1(C_1\varepsilon) - x_0(C_1\varepsilon)\|_2 \leqslant C_2\varepsilon. \tag{25}$$

The proof of this lemma is given in the appendix B.

Hence, armed with Lemma 5 and 6, it follows that

$$\|x_0 - x_1\|_2 \leqslant 2C_0C_1\varepsilon + C_2\varepsilon = (2C_0C_1 + C_2)\|y_1 - y_0\|_2 .$$

which concludes the proof.

Unfortunatly, at this point of our work, one does not have any control on the numbers C_0, C_1 and C_2 and the Lipschitz property is essentialy a theoretical result and cannot stand for a result of robustness to noise. Nevertheless, empirical findings from the numerical experiments clearly demonstrate that, most of the time, there is a real stability to a small noise. Note that since the condition $x_0 \in \mathcal{K}$ is sharp to ensure identifiability of x_0, it seems difficult to prove a strong stability to noise.

3.5 Proof of Theorem 3

In many situations such as signal processing, statistics and model selection [?, ?] for example, the observations y are corrupted by noise, $y = Ax_0 + w$, or x_0 is not exactely sparse. A way to estimate x_0 from y in this non-ideal situation, is to look at $x(\gamma)$, where γ depends on the noise level ε. That is why the solution $x(\gamma)$ of $\mathrm{P}_1(y,\gamma)$ is interesting by itself, not only to characterize the solution of $\mathrm{P}_1(y)$ by lowering γ to 0. The properties of the solutions $x(\gamma)$ to $\mathrm{P}_1(y,\gamma)$, has been already studied in statistics although in the over-determined setting $p < n$, see the homotopy method of Osborne et al. [22], and LARS/LASSO of Efron at al. [15].

Theorem 3 ensures that, if A satisfies condition (UC), $x(\gamma)$ is always uniquely defined and belongs to $\mathcal{K}$.

Proof: Let $y \in \mathbb{R}^n$, $\gamma > 0$ and x_1 and x_2 be two solutions of $\mathrm{P}_1(y,\gamma)$. One necessarily has $Ax_1 = Ax_2$. Indeed, suppose that $Ax_1 \neq Ax_2$. Let $x_3 = \frac{1}{2}(x_1 + x_2)$, from the convexity of the norm map $x \mapsto \|x\|_1$

$$\|x_3\|_1 \leqslant \frac{1}{2}(\|x_1\|_1 + \|x_2\|_1) \tag{26}$$

From the strict convexity of the mapping $Z \mapsto \|y - Z\|_2^2$

$$\|y - Ax_3\|_2^2 < \frac{1}{2}\left(\|y - Ax_1\|_2^2 + \|y - Ax_2\|_2^2\right) \tag{27}$$

and then

$$\frac{1}{2}\|y - Ax_3\|_2^2 + \gamma\|x_3\|_1 < \frac{1}{2}\|y - Ax_1\|_2^2 + \gamma\|x_1\|_1$$

which contradicts the initial definition of x_1.

Hence if there are two minimizers x_1 and x_2 we necessarily have $Ax_1 = Ax_2$ and then $\|x_1\|_1 = \|x_2\|_1$.

Since x_1 is a minimizer of $\mathrm{P}_1(y,\gamma)$, x_1 is also a minimizer of $\mathrm{P}_1(Ax_1)$. We deduce from Theorem 1 that $x_1 \in \mathcal{K}$ and that x_1 is the unique minimizer of $\mathrm{P}_1(y)$. Thus $x_2 = x_1$ and x_1 is the unique minimizer of $\mathrm{P}_1(y,\gamma)$.

■

4 Algorithm and numerical experiments

4.1 Algorithm SupportExtension

The condition $x_0 \in \mathcal{F}$ is directly verifiable, since there is an explicit formula to define $F(x_0)$ even if the computation may be instable if the matrix A_I is badly conditioned. There is however no straightforward explicit formula guaranteeing that $x_0 \in \mathcal{K}$.

This section proposes a semi-greedy algorithm to check whether $x_0 \in \mathcal{K}$. This algortihm is built upon the following proposition.

Proposition 3 *Let $x_0 \in \mathbb{R}^p$, and A a matrix satisfying condition* (UC)*, if there exists a vector $x_1 \in \mathcal{F}$ such that x_1 is an extension of x_0, and such that for all $j \in I(x_1)\backslash I(x_0)$*

$$\text{sign}\,(x_1(j)) = -\text{sign}\,(v(j))$$

where v is the vector whose support is equal to the support of x_1 defined by $\bar{v} = (A_I^t A_I)^{-1}\text{sign}\,(\bar{x}_1)$, where $I = I(x_1)$, then there is some $\gamma_0 > 0$ such that for all $\gamma \leqslant \gamma_0$, the solution $x(\gamma)$ of $P_1(Ax_0, \gamma)$ is unique and

$$x(\gamma) = x_0 - \gamma v.$$

In the following, the notation $F^+(x_0) = F(x_1)$ is used.

To prove this proposition, it is sufficient to check that $x_0 - \gamma v$ satisfies the three conditions of Lemma 3 for small γ.

The algorithm *SupportExtension* extends vector x_0 into a vector x_1 adding or removing iteratively components to x_1 in such a way that the quantity

$$F(x_1) = \max_{j \notin I = I(x_1)} |\langle a_j, A_I^{+t}\text{sign}\,(\bar{x}_1)\rangle|$$

decreases.

The main steps of the support extension algorithm are summarized as follows:

Algorithm 1 SupportExtension

1: Set $x_1 \leftarrow x_0$, $I \leftarrow I(x_1)$.
2: **while** A_I does not have full rank and $F(x_1) \geqslant 1$ **do**
3: Compute
$$j_0 = \arg\max_{j \notin I} |\langle a_j, A_I^{+t}\text{sign}\,(\bar{x}_1)\rangle|$$
4: $x_1(j_0) \leftarrow \text{sign}\left((A_I^+ a_{j_0})^t \text{sign}\,(\bar{x}_1)\right)$, $I \leftarrow I(x_1)$
5: $\bar{v} \leftarrow (A_I^t A_I)^{-1}\text{sign}\,(\bar{x}_1)$,
6: For all $k \in I\backslash I(x_0)$ such that
$$v(k)x_1(k) > 0$$
7: Set $x_1(k) \leftarrow 0$, $I \leftarrow I(x_1)$.
8: **end while**

If the algorithm terminates by finding a vector x_1 such that $F(x_1) < 1$, then $x_0 \in \mathcal{K}$. However, if the algorithm stops only because the matrix A_I has full rank, then it is possible that $x_0 \notin \mathcal{K}$. To check the efficiency of *SupportExtension*, 200 000 couples of matrices-vectors (A, x) have been randomly chosen with different matrix sizes and different sparsity levels. For each couple, the algorithm *SupportExtension* and the ℓ_1 minimization solver *SolveBP* of the matlab toolbox *SparseLab* `http://sparselab.stanford.edu` have been applied. The finding of this experiment was that identifiability as revealed by the algorithm *SupportExtension* coincided with exact recovery by *SolveBP* for all vectors, except one identifiable vector that was not recognized by *SupportExtension*.

4.2 Computational complexity

The bulk of computational complexity of this algorithm is invested in the matrix inversion $(A_I^t A_I)^{-1}$ at each step, which is a $d \times d$ matrix, where $d = |I|$. If $p \gg n$, computing all scalar products

$\langle a_j, A_I^{+t}\text{sign}\,(\bar{x}_1)\rangle$ may be more time-consuming costing $O(pn)$ flops. But this situation has not been tested. Since rank$A_I \leqslant n$ and A_I has full rank one always has $d \leqslant n$. Moreover, numerical experiments show that the removal steps corresponding to $v(k)x_1(k) > 0$ are rare and the number of steps is in practice always bounded by n. Hence the computational complexity is $O(n^3)$.
Some simplified versions of *SupportExtension* have been tested, omiting the element removal step or selecting several indices j_0 at each step. These versions are faster but may fail to recognize a small number of identifiable vectors.

4.3 Comparison to other criteria

The identifibality criteria reviewed in the introductory part of the paper, namely F, ERC and coherence C have been compared. Since these criteria can be ranked as

$$(\|x\|_0 \leqslant C(A)) \Longrightarrow \text{ERC}(I(x)) > 0 \Longrightarrow F(x) < 1 \Longrightarrow F^+(x) < 1 \tag{28}$$

they are compared pairwise. To do so, a matrix size is fixed ($e.g. n = 300, p = 1200$) and matrices A are randomly generated from the uniform spherical ensemble. For each matrix A, each support size s between 1 and 150, a vector x_0 is generated such that $\|x_0\|_0 = s$ with random signs. For each matrix A and vector x_0, the solution of $\text{P}_1(Ax_0)$ is denoted $x^\star$. The identifiability of x_0 is measured by

$$R_A(x_0) = 1 - \frac{1}{2s}\|\text{sign}\,(\bar{x}^\star) - \text{sign}\,(\bar{x}_0)\|_0$$

where $\bar{x}$ is the vector extracted from x keeping only the s largest components in magnitude. Obviously, $0 \leq R_A(x_0) \leq 1$ and $R_A(x_0) = 1$ corresponds to an identifiable vector. The quantity $F_A^+(x_0)$ is estimated by *SupportExtension* algorithm. To ease the comparison between all criteria one defines $C_A(x_0) = \frac{1}{2}(1 + 1/C(A)) - \|x_0\|_0$ and to esae the comparison between ERC, F and F^+ one also defines $F_A(x_0) = 1 - F(x_0)$ and $F_A^+(x_0) = 1 - F^+(x_0)$.

Each point on each plot of Figure 8 corresponds to a randomly generated triplet (A, s, x_0). The plots of the top row of Figure (8) compare each criterion to its successor according to the ranking relation (28). The shaded rectangle on each plot delimits the vectors x_0 for which the criterion on the abscissa fails to recognize them as identifiable, whereas the criterion on the ordinate axis succeeds in identifying them. These plots clearly confirm the ranking relation (28) of these criteria in terms of their ability to properly recognize identifiable vectors. In particular, F_A^+ is clearly better than F and is then a sharper test of exact recovery by ℓ_1 minimization.

The plots of the second row of Figure 8 depict the exact recovery success measure R_A as a function of each identifiability criterion. The criterion F^+ is strikingly better than its competitors, showing a sharp phase transition at 1 as expected. F^+ is the only criterion showing this behaviour while the other criteria fail to positively test many identifiable vectors (shown in the gray shaded rectangles).

Another way to see the gap between conditions $ERC > 0$, $F < 1$ and $F^+ < 1$ is to compare the proportion of vectors, for a given sparsity, that satisfy these conditions. By proportion, we mean that for a given sparsity d the number of half cones of vectors with the same non-zero components and the same signs is $2^d \begin{pmatrix} d \\ p \end{pmatrix}$. Among these cones, some correspond to vectors satisfying some of the three above criteria. The goal of this experiment is to estimate the proportion of these cones. In this simulation, a size of matrix is fixed (here 200×1000). For each sparsity d, 5000 couples (A, x) are randomly generated as in the previous test, and the three citeria ERC > 0, $F < 1$ and $F^+ < 1$ are computed ($F^+ < 1$ is estimated by the algorithm *SupportExtension*). Figure 9 depicts the proportions of vectors satisfying each of the three criteria as a function of the sparsity d. This figure does not ensure that there is no vectors x that are not identifiable when $\|x\|_0 \leqslant 35$, it only shows that they are not numerous. Actually one can build vectors that are not identifiable with less than 16 non-zero components using a greedy algorithm finding a sparse vector x_0 such that $\|d_0\|_2 = \left\|A_I(A_I^t A_I)^{-1}\text{sign}\,(\bar{x}_0)\right\|_2$ is as large as possible.

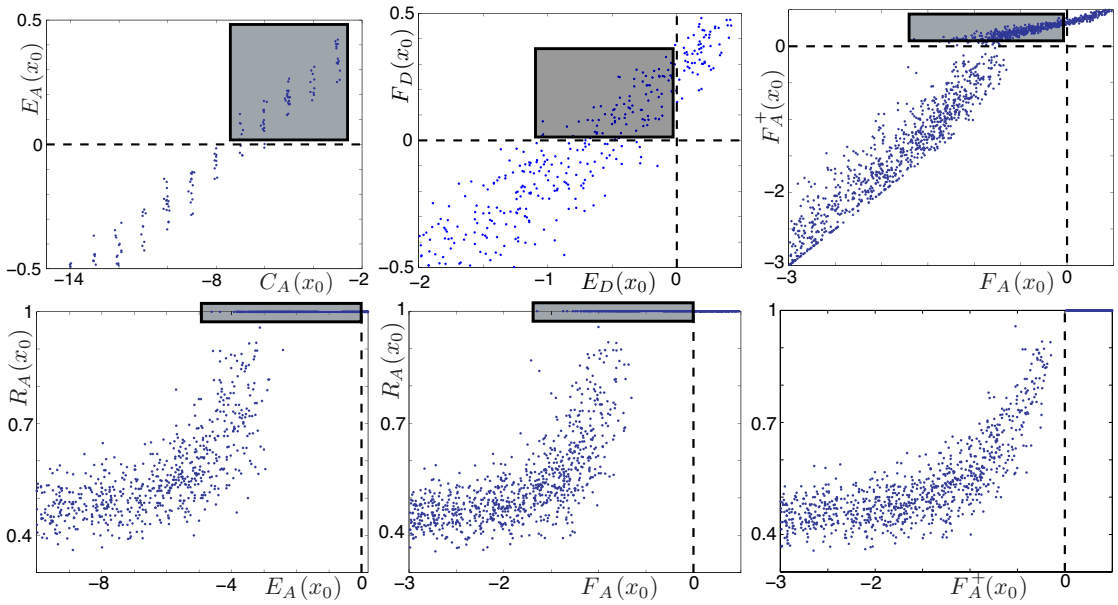


Figure 8: *First row: pairwise comparison of identifiablity criteria C, ERC, F and F^+. The shaded rectangle on each plot delimits the vectors x_0 for which the criterion on the abscissa fails to recognize them as identifiable, whereas the criterion on the ordinate axis succeeds in identifying them. Second row: exact recovery success measure R_A as a function of each identifiability criterion. The shaded rectangles show those vectors that are positively tested as identifiable by the corresponding criterion.*

JF: Acknowledgement

I would like to thank Jalal Fadili and Gabriel Peyré for their help, advice, enthusiasm and for their scientific, technical, financial and moral support. A particular thank to Gabriel for the figures and to Jalal for the hours he spent reading proofs and correcting this manuscript.

Appendix A - Proof of Lemma 5

From theorem 3, we know that $x(\gamma)$ is uniquely defined. Borrowing the arguments of the proof of proposition 2, one obtain that the support and the sign of $x(\gamma)$ varies only a finite numbers of times. More precisely $x(\gamma_1)$ and $x(\gamma_2)$ are the two minimizers of $\mathrm{P}_1(y, \gamma_1)$ and $\mathrm{P}_1(y, \gamma_2)$ with the same support and sign, one can verify using Lemma 4 that $\forall \gamma \in [\gamma_1, \gamma_2]$,

$$\bar{x}(\gamma) = A_I^+ y - \gamma (A_I^t A_I)^{-1} \mathrm{sign}\left(\bar{x}(\gamma_1)\right) \tag{29}$$

where A_I is the active matrix associated to $x(\gamma_1)$ and $x(\gamma_2)$. Hence $x(\gamma)$ is on the segment $[x(\gamma_1), x(\gamma_2)]$.

We denote $(\gamma_i)_i$ the finite sequence of values corresponding to a variation of the support of $x(\gamma)$; that is $(\gamma_i)_i$ are the values at the breakpoints of the polygonal path of $x(\gamma)$. The function $x(\gamma)$ is then locally affine, hence continuous, except at points γ_i.

It remains to show that $x(\gamma)$ is continuous on the left and on the right of points γ_i.

Let γ_{i_0} be any of these points. For all $\gamma \in]\gamma_{i_0-1}, \gamma_{i_0}[$, $x(\gamma)$ can be written $x(\gamma) = x_{i_0-1} - \gamma v_{i_0-1}$. Let's denote $x^* = x_{i_0-1} - \gamma_{i_0} v_{i_0-1}$.

By construction, the support of x^* is included in the support of $x(\gamma)$ for $\gamma \in]\gamma_{i_0-1}, \gamma_{i_0}[$. Furthermore, x^* satisfies both conditions of Lemma 4 with $\gamma = \gamma_{i_0}$. Then, $x^* = x(\gamma_{i_0})$. Using similar arguments, we can also show that $x(\gamma_{i_0})$ is the limit of $x(\gamma)$ when γ tends to γ_{i_0} on the right. We then obtain that $x(\gamma)$ is a piecewise affine and continuous function of γ.

Since for all y and $\gamma > 0$, $x(\gamma) = A_I^+ y - \gamma (A_I^t A_I)^{-1} \mathrm{sign}\left(x(\gamma)\right)$ and since $(A_I^t A_I)^{-1} \mathrm{sign}\left(x(\gamma)\right)$ can take a finite number of values, there's a real number C_0 depending on A but not on y, such that $x(\gamma)$ is C_0-Lipschitz. This concludes the proof. ∎

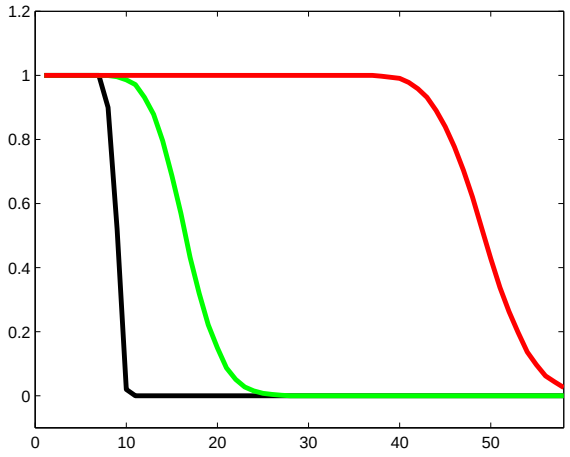


Figure 9: *Proportions of vectors satisfying each of the three criteria $ERC > 0$, $F < 1$ and $F^+ < 1$ as a function of sparsity d. The black line corresponds to the condition $ERC > 0$, the green line to $F < 1$ and the red one to $F^+ < 1$.*

Appendix B - Proof of Lemma 6

Let $y_0 \in \mathrm{Im}\,(A)$ and $x_0 = \varphi(y_0)$, the solution of $\mathrm{P}_1(y_0)$. We denote $x_0(\gamma)$ the solution of $\mathrm{P}_1(y_0, \gamma)$.
From lemma 5, $x(\gamma)$ lives on a polygonal path, and from the proof of proposition 2, one knows that $x(\gamma_n) \in \mathcal{F}$ for a sequence of γ_n tending to zero, thus one can deduce that there is a non negative real number γ_0 and a vector v such that $\forall \gamma \in]0, \gamma_0[$ such that $x_0(\gamma) \in \mathcal{F}$, $\mathrm{supp}\,(x_0) \subset \mathrm{supp}\,(x_0(\gamma)) = \mathrm{supp}\,(x_0(\gamma_0)) = I$ and $x_0(\gamma) = x_0 + \gamma v$. One can suppose $\gamma_0 \leqslant \frac{x_{\min}}{2\|v\|_\infty}$, where $x_{\min}$ denotes the smallest absolute value of non zero component of x_0.
Moreover for $\gamma \leqslant \gamma_0$ and for all inactive vector $(a_j)_{j \notin I}$,

$$|a_j^t(y - A_I \bar{x_0}(\gamma))| \leqslant \gamma F(x_0(\gamma_0)) = \gamma F^+ < \gamma. \tag{30}$$

For a matrix B, one denotes $\|B\|_{2,\infty} = \sup_{x \neq 0} \dfrac{\|Bx\|_\infty}{\|x\|_2}$.

Let $\varepsilon < \frac{\gamma_0 v_{\min}}{2\|A_I^+\|_{2,\infty}}$ and $y_1 \in \mathrm{Im}\,(A)$ such that $\|y_1 - y_2\|_2 \leqslant \varepsilon$, where $v_{\min}$ denotes the smallest absolute value of non zero component of v. For all $\gamma \in [\frac{2\varepsilon\|A_I^+\|_{2,\infty}}{v_{\min}}, \gamma_0]$, one defines vector $x_1^*(\gamma)$ whose support is equal to I and defined by

$$\bar{x}_1^*(\gamma) = \bar{x}_0 + \gamma \bar{v} + A_I^+(y_1 - y_0) \tag{31}$$

where $\bar{x}$ is obtained keeping components of x indexed by I. Hence, with this definition, $\bar{x}_0$ may have some zeros components.
To prove $x_1^*(\gamma)$ is the solution of $\mathrm{P}_1(y_1, \gamma)$, one first shows that $\mathrm{sign}\,(\bar{x}_1^*(\gamma)) = \mathrm{sign}\,(\bar{x}_0(\gamma))$.

$$\|\gamma \bar{v}\|_\infty \leqslant \frac{x_{\min}}{2} \tag{32}$$

$$\left\|A_I^+(y_1 - y_0)\right\|_\infty \leqslant \left\|A_I^+\right\|_{2,\infty} \varepsilon \leqslant \frac{\gamma v_{\min}}{2} \leqslant \frac{x_{\min}}{4} \tag{33}$$

Hence $\bar{x}_1^*(\gamma)$ is the sum of three vectors $\bar{x}_0$, $\gamma\bar{v}$ and $A_I^+(y_1 - y_0)$. The sign of this sum will be given by $\bar{x}_0(i)$ if it's non zero and by $\gamma\bar{v}(i)$ if it is. Then, if for $i \in I$, $x_0(i) \neq 0$, then

$$\mathrm{sign}\,(x_1^*(\gamma)(i)) = \mathrm{sign}\,(x_0(i)) = \mathrm{sign}\,(x_0(\gamma)(i)) \tag{34}$$

else if for $j \in I, x_0(j) = 0$, then

$$\text{sign}\left(x_1^*(\gamma)(j)\right) = \text{sign}\left(\gamma v(j)\right) = \text{sign}\left(x_0(\gamma)(j)\right) \tag{35}$$

Hence, $\text{sign}\left(\bar{x}_1^*(\gamma)\right) = \text{sign}\left(\bar{x}_0(\gamma)\right)$.
Using lemma 4 one proves that $x_1^*(\gamma)$ is the solution of $\mathrm{P}_1(y_1, \gamma)$, if moreover

$$\varepsilon \leqslant \frac{\gamma(1 - F^+)}{\max_j \|a_j\|_2} \tag{36}$$

On a first hand we have

$$\begin{aligned} A_I^t(y_1 - A_I \bar{x}_1^*(\gamma)) &= A_I^t(y_0 - A_I \bar{x}_0(\gamma)) + A_I^t(y_1 - y_0 - A_I A_I^+(y_1 - y_0)) \\ &= \gamma \text{sign}\left(\bar{x}_0(\gamma)\right) \\ &= \gamma \text{sign}\left(\bar{x}_1^*(\gamma)\right). \end{aligned}$$

On a second hand, for all inactive vector $(a_j)_{j \notin I}$, we have

$$\begin{aligned} |a_j^t(y_1 - A_I \bar{x}_1^*(\gamma))| &\leqslant |a_j^t(y_0 - A_I \bar{x}_0(\gamma))| + |a_j^t(y_1 - y_0 + A_I A_I^+(y_1 - y_0))| \\ &\leqslant \gamma F^+ + \varepsilon \|a_j\|_2 \\ &\leqslant \gamma \end{aligned}$$

and then $x_1^*(\gamma) = x_1(\gamma)$ is the solution of $\mathrm{P}_1(y_1, \gamma)$. Finnally if $\gamma \leqslant \gamma_0$ and

$$\varepsilon \leqslant \min\left(\frac{\gamma(1 - F^+)}{\max_j \|a_j\|_2}, \frac{\gamma v_{\min}}{2\left\|A_I^+\right\|_{2,\infty}}\right) = \frac{\gamma}{C_1} \tag{37}$$

then,

$$\|x_1(\gamma) - x_0(\gamma)\|_2 = \left\|A_I^+(y_1 - y_0)\right\|_2 \leqslant C_2 \varepsilon. \tag{38}$$

Constants C_1 and C_2 can take only a finite number of values and then taking $\gamma = \max_{x_0}(C_1)\varepsilon$, which is always possible if $\varepsilon \leqslant \varepsilon_0 = \frac{\gamma_0}{\max(C_1)}$.

References

[1] E.J. Candès and D.L. Donoho. A surprinsingly effective nonadaptive representation for objects with edges. Curves and Surfaces, 1999.

[2] E.J. Candès and J.Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. Found. of Comput. Math., 6:227–254, 2004.

[3] E.J. Candès, J. Romberg, and T.Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory, 52:489–509, 2004.

[4] E.J. Candès, J. Romberg, and T.Tao. Stable signal recovery from incomplete and inacurate measurements. Pure Appl. Math., 59, 2005.

[5] E.J. Candès and T.Tao. Near optimal signal recovery from random projections and universal encoding strategies. IEEE trans. Inform. Theory, 52:5406–5425, 2004.

[6] E.J. Candès and T.Tao. The dantzig selector: statistical estimation when p is much larger than n. 2005.

[7] E.J. Candès and T.Tao. Decoding by linear programming. IEEE Trans. Inform. Theory., 51, 2005.

[8] E. J. Candès. Compressive sampling. Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006.

[9] D.L. Donoho. Neighborly Polytopes and Sparse Solutions of Underdetermined Linear Equations. Tech. Report, Department of statistics, Stanford University, 2004.

[10] D.L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionnaries via l^1 minimization. PNAS, 100:2197–2202, 2002.

[11] D.L. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inform. Theory, 52:6–18, 2006.

[12] D.L. Donoho and X. Huo. Uncertainty principles and ideal decomposition. IEEE Trans. Inform. Theory, 47:2845–2862, 2001.

[13] D.L. Donoho and J.Tanner. Sparse Nonnegative Solutions of Undeterminned Linear Equations by linear Programming. 2005.

[14] D.L. Donoho and J. Tanner. Neighborliness of Randomly-Projected Simplices in High Dimensions. 2005.

[15] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least Angle Regression. Annals of statistics, 32(2):407–499, 2004.

[16] M. Elad and A.M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. IEEE transactions on information theory, 48, 2002.

[17] R. Fletcher. Practical methods of optimization. 1987.

[18] J.J. Fuchs. On sparse representations in arbitrary redundant bases. IEEE Trans. Inform. Theory, 50:1341–1344, 2004.

[19] J.J. Fuchs. Recovery of exact sparse representations in the presence of bounded noise. IEEE Trans. Inform. Theory, 51:3601–3608, 2005.

[20] R. Gribonval and M. Nielsen. Sparse representation in unions of bases. IEEE Trans. Inform. Theory, 49:3320–3325, 2003.

[21] S. Mallat. A wavelet tour of signal processing. 1998.

[22] M. R. Osborne, Brett Presnell, and B. A. Turlach. A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis, 20(3):389–403, 2000.

[23] J.A. Tropp. Just Relax: Convex Programming Methods for Identifying Sparse Signals in Noise. IEEE transactions on informnation theory, 2005.

[24] R.A. De Vore. Deterministic constructions of compressed sensing matrices. 2007.