Sparsest representations and approximations of a high-dimensional linear system

Patrick Tardivel, Remi Servien, Didier Concordet

Sparsest representations and approximations of a high-dimensional linear system

Patrick J.C. Tardivel*, Rémi Servien and Didier Concordet
Toxalim, Université de Toulouse, INRA, ENVT, Toulouse, France.

Abstract

In a high-dimensional linear system of equations, constrained l^1 minimization methods such as the basis pursuit or the lasso are often used to recover one of the sparsest representations or approximations of the system. The null space property is a sufficient and "almost" necessary condition to recover a sparsest representation with the basis pursuit. Unfortunately, this property cannot be easily checked. On the other hand, the mutual coherence is an easily checkable sufficient condition insuring the basis pursuit to recover one of the sparsest representations. Because the mutual coherence condition is too strong, it is hardly met in practice. Even if one of these conditions holds, to our knowledge, there is no theoretical result insuring that the lasso solution is one of the sparsest approximations. In this article, we study a novel constrained problem that gives, without any condition, one of the sparsest representations or approximations. To solve this problem, we provide a numerical method and we prove its convergence. Numerical experiments show that this approach gives better results than both the basis pursuit problem and the reweighted l^1 minimization problem.

Keywords: Basis pursuit, Lasso, Sparsest representations, Sparsest approximations.

1 Introduction

We consider a vector $y \in \mathbb{R}^n$ and a family of vectors $\mathcal{D} = \{d_1, \dots, d_p\}$ spanning $\mathbb{R}^n$. An $\epsilon-$approximation of y in $\mathcal{D}$ is a vector $x = (x_1, \dots, x_p)$ such that $\|y - (x_1 d_1 + \dots + x_p d_p)\|^2 \leq \epsilon$. The aim of this article is to find at least one of the sparsest $\epsilon-$approximations of y when $p > n$. These sparsest $\epsilon-$approximations are defined as the solutions of

$$S_0^\epsilon := \text{argmin } \|x\|_0 \text{ subject to } \|y - Dx\|^2 \leq \epsilon \qquad (\mathcal{P}_0^\epsilon)$$

where $\|x\|_0 := \text{Card}\{i \in [\![1,p]\!] \mid x_i \neq 0\} = \sum_{i=1}^p \mathbb{1}_{x_i \neq 0}$ is the l^0 "norm" of x and $D := (d_1|\dots|d_p)$ is the $n \times p$ matrix whose columns are the vectors $(d_j)_{1 \leq j \leq p}$.

A first simplified problem is to look for the sparsest representations of y in $\mathcal{D}$ corresponding to the solutions of $\mathcal{P}_0^0$ namely

$$S_0 := \text{argmin } \|x\|_0 \text{ subject to } Dx = y. \qquad (\mathcal{P}_0)$$

*corresponding author: patrick.tardivel@inra.fr

A simple way to solve $\mathcal{P}_0$ is to compute $\tilde{x} = \tilde{D}^{-1}y$ for all $n \times n$ invertible submatrices $\tilde{D}$ of D and to select the $\tilde{x}$ with the lowest l^0 "norm". The number of such $n \times n$ submatrices of D is $\binom{p}{n}$. When $p \gg n$ this number is huge rending the previous approach intractable. So, other approaches such as the basis pursuit problem, denoted $\mathcal{P}_1$, have been proposed [13, 14, 18]. Under some conditions, given hereafter, the problem

$$\text{argmin } \|x\|_1 \text{ subject to } Dx = y \tag{$\mathcal{P}_1$}$$

has a unique solution that is also a solution of $\mathcal{P}_0$. The standard approach to know if a solution of $\mathcal{P}_1$ is also a solution of $\mathcal{P}_0$ is to compute a solution for $\mathcal{P}_1$ and to check whether or not one of these conditions holds for it. When the solution of $\mathcal{P}_1$ does not meet any of these conditions, we do not know if it belongs to S_0.

The null space property [11, 13, 14, 18] is probably the most known condition. However, as pointed out by Tillmann et al. [32], this condition is uncheckable. Another condition is the restricted isometry property detailed in [5, 6, 7, 8, 17]. However, this condition is not easy to use because the computation of the restricted isometry constant is intractable [32]. On the contrary, the mutual coherence and the spark conditions [13, 18] are easily checkable. Unfortunately, none of these four conditions (null space property, restricted isometry property, mutual coherence and spark conditions) hold for the basis pursuit solution as soon as its l^0 "norm" is greater or equal to $(n+1)/2$. In this case, the solutions of $\mathcal{P}_1$ does not give any information on those of $\mathcal{P}_0$. Moreover, even if the l^0 "norm" of the sparsest representation is strictly smaller than $(n+1)/2$, the numerical comparisons of [9] illustrate that the solution of the basis pursuit may not be a solution of $\mathcal{P}_0$.

An intuitive alternative approach consists in the approximation of the l^0 "norm" in $\mathcal{P}_0$ by a surrogate function with nice properties. As an example, the function $\sum_{i=1}^{p} \ln(1 + |x_i|/\delta)$ has been studied as an approximation of the l^0 "norm" [9, 24], leading to the following problem

$$\text{argmin } \sum_{1 \leqslant i \leqslant p} \ln(1 + |x_i|/\delta) \text{ subject to } Dx = y. \tag{1}$$

An iterative method converging to a stationary point of the problem (1) is provided in [24]. With some well chosen δ, simulations show that this heuristic approach gives better results than the basis pursuit. However, nothing guarantees that the solutions of (1) are also solutions of $\mathcal{P}_0$ and the choice of δ plays a major role on the performances of the method.

When $\epsilon > 0$, the problem $\mathcal{P}_0^\epsilon$ is even more complicated and still intractable. Similarly to the basis pursuit problem $\mathcal{P}_1$, one can substitute in $\mathcal{P}_0^\epsilon$ the l^0 "norm" by a l^1 norm. This leads to the following problem

$$\text{argmin } \|x\|_1 \text{ subject to } \|y - Dx\|_2^2 \leqslant \epsilon. \tag{$\mathcal{P}_1^\epsilon$}$$

This problem $\mathcal{P}_1^\epsilon$ can be rewritten as a lasso problem [30]:

$$\text{argmin } \|y - Dx\|^2 + \lambda\|x\|_1. \qquad (\mathcal{P}(\lambda))$$

Actually, there exists a (not explicit) bijection between λ et ϵ guaranteeing that both problems have the same solution ; see [1] (chapter 5.3) for more details.

To our knowledge, there is no theoretical result insuring that $x(\lambda)$, the unique solution of $\mathcal{P}(\lambda)$, is an element of S_0^ϵ. Instead, there exists a lot of conditions that state the convergence of $x(\lambda)$ to a solution $x^* \in S_0$ when λ converges to 0 [4, 14, 15, 33, 34]. Among these conditions (for an exhaustive list, see [3] page 177), the two most known are probably the irrepresentable condition [26, 37, 38] and the compatibility condition [33]. In practice all these conditions are not easily checkable. Furthermore, when these conditions do not hold the solution obtained with the basis pursuit or with the lasso can be very far from the set S_0^ϵ we wish to recover.

The aim of this article is to propose a new tractable problem which allows to catch one of the sparsest representations (element of S_0) or one of the sparsest $\epsilon-$approximations (element of S_0^ϵ). To obtain such solutions, we define and solve the following problem

$$S_{f_\alpha}^\epsilon := \text{argmin} \sum_{i=1}^{p} f_\alpha(|x_i|) \text{ subject to } \|y - Dx\|^2 \leq \epsilon.$$

We provide functions $f_\alpha : \mathbb{R}_+ \to \mathbb{R}$, depending on a parameter $\alpha > 0$, guaranteeing without any condition that

- when $\epsilon = 0$, there exists α_0 such that whatever $0 < \alpha \leq \alpha_0$, the previous problem is "almost equivalent" to $\mathcal{P}_0$ since $S_{f_\alpha}^0 \subset S_0$,
- when $\epsilon > 0$, $S_{f_\alpha}^\epsilon$ becomes arbitrary close to S_0^ϵ when α converges to 0.

This article is organized as follows. In section 2, we study the case $\epsilon = 0$. We prove that there exists α_0 such that, whatever $\alpha \leq \alpha_0$, each element of $S_{f_\alpha}^0$ is a solution of $\mathcal{P}_0$ and that a Maximisation Minimisation (MM) method provides an iterative sequence which converges to a local minimum of $\mathcal{P}_0$. Section 3 is dedicated to the case $\epsilon > 0$. We prove that $S_{f_\alpha}^\epsilon$ becomes arbitrary close to the set S_0^ϵ when α converges to 0 and we give necessary conditions that must satisfy the limit points of the iterative sequence provided by the MM method. We also exhibit a subset of S_0^ϵ that fulfilled these necessary conditions. The section 4 is devoted to simulations. Numerical experiments show that this approach gives better results to recover one of the sparsest representations than both the basis pursuit problem $\mathcal{P}_1$ and the reweighted l^1 minimization problem.

2 A sparsest representation

As already explained, solve $\mathcal{P}_0$ is difficult. Replacing the l^0 "norm" by a l^1 norm leads to the problem $\mathcal{P}_1$ which provides sparse solutions. However, the conditions guaranteeing that a solution of $\mathcal{P}_1$ is also a solution of $\mathcal{P}_0$ are unverifiable. The substitution in $\mathcal{P}_0$ of the l^0 "norm" by a l^α "norm" with $\alpha < 1$ gives the following problem $\mathcal{P}_\alpha$ which also has sparse solutions

$$S_\alpha := \text{argmin } \|x\|_\alpha \text{ subject to } Dx = y, \tag{$\mathcal{P}_\alpha$}$$

where $\|x\|_\alpha = \left(\sum_{i=1}^p |x_i|^\alpha\right)^{1/\alpha}$ is the l^α "norm" of the vector x. The study of this problem has been the subject of an abundant literature [10, 16, 18, 19, 22, 28, 36]. The problem $\mathcal{P}_\alpha$ provides a sparsest representation as soon as the null space property condition [18] or the restricted isometry property [10, 16, 22, 28] hold. As for the basis pursuit, these conditions are uncheckable.

In this section we show that there exists $\alpha_0 > 0$ such that the solutions of $\mathcal{P}_\alpha$ are also solutions of $\mathcal{P}_0$ as soon as $\alpha < \alpha_0$. When $\alpha < 1$, the function $x = (x_1, \dots, x_p) \mapsto \|x\|_\alpha$ is a concave function on each domain of the form $I_1 \times \cdots \times I_p$, with $I_k =]-\infty, 0]$ or $I_k = [0, +\infty[$. Solving $\mathcal{P}_\alpha$ leads to minimize a locally concave function on a convex set. This is not a convex optimization problem. In this respect, we propose in this section a numerical method to solve it. We can generalize the problem $\mathcal{P}_\alpha$ by substituting the function $|x_i|^\alpha$ by a function $f_\alpha(|x_i|)$. This modification leads to minimize an expression of the form $\sum_{i=1}^p f_\alpha(|x_i|)$. Intuitively, by comparing $\sum_{i=1}^p f_\alpha(|x_i|)$ with the l^α "norm", one sees that the function $\sum_{i=1}^p f_\alpha(|x_i|)$ should simply converge to $\|.\|_0$ and should have level sets that look like spheres for the l^α "norm". A geometric interpretation linking the shape of the spheres of the l^α "norm" to the sparseness of the solutions of $\mathcal{P}_\alpha$ is given in [20]. In the theorem 1, we focus on the following problem

$$S_{f_\alpha} := \text{argmin } \sum_{1 \leqslant i \leqslant p} f_\alpha(|x_i|) \text{ subject to } y = Dx. \tag{$\mathcal{P}_{f_\alpha}$}$$

Without any condition, we prove that the solutions of $\mathcal{P}_{f_\alpha}$ are also solutions of $\mathcal{P}_0$ as soon as α is small enough.

Theorem 1 *Let f_α be a function defined on $\mathbb{R}_+$ strictly increasing and strictly concave such that*

$$\forall x \in \mathbb{R}_+, \lim_{\alpha \to 0} f_\alpha(x) = \mathbb{1}_{x \neq 0}.$$

Then, there exists $\alpha_0 > 0$ such that for all $\alpha \in (0, \alpha_0)$, $S_{f_\alpha} \subset S_0$.

The α_0 threshold depends on D and y and its value is quite hard to infer except in few cases (see [29]). However, since the $\mathcal{P}_{f_\alpha}$ allows to capture a part of S_0 for all $\alpha < \alpha_0$, one can choose *a priori* a very small α so that we can expect it is less than α_0. A study of the problem $\mathcal{P}_{f_\alpha}$ where the functions f_α have different properties that those given in the theorem 1 is given in [35]. The authors proved that the problem $\mathcal{P}_{f_\alpha}$ catches an element of S_0 under the conditions that the l_0 "norm" of the sparsest representation is smaller than $n/2$ and that the matrix D satisfies the unique representation property.

In the theorem 1, we made relatively weak assumptions on the f_α functions. Indeed, a function f_α for which the properties of the theorem 1 hold can be not derivable on $(0, +\infty)$ or not continuous in 0. Because the numerical resolution of the problem $\mathcal{P}_{f_\alpha}$ requires some regularity, we restrict ourselves to functions f_α which are differentiable on $(0, +\infty)$. Numerically, we solve the problem $\mathcal{P}_{f_\alpha}$ using a MM method [21] popularized in statistics by the EM algorithm [12]. This method iteratively alternates two steps. First a function that majorizes the function $\sum_{1 \leqslant i \leqslant p} f_\alpha(|x_i|)$ is defined. Then this majorazing function is minimized.

In a similar way as in [9, 24], we define a sequence $(x^{(k)})_{k\in\mathbb{N}}$ by "linearising" the function $\sum_{1\leqslant i\leqslant p} f_\alpha(|x_i|)$ at the point $x^{(k)} \in \mathbb{R}^p$. This "linearisation" (we use quotation because this function is not affine) gives the function $x \in \mathbb{R}^p \mapsto \sum_{1\leq i\leq p} f_\alpha(|x_i^{(k)}|) + f'_\alpha(|x_i^{(k)}|)(|x_i| - |x_i^{(k)}|)$. Because f is concave on $\mathbb{R}_+$, we have

$$\forall x \in \mathbb{R}^p, \sum_{1\leqslant i\leqslant p} f_\alpha(|x_i|) \leq \sum_{1\leq i\leq p} f_\alpha(|x_i^{(k)}|) + f'_\alpha(|x_i^{(k)}|)(|x_i| - |x_i^{(k)}|).$$

Then, this majorizing function is minimized with respect to x leading to $x^{(k+1)}$. More precisely, we choose $x^{(0)} \in \mathbb{R}^p$ and we set $x^{(k+1)}$ as the solution of the following weighted basis pursuit problem

$$\begin{aligned} x^{(k+1)} &:= \operatorname{argmin} \sum_{1\leq i\leq p} f_\alpha(|x_i^{(k)}|) + f'_\alpha(|x_i^{(k)}|)(|x_i| - |x_i^{(k)}|) \text{ subject to } Dx = y, \\ &= \operatorname{argmin} \sum_{i=1}^{p} f'_\alpha(|x_i^{(k)}|)|x_i| \text{ subject to } Dx = y. \end{aligned}$$

Note that without any other consideration, nothing guarantees that $x^{(k+1)}$ is unique. The general position condition for D (as defined in [31]) is a sufficient condition for the uniqueness of $x^{(k+1)}$ [27]. The general position condition is very weak. Indeed, when D is a random matrix with a continuous distribution on the set of the $n \times p$ matrix, the general position condition holds almost surely [31]. Consequently, in practice, the uniqueness of the basis pursuit solution always holds.

The first iteration of the previous MM method gives a vector $x^{(1)}$ solution of the weighted basis pursuit problem. This vector has a large number of null components. When f is right differentiable at 0, as for small α the quantity $f'_\alpha(0)$ is very large (because $\lim_{\alpha\to 0} f'_\alpha(0) = +\infty$), the null components of $x^{(1)}$ will be strongly weighted implying that the algorithm will get stuck at this point. To avoid this problem, we propose to iteratively solve the following approximate problem that gives less weight on null components

$$x^{(k+1)} := \operatorname{argmin} \sum_{1\leqslant i\leqslant p} f'_\alpha(|x_i^{(k)}| + \Delta)|x_i| \text{ subject to } Dx = y. \tag{2}$$

The theoretical results justifying the introduction of Δ are provided in the theorem 2 and proposition 1.

Theorem 2 *For every $x^{(0)} \in \mathbb{R}^p$, for every $\Delta > 0$, there exists an integer k_0 such that $\forall k \geqslant k_0$, the sequence $x^{(k)}$ defined in (2) is so that $x^{(k)} = x^{(k_0)}$.*

A similar theorem that deals only with the convergence of the iterative method in the special case where

$f_\alpha(x) = \log(1 + x/\alpha)$ already denoted as (1) is given in [24]. This theorem shows that the iterative sequence converges onto a stationary point of the problem $\min \sum_{1 \leqslant i \leqslant p} \log(1 + |x_i|/\alpha)$ subject to $Dx = y$ which is not *a priori* a local minimum of $\mathcal{P}_0$. Moreover, the proposed proof in [24] seems incorrect because even for a bounded sequence, the fact that $\lim_{k \to +\infty} x_i^{(k+1)} - x_i^{(k)} = 0$ does not imply the convergence of $(x_i^{(k)})_{k \in \mathbb{N}}$. The proposition 1 states the limit of the sequence $(x^{(k)})_{k \in \mathbb{N}}$ defined in (2) is a local minimum of the problem $\mathcal{P}_0$.

Proposition 1 *Let $(x^{(k)})_{k \in \mathbb{N}}$ be the sequence defined in (2) and l its limit then, there exists a radius $r > 0$ such that $\forall x \in B_\infty(l, r)$ with $Dx = y$ and $x \neq l$, we have $\|x\|_0 > \|l\|_0$.*

The limit l given in the previous proposition depends on $x^{(0)} \in \mathbb{R}^p$ and $\Delta > 0$. In Section 4 we discuss the choice of the initial point $x^{(0)}$ and we propose to test different values for Δ in order to keep the local minimum having the lowest l^0 "norm".

3 Sparsest $\epsilon-$approximations

In the previous section, we obtained one of the sparsest representations of y by solving the problem $\mathcal{P}_{f_\alpha}$ instead of $\mathcal{P}_0$ with α small enough. Similarly, to solve the intractable problem $\mathcal{P}_0^\epsilon$, one substitutes the constraint $Dx = y$ that appears in the problem $\mathcal{P}_{f_\alpha}$ by the constraint $\|y - Dx\|_2^2 \leq \epsilon$. This modification leads to consider

$$S_{f_\alpha}^\epsilon := \operatorname{argmin} \sum_{1 \leqslant i \leqslant p} f_\alpha(|x_i|) \text{ subject to } \|y - Dx\|^2 \leq \epsilon. \qquad (\mathcal{P}_{f_\alpha}^\epsilon)$$

The following theorem 3 shows that, when α is small enough, the set $S_{f_\alpha}^\epsilon$ is arbitrary close to the set S_0^ϵ of solutions of $\mathcal{P}_0^\epsilon$. This justifies to solve $\mathcal{P}_{f_\alpha}^\epsilon$ instead of $\mathcal{P}_0^\epsilon$. There are situations in which solving $\mathcal{P}_{f_\alpha}^\epsilon$, with a small enough α, gives one of the sparsest approximations. However, there are situations in which it is not the case. Unfortunately, we do not have any general criterion separating these two cases. This is the reason why, we propose the following theorem that states that the solutions of $\mathcal{P}_{f_\alpha}^\epsilon$ are arbitrarily close to S_0^ϵ. For this theorem, we introduce the $\eta-$magnification of the set S_0^ϵ. It is defined as the open set $G_\eta := \bigcup_{x \in S_0^\epsilon} B(x, \eta)$, where $B(x, \eta)$ is an l^2 open ball of radius $\eta > 0$ centered in x.

Theorem 3 *Let $(f_\alpha)_{\alpha > 0}$ be a family of strictly increasing, strictly concave and continuous functions defined on $\mathbb{R}_+$ such that*

$$0 < \alpha \leq \alpha' \Rightarrow f_\alpha \geq f_{\alpha'} \text{ and } \forall x \in \mathbb{R}_+ \lim_{\alpha \to 0} f_\alpha(x) = \mathbb{1}_{x \neq 0}.$$

Then, for all $\eta > 0$, there exists $\alpha_0 > 0$ such that the following inclusion holds

$$\forall \alpha \leq \alpha_0, S_{f_\alpha}^\epsilon \subset G_\eta.$$

Such families of functions may appear difficult to build, but this is not the case. As an example, the assumptions of theorem 3 hold for the families of functions $f_\alpha : x \in \mathbb{R}_+ \mapsto x/(\alpha + x)$ and $f_\alpha : x \in \mathbb{R}_+ \mapsto \arctan(x/\alpha)$. The figure 1 illustrates this result in two different cases. In the first case, with a small enough α, the problem $\mathcal{P}^{\epsilon}_{f_\alpha}$ captures one of the sparsest approximations. In the second case, whatever $\alpha > 0$, the solution of the problem $\mathcal{P}^{\epsilon}_{f_\alpha}$ is not one of the sparsest approximations but stays close to S^{ϵ}_0.

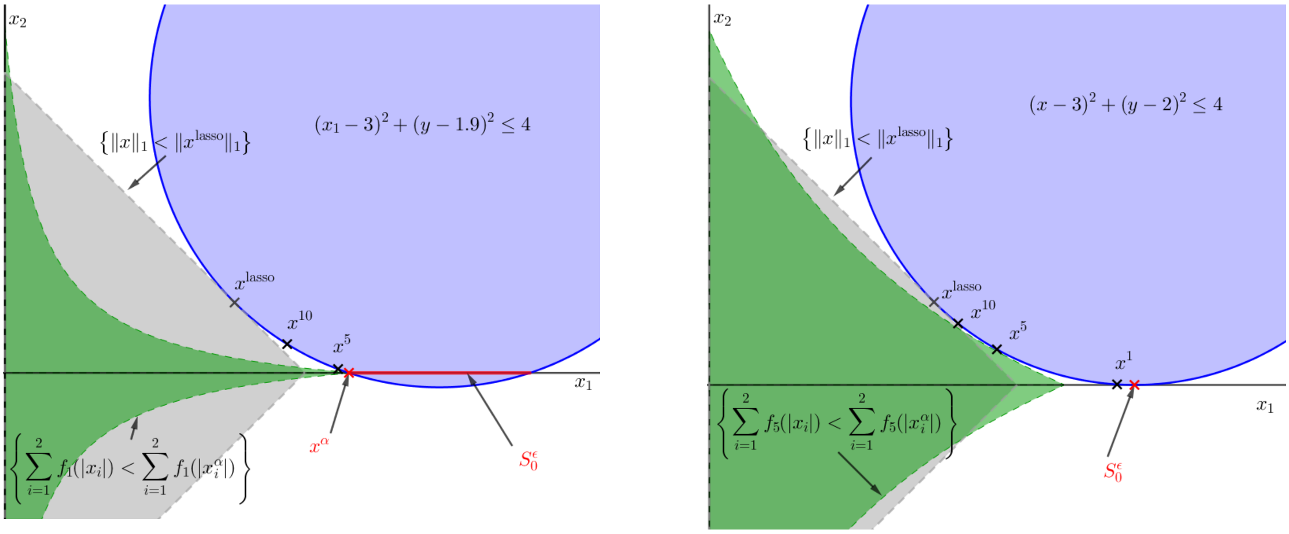

Figure 1: Let f_α be the function $f_\alpha : x \in \mathbb{R}_+ \mapsto x/(x + \alpha)$ with $\alpha > 0$. On the left, we represent the solution of the problem argmin $\sum_{i=1}^{2} f_\alpha(|x_i|)$ subject to $(x_1 - 3)^2 + (x_2 - 1.9)^2 \leq 4$ for several values of α and the solution of the lasso problem argmin $\sum_{i=1}^{2} |x_i|$ subject to $(x_1 - 3)^2 + (x_2 - 1.9)^2 \leq 4$ denoted x^{lasso}. The points x^{10}, x^5 and x^α are the solutions of the first problem when $\alpha = 10$, $\alpha = 5$ and $\alpha \leq \alpha_0$ with $\alpha_0 \approx 4.5$. Geometrically, x^α and x^{lasso} are respectively the unique solution of the first problem with $\alpha = 1$ and of the lasso problem because the "open balls" $\{\sum_{i=1}^{2} f_1(|x_i|) < \sum_{i=1}^{2} f_1(|x_i^\alpha|)\}$ (in green) and $\{\|x\|_1 < \|x^{\text{lasso}}\|_1\}$ (in grey) do not share any point with the constraint set $(x_1 - 3)^2 + (x_2 - 1.9)^2 \leq 4$ (in blue). Note that when $\alpha \leq \alpha_0$, the first problem catches an element x^α of S^{ϵ}_0 (in red). On the right, we represent the solution of the lasso problem and the solutions x^{10}, x^5, x^1 of the problem argmin $\sum_{i=1}^{2} f_\alpha(|x_i|)$ subject to $(x_1 - 3)^2 + (x_2 - 2)^2 \leq 4$ when $\alpha = 10, \alpha = 5$ and $\alpha = 1$. In addition we draw the "open balls" $\{\sum_{i=1}^{2} f_5(|x_i|) < \sum_{i=1}^{2} f_5(|x_i^5|)\}$ (in green) and $\{\|x\|_1 < \|x^{\text{lasso}}\|_1\}$ (in grey). When α is small the solution is close to S^{ϵ}_0. However, one can prove that whatever $\alpha > 0$, this second problem never catches exactly an element of S^{ϵ}_0.

In the previous section, we have seen that a MM method provides a sequence (2) which is stationary from a certain rank onto a local minimum of the problem $\mathcal{P}_0$. To solve the problem $\mathcal{P}^{\epsilon}_{f_\alpha}$, one uses the same MM method as in (2) leading to the iterative sequence given hereafter. Let $x^{(0)} \in \mathbb{R}^p$ and define the sequence $(x^{(k)})_{k \in \mathbb{N}}$ as follows

$$x^{(k+1)} := \text{argmin} \sum_{1 \leqslant i \leqslant p} f'_\alpha(|x_i^{(k)}| + \Delta)|x_i| \text{ subject to } \|y - Dx\|^2 \leq \epsilon. \tag{3}$$

Similarly to the basis pursuit problem, the lasso problem (3) does not always have an unique solution. However, the general position condition for D is sufficient to insure the uniqueness of the lasso solution [27, 31]. As already explained, the general position condition is very weak [31] and, in practice, the uniqueness of the lasso solution always occurs.

In the theorem 4, we prove that the sequence $(x^{(k)})_{k\in\mathbb{N}}$, as defined in (3), is bounded that is, when k is large enough, $x^{(k)}$ is close to a limit point. The theorem 4 shows that the optimality conditions hold for the limit points of the sequence $(x^{(k)})_{k\in\mathbb{N}}$.

Theorem 4 *Let $(f_\alpha)_{\alpha>0}$ be a family of increasing, concave and two times differentiable functions defined on $(0,+\infty)$ such that $\forall\alpha>0, f'_\alpha$ is convex and*

$$\forall x\in\mathbb{R}_+ \lim_{\alpha\to 0} f_\alpha(x) = \mathbb{1}_{x\neq 0}.$$

Then :

1. *The sequence $(x^{(k)})_{k\in\mathbb{N}}$ described in (3) is bounded.*
2. *For any limit point $\tilde{x}$ of the sequence $(x^{(k)})_{k\in\mathbb{N}}$, we have*
   - *i) The vector $\tilde{x}$ is on the boundary of the constraints' set thus, $\|y-D\tilde{x}\|^2=\epsilon$.*
   - *ii) The family of D matrix columns $(d_i)_{i\in\mathrm{supp}(\tilde{x})}$ is linearly independent.*
   - *iii) The vectors $(d_i^T(y-D\tilde{x}))_{i\in\mathrm{supp}(\tilde{x})}$ and $(f'_\alpha(|\tilde{x}_i|+\Delta))_{i\in\mathrm{supp}(\tilde{x})}$ are collinear.*

As for the theorem 3, the assumptions on f_α given in theorem 4 hold for the functions $f_\alpha : x\in\mathbb{R}_+\mapsto x/(\alpha+x)$ and $f_\alpha : x\in\mathbb{R}_+\mapsto \arctan(x/\alpha)$. The points for which the properties i), ii) and iii) hold are kind of "critical points" of the problem $\mathcal{P}^\epsilon_{f_\alpha}$. The properties i), ii), iii) described in the previous theorem are verified at all points x^α of $S^\epsilon_{f_\alpha}$.

Actually, a proof similar to the proof of the lemma 9 shows that x^α is on the boundary of the constraint $\|y-Dx\|^2\leq\epsilon$. Consequently, the property i) holds for x^α.

By the lemma 1, the family $(d_i)_{i\in\mathrm{supp}(x^\alpha)}$ is linearly independent thus property ii) holds.

Finally, because x^α is a solution of the problem $\mathcal{P}^\epsilon_{f_\alpha}$, $(x^\alpha)_{i\in\mathrm{supp}(x^\alpha)}$ is also a solution of the problem

$$\text{argmin} \sum_{i\in\mathrm{supp}(x^\alpha)} f_\alpha(|x_i|) \text{ subject to } \|y-\tilde{D}x\|^2\leq\epsilon \text{ where } \tilde{D} \text{ is the matrix with columns } (d_i)_{i\in\mathrm{supp}(x^\alpha)}. \quad (4)$$

Consequently $(x^\alpha)_{\mathrm{supp}(x^\alpha)}$ is a stationary point of a Lagrangian function ([23] page 71, [2] page 243) implying thus the property iii) to hold with $\Delta=0$. The previous remark and the theorem 3 have a nice geometric interpretation illustrated on figure 2 for $p=3$ and $n=2$.

Because for each element x^α in $S^\epsilon_{f_\alpha}$, the property iii) holds with $\Delta=0$, this value of Δ could appear as the ideal value. It is not the case. Indeed, if we define the set L^α by

$$L^\alpha := \underset{x\in S^\epsilon_0}{\text{argmin}} \sum_{i=1}^{p} f_\alpha(|x_i|+\Delta), \quad (5)$$

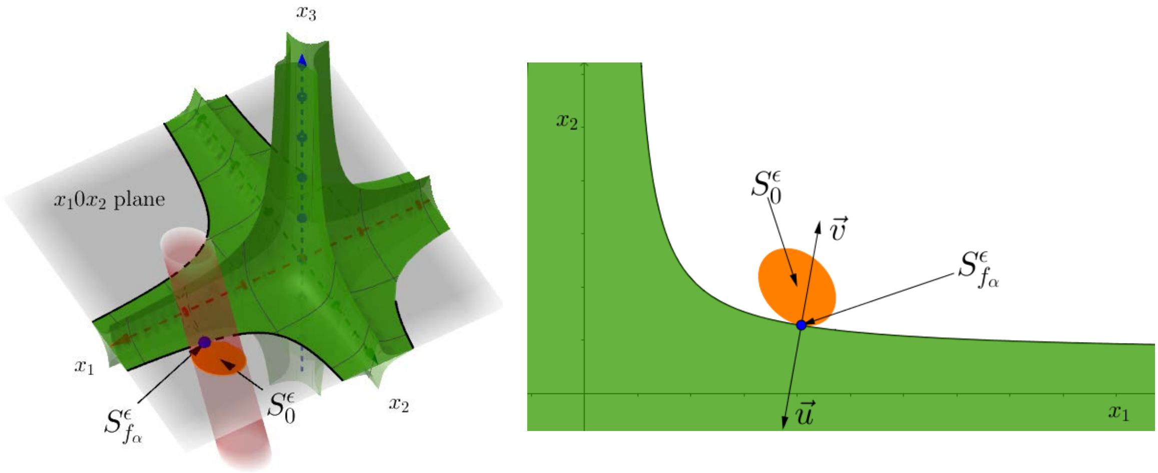

Figure 2: In the left panel the set of constraints $\|y - Dx\|^2 \leq \epsilon$ (in orange) and the "ball" $\sum_{i=1}^{3} f_\alpha(|x_i|) \leq R$ (in green) are represented. The radius R is the smallest positive number for which the cylinder $\|y - Dx\|^2 \leq \epsilon$ and the "ball" $\sum_{i=1}^{3} f_\alpha(|x_i|) \leq R$ share at least one common point. The set S_0^ϵ is a union of three ellipsoids which are the intersection of the cylinder $\|y - Dx\|^2 \leq \epsilon$ with the planes $x_1 0 x_2, x_1 0 x_3$ and $x_2 0 x_3$. To keep this illustration understandable, we only plot the intersection of the cylinder $\|y - Dx\|^2 \leq \epsilon$ and the plane $x_1 0 x_2$. The set $S_{f_\alpha}^\epsilon = \{x^\alpha\}$, represented as a blue point in the left figure, is a singleton of S_0^ϵ. This illustrates theorem 3 showing that whatever $\eta > 0$ $S_{f_\alpha}^\epsilon \subset G_\eta$. In the right panel, we focus on the intersection of the cylinder $\|y - Dx\|^2 \leq \epsilon$ and the intersection of the "ball" $\sum_{i=1}^{3} f_\alpha(|x_i|) \leq R$ with the plane $x_1 0 x_2$. The vectors $\vec{u} = \left(-d_i^T(y - Dx^\alpha)\right)_{1 \leq i \leq 2} = \left(\frac{\partial \|y-Dx\|^2}{\partial x_i}(x^\alpha)\right)_{1 \leq i \leq 2}$ and $\vec{v} = (\text{sign}(x_i) f'_\alpha(|x_i^\alpha|))_{1 \leq i \leq 2} = \left(\frac{\partial \sum_{i=1}^{3} f_\alpha(|x_i|)}{\partial x_i}(x^\alpha)\right)_{1 \leq i \leq 2}$ represent respectively the normalized normal vectors to the ellipsoid and the "ball". Note that the solution x^α of the problem (4) is i) on the boundary of the cylinder ii) completely included in the plane $(x_1 0 x_2)$, and iii) that at this point, the normal vectors $\vec{u}$ and $\vec{v}$ are collinear.

for an arbitrary $\Delta > 0$, the proposition 2 shows that L^α is a set of "critical points" such that $L^\alpha \subset S_0^\epsilon$. Consequently, whatever Δ, when $x^{(0)}$ is well chosen, one can expect that for k large enough, $x^{(k)}$ is close to the set L^α.

The proposition 2 shows that every element of L^α satisfies the property i), ii) and iii).

Proposition 2 *Let x^α be an arbitrary element of L^α. Then, the three following properties hold for x^α.*

i) The vector x^α is on the boundary of the constraint thus, $\|y - Dx^\alpha\|^2 = \epsilon$.

ii) The family $(d_i)_{i \in \text{supp}(x^\alpha)}$ is linearly independent.

iii) The vectors $(d_i^T(y - Dx^\alpha))_{i \in \text{supp}(x^\alpha)}$ and $(f'_\alpha(|x_i^\alpha| + \Delta))_{i \in \text{supp}(x^\alpha)}$ are collinear.

4 Numerical experiments

In the previous section, we developed a new method able to recover at least one solution of $\mathcal{P}_0$ or $\mathcal{P}_0^\epsilon$. Currently, the basis pursuit $\mathcal{P}_1$ is the reference method to recover a solution of $\mathcal{P}_0$. An alternative to the basis pursuit is the reweighted l^1 minimization [9]. In this section, we compare our method with both the basis pursuit and the reweighted l^1 minimization. For this numerical study, we use the same simulation framework as [9]. The family $\mathcal{D} = \{d_1, \ldots, d_p\}$ owns $p = 256$ vectors of $\mathbb{R}^n$ with $n = 100$. Whatever $i \in [\![1, 256]\!]$, the vector d_i is random vector $d_i := X_i/\|X_i\|$ with X_i i.i.d $\mathcal{N}(0, Id_{100})$. Consequently, the vectors $d_1, \ldots, d_p$ are independent and uniformly distributed on the $\mathbb{R}^n$ sphere. The vector $y \in \mathbb{R}^{100}$ that appears in the constraint $y = Dx$ is such that $y = D\tilde{x}$. For a given $s \in [\![1, n-1]\!]$, we choose $\tilde{x}$ as a random vector constructed as follows. Let $Z_1, \ldots, Z_s$ be i.i.d random variables $\mathcal{N}(0,1)$ distributed, we set $\forall i \notin [\![1, s]\!], \tilde{x}_i = 0$ and $\forall i \in [\![1, s]\!], \tilde{x}_i := Z_{(i)}$, where $Z_{(1)}, \ldots, Z_{(s)}$ are ordered variables such that $|Z_{(1)}| \geq \cdots \geq |Z_{(s)}|$. Because, by construction, almost surely the unique representation property holds for D (*i.e.* with a probability 1, $\text{spark}(D) = n+1$), when $s < (n+1)/2$ $\tilde{x}$ is almost surely the unique sparsest representation of y in D [35]. When $s \in [\![(n+1)/2, n-1]\!]$, one can show that $\tilde{x}$ is still the unique sparsest representation of y in D. The proposed MM method aims to find the sparsest representation of y in D which correspond to $\tilde{x}$.

In this section, we propose to slightly modify as follows the MM method given in (2).

$$\text{Let } a := \operatorname{argmin} \sum_{1 \leqslant i \leqslant p} f'_\alpha(|x_i^{(k)}| + \Delta)|x_i| \text{ subject to } Dx = y \text{ and set } \begin{cases} x^{(k+1)} = a \text{ if } \|a\|_0 \leq \|x^{(k)}\|_0 \\ x^{(k+1)} = x^{(k)} \text{ otherwise} \end{cases} . \quad (6)$$

As for the sequence given in (2), when k is large enough, the sequence (6) is stationary onto a point l. As defined in (6) the sequence $(\|x^k\|_0)_{k \in \mathbb{N}}$ is decreasing, consequently, $\|l\|_0 \leq \|x^{(0)}\|_0$. In particular when the initial point is the solution of $\mathcal{P}_1$, denoted hereafter x^{bp}, the modified MM method allows to catch a representation l better than x^{bp} in the sense that $\|l\|_0 \leq \|x^{\text{bp}}\|_0$. Whereas by taking $x^{(0)} = x^{\text{bp}}$ the performances of the modified MM method to solve $\mathcal{P}_0$ are better than the performances of the basis pursuit, x^{bp} is not the better initial point. The following section provides a smart initial point $x^{(0)}$.

4.1 Choice of the initial point $x^{(0)}$

Because the MM algorithm converges to a local minimum of $\mathcal{P}_0$, the choice of its initial point is critical. Candès et al [9] took the solution of problem $\mathcal{P}_1$ as the initial point for the iterative sequence (2). Another way to choose this initial point is based on the following two remarks.

1) Intuitively, the largest components of $\tilde{x}$ are more easily recovered than the smallest one. This intuition is confirmed by the right panel of the figure 3 which illustrates that x^{bp} catch easily the largest components of $\tilde{x}$.

2) When $\mathcal{A}$ is a known set that owns the largest components of $\tilde{x}$, the expression $\sum_{i \notin \mathcal{A}} |\tilde{x}_i|$ becomes small. As a consequence, substituting in $\mathcal{P}_1$ the function $\sum_{i=1}^p |x_i|$ by $\sum_{i \notin \mathcal{A}} |\tilde{x}_i|$ should provide a solution closer to $\tilde{x}$ than x^{bp}. So, to insure the uniqueness of the solution, instead of $\sum_{i \notin \mathcal{A}} |x_i|$ we could minimize the

expression $\omega \sum_{i\in\mathcal{A}} |x_i| + \sum_{i\notin\mathcal{A}} |x_i|$, with ω very small. This leads to the problem

$$\text{argmin } \omega \sum_{i\in\mathcal{A}} |x_i| + \sum_{i\notin\mathcal{A}} |x_i| \text{ subject to } Dx = y. \qquad (\mathcal{P}_{\mathcal{A}})$$

provides a closer solution of $\tilde{x}$ than the problem $\mathcal{P}_1$.

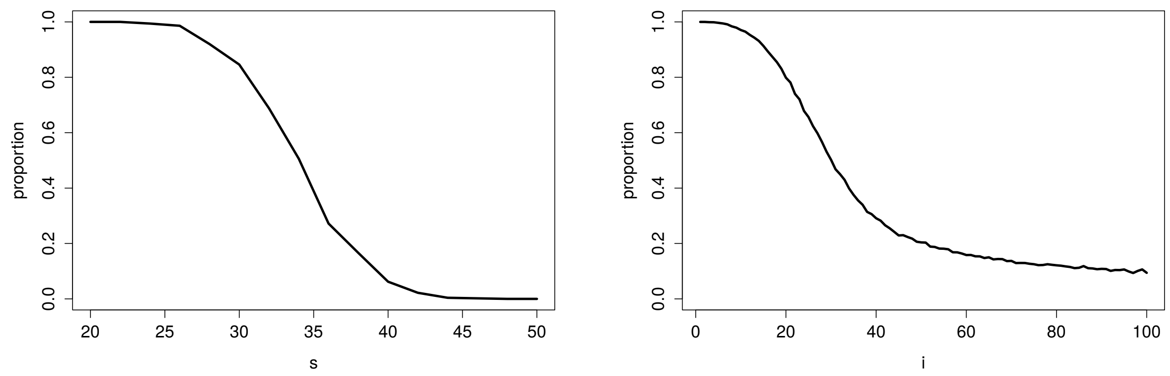

Figure 3: In this figure, $\tilde{x}$ is a random vector such that $\text{supp}(\tilde{x}) = [\![1, s]\!]$, with $s \in \{20, 22, \ldots, 50\}$ and $|\tilde{x}_1| \geq \cdots \geq |\tilde{x}_s|$. For every $s \in \{20, 22, \ldots, 50\}$, a sample of 500 families $\mathcal{D} = \{d_1, \ldots, d_{256}\}$ and 500 observations of the random vectors $\tilde{x}$ have been simulated. For each family and observation of $\tilde{x}$, we compute the solution x^{bp} of the basis pursuit problem $\mathcal{P}_1$. On the left panel, we have the representation of the proportion of times when $x^{\text{bp}} = \tilde{x}$ as a function of s. One notices that when $s \geq 45$, the event $x^{\text{bp}} = \tilde{x}$ is never observed. In the right panel, we set $s = 50$ and r is a permutation of $[\![1, 100]\!]$ such that $|x^{\text{bp}}_{r(1)}| \geq \cdots \geq |x^{\text{bp}}_{r(100)}|$ (by lemma 3, $\text{Card}(\text{supp}(x^{\text{bp}})) \leq 100$). For each $i \in [\![1, 100]\!]$ in the x-axis, the y-axis represents the proportion of times for which $r(i) \in \text{supp}(\tilde{x})$. Note that largest components of x^{bp} are elements of $\text{supp}(\tilde{x})$.

The figure 4 gives an algorithm which describes how to choose $x^{(0)}$. The input of the algorithm is x^{bp}. Ideally, when $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots \subset \text{supp}(\tilde{x})$, the solutions $x^{\text{init},(1)}, x^{\text{init},(2)} \ldots$ of the problems $\mathcal{P}_{\mathcal{A}_1}, \mathcal{P}_{\mathcal{A}_2}, \ldots$ should be increasingly closed to $\tilde{x}$. As already mentioned, the sparsest representation of y in D has a l^0 "norm" smaller than n. Consequently, the previous inclusion can not hold after the n^{th} iteration. So we stop the algorithm no later than the n^{th} iteration. When at the j^{th} iteration $\text{Card}(\text{supp}(x^{\text{init},(j)}) \setminus \mathcal{A}_j) = 0$, it is not possible to find an element i_j to construct the set $\mathcal{A}_{j+1}$ and the algorithm stops.

4.2 Comparisons

The simulations were performed for each $s \in \{24, 26, \ldots, 72\}$ using 500 random vectors $\tilde{x}$ such that $\text{supp}(\tilde{x}) = [\![1, s]\!]$, and 500 families $\mathcal{D} = \{d_1, \ldots, d_{256}\}$. These random vectors were ordered so that $|\tilde{x}_1| \geq \cdots \geq |\tilde{x}_s|$. For each family and each $\tilde{x}$, we compute the basis pursuit solution (x^{bp}) of $\mathcal{P}_1$, the reweighted l^1 minimization solution [9] and the solution given by our method as defined by (6). The reweighted l^1 solution is the limit of

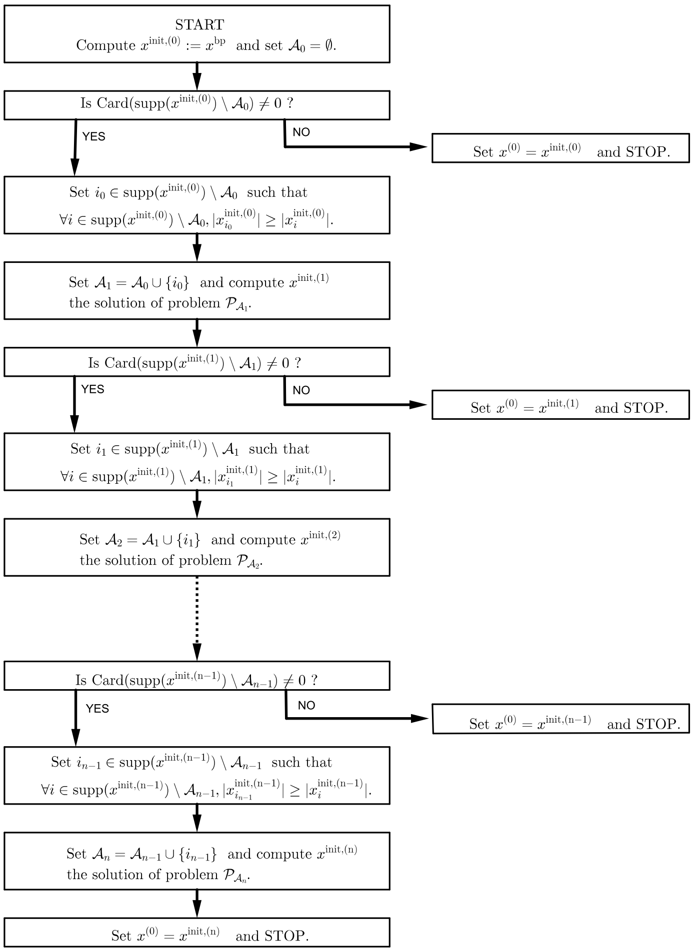

Figure 4: In this figure, we give the different steps of the algorithm to obtain the initial point $x^{(0)}$.

the sequence $(x^{\text{l1},(\text{k})})_{k\in\mathbb{N}}$ defined by $x^{\text{l1},(0)} = x^{\text{bp}}$ and

$$x^{\text{l1},(\text{k}+1)} := \text{argmin} \sum_{i=1}^{p} \frac{1}{|x_i^{\text{l1},(\text{k})}| + \delta} |x_i| \text{ subject to } Dx = y, \text{ with } y = D\tilde{x}.$$

As in [9] we set $\delta = 0.1$. The number of iterations was set to $k_0 = 8$ for both the reweighted l^1 minimization method and our method. We choose $f_\alpha(x) = x^\alpha$ with $\alpha = 0.01$ and the initial point of (6) was computed using the algorithm described previously. After 8 iterations, we keep the sparsest solution among the one obtained with $\Delta \in \{0.01, 0.1, 0.5, 1, 2, 4\}$.

The figure 5 shows the performances of the basis pursuit, the reweighted l^1 minimization and our method.

Numerical experiments given in the figure 5 show that when $\|\tilde{x}\|_0 \leq 22$, $\tilde{x}$ is always recovered by all these three methods. No method recovered $\tilde{x}$ when $\|\tilde{x}\|_0 \geq 68$. When $22 \leq \|\tilde{x}\|_0 \leq 68$, the proportion of times for which our method recovers $\tilde{x}$ is greater than the proportion given by the two other methods. These numerical experiments illustrate that the performances of our method are better than those of the basis pursuit and the reweighted l^1 minimization.

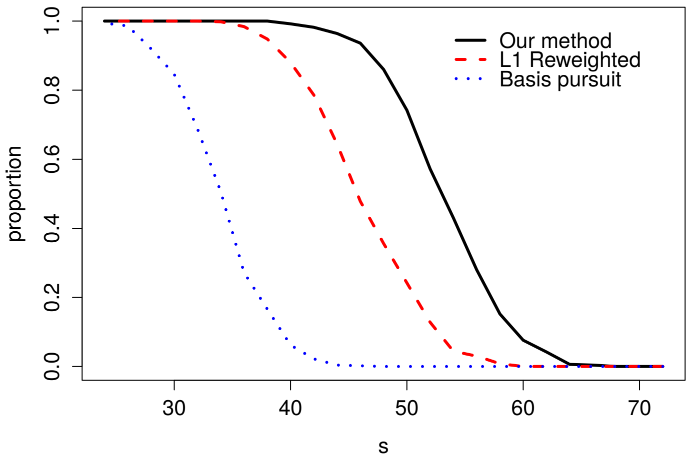

Figure 5: The performances of the three competing methods are represented by the proportions of realisations of the events $x^{\mathrm{bp}} = \tilde{x}$, $x^{\mathrm{l1},(8)} = \tilde{x}$ and $x^{(8)} = \tilde{x}$ as a function of the number of non null components of $\tilde{x}$ denoted s. One notices that the graph of the reweighted l1 minimization method is almost the same as those given in [9].

5 Conclusion

In this article, we studied the problems $\mathcal{P}_{f_\alpha}$ and $\mathcal{P}^{\epsilon}_{f_\alpha}$ which recover respectively one of the sparsest representations or one of the sparsest approximations of a high-dimensional linear system. Theoretical results are proved and a MM method is then used to solve these problems. Numerical experiments highligh the performances of our method compared to the basis pursuit and the reweighted l^1 minimization ones. In this study, the vector y is not corrupted by any noise. When y is a random vector, [25] provides an estimation of the representation of its expectation which has the smallest l^1 norm. In a future work, this work could be extended to estimate the sparsest representation of the expectation of y.

6 Appendix

6.1 Proof of the theorem 1

By construction, the function to be minimized in the problem $\mathcal{P}_{f_\alpha}$ converges pointwise to the l^0 "norm" when α goes to 0. As the l^0 norm is not continuous, this convergence can not be uniform onto $\mathbb{R}^p$. However, a straightforward consequence of the lemma 1 is that the number of possible solutions of the problem $\mathcal{P}_{f_\alpha}$ is finite and the convergence of $\sum_{i=1}^{p} f_\alpha(|x_i|)$ to $\|x\|_0$ is therefore uniform onto this finite set. The proof of theorem 1 is based on this uniform convergence.

Lemma 1 *Let f_α be a function defined on $\mathbb{R}_+$ strictly increasing and strictly concave such that*

$$\forall x \in \mathbb{R}_+, \lim_{\alpha \to 0} f_\alpha(x) = \mathbb{1}_{x \neq 0}.$$

Denote x^α a solution of the problem $\mathcal{P}_{f_\alpha}$ (resp. $\mathcal{P}^\epsilon_{f_\alpha}$) then the family $(d_i)_{i \in \mathrm{supp}(x^\alpha)}$ is linearly independent.

Proof : Let us assume that the family $(d_i)_{i \in \mathrm{supp}(x^\alpha)}$ is not linearly independent. There exist coefficients $(\gamma_i)_{i \in \mathrm{supp}(x^\alpha)}$ not simultaneously null such that

$$\sum_{i \in \mathrm{supp}(x^\alpha)} \gamma_i d_i = \vec{0}.$$

To provide a contradiction, we are going to show that$\sum_{i=1}^p f_\alpha(|x_i^\alpha|)$ is no longer minimal. That is, there exists an admissible point z so that $\sum_{i=1}^p f_\alpha(|z_i|) < \sum_{i=1}^p f_\alpha(|x_i^\alpha|)$. Let us define $\{i_1, \ldots, i_s\} := \{i \in \mathrm{supp}(x^\alpha) \mid \gamma_i \neq 0\}$, the set of non-null components of γ. We are looking for z among the admissible points $x(t)$ defined by

$$\forall t \in \mathbb{R}, x_i(t) = x_i^\alpha + t\gamma_i \text{ if } i \in \{i_1, \ldots, i_s\} \text{ and } x_i(t) = x_i^\alpha \text{ otherwise.}$$

For all $i \in \{i_1, \ldots, i_s\}$, let us denote $t_i = -x_i^\alpha / \gamma_i$. Without loss of generality, we assume that $t_{i_1} \leqslant \ldots \leqslant t_{i_s}$. The function $t \in \mathbb{R} \mapsto f_\alpha(|x_i(t)|)$ is strictly decreasing and strictly concave on $(-\infty, t_i]$ and strictly increasing and strictly concave on $[t_i, +\infty)$ when $i \in \{i_1, \ldots, i_s\}$.

Assume that $0 \notin [t_{i_1}, t_{i_s}]$; because each function $t \in \mathbb{R} \mapsto f_\alpha(|x_i(t)|)$ with $i \in \{i_1, \ldots, i_s\}$ is strictly decreasing on $(-\infty, t_i]$ (resp. strictly increasing on $[t_i, +\infty)$), one deduces that $t \in \mathbb{R} \mapsto \sum_{i=1}^p f_\alpha(|x_i(t)|)$ is strictly decreasing on $(-\infty, t_{i_1}]$ (resp. strictly increasing on $[t_{i_s}, +\infty)$). These monotony results imply that

$$\sum_{i=1}^p f_\alpha(|x_i(0)|) = \sum_{i=1}^p f_\alpha(|x_i^\alpha|) > \min\left\{\sum_{i=1}^p f_\alpha(|x(t_{i_1})|), \sum_{i=1}^p f_\alpha(|x(t_{i_s})|)\right\},$$

which provides a contradiction for the minimality of $\sum_{i=1}^p f_\alpha(|x_i^\alpha|)$.

Assume that $0 \in [t_{i_1}, t_{i_s}]$ then, there exists i_k such that $0 \in (t_{i_k}, t_{i_{k+1}})$ (note that t_{i_k} and $t_{i_{k+1}}$ are not null). Because each function $t \in \mathbb{R} \mapsto f_\alpha(|x_i(t)|)$ with $i \in \{i_1, \ldots, i_s\}$ is strictly concave on $[t_{i_k}, t_{i_{k+1}}]$, one deduces that $t \in \mathbb{R} \mapsto \sum_{i=1}^p f_\alpha(|x(t)|)$ is also strictly concave on $[t_{i_k}, t_{i_{k+1}}]$. Consequently, the restriction of the function $t \in \mathbb{R} \mapsto \sum_{i=1}^p f_\alpha(|x(t)|)$ to the set $[t_{i_k}, t_{i_{k+1}}]$ reaches its minimum at t_{i_k} or $t_{i_{k+1}}$ and nowhere else. This concavity result implies that

$$\sum_{i=1}^p f_\alpha(|x_i(0)|) = \sum_{i=1}^p f_\alpha(|x_i^\alpha|) > \min\left\{\sum_{i=1}^p f_\alpha(|x_i(t_{i_k})|), \sum_{i=1}^p f_\alpha(|x_i(t_{i_{k+1}})|)\right\},$$

which provides a contradiction for the minimality of $\sum_{i=1}^{p} f_\alpha(|x_i^\alpha|)$. □

We now consider the set $\mathcal{E}$ of subsets $I \subset [\![1,p]\!]$ such that

- The family $(d_i)_{i\in I}$ is linearly independent.
- $y \in Vect(d_i)_{i\in I}$.

Given a subset $I \in \mathcal{E}$, let x_I be the unique vector such that $\mathrm{supp}(x_I) = I$ and $Dx_I = y$. Let us introduce $S := \{x_I, I \in \mathcal{E}\}$. As $\mathcal{E}$ is finite, this set of vectors is finite.

Whatever the function f_α satisfying the properties of the lemma 1, the lemma 1 shows that the family $(d_i)_{i\in\mathrm{supp}(x^\alpha)}$ is linearly independent. As x^α is admissible, $y \in Vect(d_i)_{i\in\mathrm{supp}(x^\alpha)}$. It follows that for all $x^\alpha \in S_{f_\alpha}$, $x^\alpha \in S$; that is $S_{f_\alpha} \subset S$. The next lemma shows that the solutions of the problem $\mathcal{P}_0$ are also included in S.

Lemma 2 *The set S_0 of solutions of $\mathcal{P}_0$ satisfies $S_0 \subset S$.*

Proof : Let x^* be a solution of $\mathcal{P}_0$, we have $Dx^* = y$. To show that $x^* \in S$, it remains to prove that the family $(d_i)_{i\in\mathrm{supp}(x^*)}$ is linearly independent. Suppose that this family is not linearly independent then there exist coefficients $(\gamma_i)_{i\in\mathrm{supp}(x^*)}$ not simultaneously null such that

$$\sum_{i\in\mathrm{supp}(x^*)} \gamma_i d_i = \vec{0}.$$

To provide a contradiction for the minimality of $\|x^*\|_0$, we are going to prove that there exists an admissible point z such that $\|z\|_0 < \|x^*\|_0$. We are looking for z among admissible points $x(t)$ defined by

$$\forall t \in \mathbb{R}, x_i(t) = x_i^* + t\gamma_i \text{ if } i \in \mathrm{supp}(x^*) \text{ and } x_i(t) = x_i^* = 0 \text{ otherwise.}$$

By construction, we have $\forall t \in \mathbb{R}, \mathrm{supp}(x(t)) \subset \mathrm{supp}(x^*)$. To conclude this proof, we have to find $t_0 \in \mathbb{R}$ for which the inclusion is strict. Let $i_0 \in \mathrm{supp}(x^*)$ such that $\gamma_{i_0} \neq 0$ and define $t_0 = -x_{i_0}^*/\gamma_{i_0}$. The i_0^{th} component of $x(t_0)$ is null. Consequently, $\|x(t_0)\|_0 < \|x^*\|_0$ which provides a contradiction to the fact that x^* is a solution of $\mathcal{P}_0$. □

Proof of theorem 1: By the lemma 1 and 2, we have $S_{f_\alpha} \subset S$ and $S_0 \subset S$. If the elements of $S \setminus S_0$ are not solution of $\mathcal{P}_{f_\alpha}$, one deduces that $S_{f_\alpha} \subset S_0$. Let x and x^* be respectively an arbitrary element of $S \setminus S_0$ and of S_0. A straightforward consequence of the inequality $\sum_{i=1}^{p} f_\alpha(|x_i|) > \sum_{i=1}^{p} f_\alpha(|x_i^*|)$ is that x is not a solution of $\mathcal{P}_{f_\alpha}$. We are going prove that this inequality holds when α is small enough. We have that

$\forall x \in S \setminus S_0$,

$$\sum_{i=1}^{p} f_\alpha(|x_i|) - \sum_{i=1}^{p} f_\alpha(|x_i^*|) = \sum_{i=1}^{p} f_\alpha(|x_i|) - \|x\|_0 + \|x\|_0 - \|x^*\|_0 + \|x^*\|_0 - \sum_{i=1}^{p} f_\alpha(|x_i^*|).$$

Because x is not a solution of $\mathcal{P}_0$ contrarily to x^*, one has $\|x\|_0 - \|x^*\|_0 \geqslant 1$. Furthermore, the uniform convergence of $\sum_{i=1}^{p} f_\alpha(|x_i|)$ to $\|x\|_0$ onto the set S gives $\alpha_0 > 0$ such that

$$\forall \alpha \in (0, \alpha_0), \forall x \in S, \left| \sum_{i=1}^{p} f_\alpha(|x_i|) - \|x\|_0 \right| < 1/2.$$

Consequently, one obtains

$$\forall \alpha \in (0, \alpha_0), \forall x \in S \setminus S_0, \sum_{i=1}^{p} f_\alpha(|x_i|) > \sum_{i=1}^{p} f_\alpha(|x_i^*|).$$

Thus, as soon as $\alpha < \alpha_0$, the solution of $\mathcal{P}_{f_\alpha}$ satisfies $S_{f_\alpha} \subset S_0$ □

6.2 Proof of the theorem 2 and of the proposition 1

The main consequence of lemma 3, is that the iterative sequence $(x^{(k)})_{k\geq 1}$ provided by the MM method (2) satisfies $\forall k \geq 1, x^{(k)} \in S$. Because S is a finite set, this result is useful for the proof of the theorem 2.

Lemma 3 *Let us denote*

$$S_\omega := \operatorname{argmin} \sum_{i=1}^{p} w_i |x_i| \text{ subject to } y = Dx, \text{with } \forall i \in [\![1, p]\!], \omega_i > 0 \tag{7}$$

and

$$S_\omega^\epsilon := \operatorname{argmin} \sum_{i=1}^{p} w_i |x_i| \text{ subject to } \|y - Dx\|_2^2 \leqslant \epsilon, \text{with } \forall i \in [\![1, p]\!], \omega_i > 0. \tag{8}$$

Then, there exists an element $x^\omega \in S_\omega$ (resp. $x^\omega \in S_\omega^\epsilon$) such that the family $(d_i)_{i \in \mathrm{supp}(x^\omega)}$ is linearly independent.

Proof : When the set S_ω (resp. S_ω^ϵ) is not a singleton, we set x^ω an element of S_ω (resp. S_ω^ϵ) with a minimal l^0 norm. Assume that $(d_i)_{i \in \mathrm{supp}(x^\omega)}$ is not linearly independent. There exist coefficients $(\gamma_i)_{i \in \mathrm{supp}(x)}$ not simultaneously null such that $\sum_{i \in \mathrm{supp}(x^\omega)} \gamma_i d_i = \vec{0}$. Let us set $\mathcal{A}' := \{i \in \mathrm{supp}(x^\omega) \text{ such that } \gamma_i \neq 0\}$. One defines the admissible $x(t)$ of the problem (7) (resp. (8)) as follows

$$x_i(t) := \begin{cases} x_i^\omega + t\gamma_i \text{ if } i \in \mathcal{A}', \\ x_i^\omega \text{ otherwise.} \end{cases}$$

By definition, the point $x(t)$ satisfies $\text{supp}(x(t)) \subset \text{supp}(x^\omega)$. To provide a contradiction for the minimality of the l^0 "norm" of the solution x^ω, we could build an element $x(t_0) \in S_\omega$ (resp. S_ω^ϵ) with a strictly lower l^0 "norm".

Let f be the function $\forall t \in \mathbb{R}, f(t) := \sum_{i=1}^p w_i |x_i(t)|$. This function is equal to $f(t) = \sum_{i\in\mathcal{A}'} \omega_i |x_i + t\gamma_i| + \sum_{i\notin\mathcal{A}'} \omega_i |x_i|$. The minimum of f is reached on the set $\{-x_i/\gamma_i\}_{i\in\mathcal{A}'}$. If $t_0 := -x_{i_0}/\gamma_{i_0}$, with $i_0 \in \mathcal{A}'$, is a value for which the minimum of f is reached, one sees that $x_{i_0}(t_0) = 0$. This shows $\|x(t_0)\|_0 < \|x^\omega\|_0$ and $x(t_0)$ is an admissible point for which $\sum_{i=1}^p \omega_i |x_i(t_0)| \leq \sum_{i=1}^p \omega_i |x_i(0)| = \sum_{i=1}^p \omega_i |x_i^\omega|$. Consequently, $x(t_0)$ is point of S_ω (resp. S_ω^ϵ) with a strictly smaller l^0 "norm" than the one of x^ω which contradicts the minimality of $\|x^\omega\|_0$. □

Remind that for each $k \geq 1$, $x^{(k)}$ defined in (2) is the solution of a weighted basis pursuit problem. We have already noted that in practice weighted basis pursuit problem admits a unique solution. Consequently, by the lemma 3 the family $(d_i)_{i\in\text{supp}(x^{(k)})}$ is linearly independent and, on the other hand, $y = Dx^{(k)}$ which implies that $x^{(k)} \in S$.

Proof of theorem 2 : The MM method for the function $x \in \mathbb{R}^p \mapsto \sum_{1\leqslant i\leqslant p} f_\alpha(|x_i| + \Delta)$ provides the sequence $(x^{(k)})_{k\geq 0}$ defined in (2). In the following, we prove that the sequence $(u_k)_{k\in\mathbb{N}}$ with $u_k := \sum_{1\leqslant i\leqslant p} f_\alpha(|x_i^{(k)}| + \Delta)$ is stationary.

For $k \geq 1$, the vector $x^{(k)}$ is a solution of a weighted basis pursuit problem. Consequently, the lemma 3 insures that $x^{(k)} \in S$. Since S is a finite set, the sequence $(u_k)_{k\leq 1}$ can only take a finite number of values

$$\forall k \in \mathbb{N}^*, u_k \in \left\{ \sum_{1\leqslant i\leqslant p} f_\alpha(|x_i^I| + \Delta), I \in \mathcal{E} \right\}.$$

If we show that the sequence $(u_k)_{k\in\mathbb{N}}$ is decreasing that implies its stationary for a large enough k. We follow the proof given in [21, 23]. Remind that $x^{(k+1)}$ is defined as follow

$$x^{(k+1)} := \text{argmin} \sum_{1\leqslant i\leqslant p} f_\alpha(|x_i^{(k)}| + \Delta) + f'_\alpha(|x_i^{(k)}| + \Delta)(|x_i| - |x_i^{(k)}|).$$

Let us set $L_{x^{(k)}}(x) := \sum_{1\leqslant i\leqslant p} f_\alpha(|x_i^{(k)}| + \Delta) + f'_\alpha(|x_i^{(k)}| + \Delta)(|x_i| - |x_i^{(k)}|)$. The concavity of the function $x \in \mathbb{R} \mapsto f_\alpha(x + \Delta)$ on $\mathbb{R}_+$ implies that

$$\forall x \in \mathbb{R}^p, \sum_{1\leqslant i\leqslant p} f_\alpha(|x_i| + \Delta) \leqslant L_{x^{(k)}}(x).$$

Because, the minimum of $L_{x^{(k)}}(x)$ is reached at $x^{(k+1)}$, one obtains the following property

$$u_{k+1} = \sum_{1\leqslant i\leqslant p} f_\alpha(|x_i^{(k+1)}| + \Delta) \leqslant L_{x^{(k)}}(x^{(k+1)}) \leqslant L_{x^{(k)}}(x^{(k)}) = \sum_{1\leqslant i\leqslant p} f_\alpha(|x_i^{(k)}| + \Delta) = u_k.$$

Since the sequence $(u_k)_{k\in\mathbb{N}}$ is decreasing, there exists $k_0 \geq 0$ such that $(u_k)_{k\in\mathbb{N}}$ is stationary for $k \geq k_0$.

The strict concavity of the function $x \in \mathbb{R}_+ \mapsto f(x+\Delta)$ implies that

$$f_\alpha(|x_i^{(k_0+1)}|+\Delta) \leq f_\alpha(|x_i^{(k_0)}|+\Delta) + f'_\alpha(|x_i^{(k_0)}|+\Delta)(|x_i^{(k_0+1)}| - |x_i^{(k_0)}|),$$

with a strict inequality when $|x_i^{(k_0+1)}| \neq |x_i^{(k_0)}|$. Thus, if there exists $i_0 \in [\![1,p]\!]$ such that $|x_{i_0}^{(k_0+1)}| \neq |x_{i_0}^{(k_0)}|$, $u_{k_0+1} < L_{x^{(k_0)}}(x^{(k_0+1)}) \leq u_{k_0}$ which provides a contradiction for the stationary of the sequence $(u_k)_{k\in\mathbb{N}}$. Consequently, we have

$$\forall i \in [\![1,p]\!], |x_i^{(k_0+1)}| = |x_i^{(k_0)}|.$$

This equality gives that $\text{supp}(x^{(k_0)}) = \text{supp}(x^{(k_0+1)})$. Because $x^{(k_0)}$ and $x^{(k_0+1)}$ are admissible points,

$$\sum_{i\in\text{supp}(x^{(k_0)})} x_i^{(k_0)} d_i = \sum_{i\in\text{supp}(x^{(k_0)})} x_i^{(k_0+1)} d_i.$$

Finally, the lemma 3 implies that the family $(d_i)_{i\in\text{supp}(x^{(k_0)})}$ is linearly independent. One deduces that $x^{(k_0)} = x^{(k_0+1)}$. A straightforward consequence is that the sequence $(x^{(k)})_{k\in\mathbb{N}}$ is stationary when $k \geq k_0$. □

Proof of proposition 1 : Remind that l is the limit of the sequence $x^{(k)}$ given in (2). Let us defined $r := \min\{|l_i|, i \in \text{supp}(l)\}$. One can check that $\forall x \in B_\infty(l,r)$ we have $x_i \neq 0$ once $l_i \neq 0$. Consequently, $\text{supp}(l) \subset \text{supp}(x)$. Assume $\text{supp}(x) = \text{supp}(l)$. Since $Dx = Dl$, one deduces that

$$\sum_{i\in\text{supp}(l)} x_i d_i = \sum_{i\in\text{supp}(l)} l_i d_i.$$

Since the family $(d_i)_{i\in\text{supp}(l)}$ is linearly independent, one deduces that $x = l$. Consequently, $\forall x \in B_\infty(l,r)$ such that $x \neq l$, we have $\text{supp}(l) \subsetneq \text{supp}(x)$ thus, $\|l\|_0 < \|x\|_0$. □

6.3 Proof of the theorem 3

By the lemma 1, for any x^* in $S^\epsilon_{f_\alpha}$, the family $(d_i)_{i\in x^*}$ is linearly independent. Moreover, x^* is an admissible point, thus $\|y - Dx^*\|^2 \leq \epsilon$. Consequently, $x^* \in \bigcup_{I\in\mathcal{I}} E_I$, where

$$\mathcal{I} := \{I \subset [\![1,p]\!] \mid (d_i)_{i\in I} \text{ is linearly independent } \} \text{ and } E_I := \{x \in \mathbb{R}^p \mid \text{supp}(x) \subset I \text{ and } \|y - Dx\|^2 \leq \epsilon\}.$$

Let us denotes $E := \bigcup_{I\in\mathcal{I}} E_I$.

Lemma 4 *The set E is compact.*

Proof : Let us denote $\bar{x} \in \mathbb{R}^p$ with $\text{supp}(\bar{x}) \subset I$ such that $D\bar{x}$ is the orthogonal projection of y onto the space $\text{Vect}(d_i)_{i\in I}$. If $\|y - D\bar{x}\|^2 > \epsilon$ then the set E_I is empty. Otherwise,

$$E_I = \{x \in \mathbb{R}^p \mid \text{supp}(x) \subset I \text{ and } \|D(x - \bar{x})\|^2 \leq \epsilon'\}, \text{ with } \epsilon' = \epsilon - \|y - D\bar{x}\|^2.$$

Since $\text{supp}(x) \subset I$ and $\text{supp}(\bar{x}) \subset I$, one shows that

$$\|D(x - \bar{x})\|^2 = \|D_S(x_I - \bar{x}_I)\|^2,$$

with $x_I := (x_i)_{i\in I}$, $\bar{x}_I := (\bar{x}_i)_{i\in I}$ and D_I is matrix whose columns are $(d_i)_{i\in S}$. Because the family $(d_i)_{i\in I}$ is linearly independent, the Gram matrix $D_I^T D_I$ is invertible thus, $\|D_I(x_I - \bar{x}_I)\|^2 \leq \epsilon'$ is an ellipsoid of $\mathbb{R}^{\text{Card(I)}}$. Therefore, E_I is a compact. Consequently, the finite union of compact set $\bigcup_{I\in\mathcal{I}} E_I$ is a compact set. □

In the lemma 5 and the theorem 3, we denote $s_0 := \min \|x\|$ subject to $\|y - Dx\|^2 \leq \epsilon$.

Lemma 5 *For $\eta > 0$, let us denote G_η the open set $G_\eta = \bigcup_{x\in S_0^\epsilon} B(x, \eta)$. The function*

$$F_\alpha : x \in E \setminus G_\eta \mapsto \min\left\{s_0 + 1, \sum_{i=1}^{p} f_\alpha(|x_i|)\right\}$$

converges uniformly to the function $F : x \in E \setminus G_\eta \mapsto s_0 + 1$ when α converges to 0.

Proof : Let $(\alpha_n)_{n\in\mathbb{N}}$ be a decreasing sequence converging toward 0. Because $f_\alpha \geq f_{\alpha'}$ once $\alpha \leq \alpha'$, $(F_{\alpha_n})_{n\in\mathbb{N}}$ is a monotonic sequence of continuous functions. Furthermore, on the compact set $E\setminus G_\eta$, this sequence converges pointwise toward the continuous function $F : x \in E \setminus G_\eta \mapsto s_0 + 1$. Consequently, the Dini's theorem gives the uniform convergence of $(F_{\alpha_n})_{n\in\mathbb{N}}$. Therefore, for all $\delta > 0$, there exists n_0 such that

$$\forall n \geq n_0, \sup_{x\in E\setminus G_\eta} \{|F_{\alpha_n}(x) - s_0 - 1|\} \leq \delta.$$

Finally, if $\alpha \leq \alpha_{n_0}$, for all $x \in E \setminus G_\eta$ we have the following inequalities

$$-\delta \leq F_{\alpha_{n_0}}(x) - s_0 - 1 \leq F_\alpha(x) - s_0 - 1 \leq 0.$$

Consequently, one obtains

$$\sup_{x\in E\setminus G_\eta} \{|F_\alpha(x) - s_0 - 1|\} \leq \sup_{x\in E\setminus G_\eta} \left\{|F_{\alpha_{n_0}}(x) - s_0 - 1|\right\} \leq \delta,$$

which shows the uniform convergence. □

Proof of theorem 3 : Let x^* be an arbitrary element of S_0^ϵ, we are going to prove that for $\alpha > 0$ small enough,

$$\forall x \in E \setminus G_\eta, \sum_{i=1}^{p} f_\alpha(|x_i|) > \sum_{i=1}^{p} f_\alpha(|x_i^*|). \tag{9}$$

If the inequality (9) holds then $S^\epsilon_{f_\alpha} \subset G_\eta$. Actually, by definition, $S^\epsilon_{f_\alpha} \subset E$ and by the inequality (9), the elements of $E \setminus G_\eta$ are not solution of $\mathcal{P}^\epsilon_{f_\alpha}$. The convergence of $\sum_{i=1}^p f_\alpha(|x_i^*|)$ toward s_0 once α converges to 0 implies that

$$\exists \alpha_1 > 0 \text{ such that } \forall \alpha \leq \alpha_1, \sum_{i=1}^{p} f_\alpha(|x_i^*|) < s_0 + 1/2.$$

The uniform convergence given in the previous lemma 5 implies that

$$\exists \alpha_2, \forall \alpha \leq \alpha_2, \forall x \in E \setminus G_\eta \min\left\{ s_0 + 1, \sum_{i=1}^{p} f_\alpha(|x_i|) \right\} > s_0 + 1/2.$$

Finally, if we set $\alpha_0 = \min\{\alpha_1, \alpha_2\}$, we have

$$\forall \alpha \leq \alpha_0, \forall x \in E \setminus G_\eta, \min\left\{ s_0 + 1, \sum_{i=1}^{p} f_\alpha(|x_i|) \right\} - \sum_{i=1}^{p} f_\alpha(|x_i^*|) > 0,$$

which implies

$$\forall \alpha \leq \alpha_0, \forall x \in E \setminus G_\eta, \sum_{i=1}^{p} f_\alpha(|x_i|) > \sum_{i=1}^{p} f_\alpha(|x_i^*|).$$

□

6.4 Proof of the theorem 4 and of the proposition 2

Let $(x^{(\phi(k))})_{k\geq 0}$ be a subsequence of $x^{(k)}$ (defined in 3) that converges to $\tilde{x}$. The lemmas 6, 7 and 8 are used to prove that the sequence $(x^{(\phi(k)+1)})_{k\geq 0}$ has the same limit as $(x^{(\phi(k))})_{k\geq 0}$.

Lemma 6 *Let $f : \mathbb{R}_+ \to \mathbb{R}$ be an striclty increasing, strictly concave and two times differentiable function such that f' is convex then,*

$$\forall \eta > 0, \exists \epsilon > 0 \text{ such that } \forall a \in [0, a_0], \forall b \in \mathbb{R}_+, |a - b| > \eta \Rightarrow f(a) + f'(a)(b - a) - f(b) > \epsilon. \tag{10}$$

Proof : Let us defined the function $g_{a_0}(h)$ as follows

$$\forall h \geq 0, g_{a_0}(h) := f(a_0) + f'(a_0)h - f(a_0 + h).$$

We are going to prove that (10) holds when $\epsilon = g_{a_0}(\eta)$. In a first step, let us prove that $f(a)+f'(a)(b-a)-f(b) \geq g_{a_0}(|b - a|)$. We set $t = b - a$, the convexity of f' gives

$$\frac{\partial}{\partial a} \left(f(a) + f'(a)|t| - f(a + |t|) \right) = f'(a) + f''(a)|t| - f'(a + |t|) \leq 0.$$

The concavity of f gives

$$f(a)+f'(a)t-f(a+t)\geq f(a)+f'(a)|t|-f(a+|t|).$$

Indeed, when $t\geq 0$, the result is obvious otherwise, when $t<0$, we have $t=-|t|$, the previous inequality is a consequence of the next one

$$\frac{f(a)-f(a-|t|)}{|t|}\geq f'(a)\geq \frac{f(a+|t|)-f(a)}{|t|}$$

From these inequalities, one deduces that

$$f(a)+f'(a)t-f(a+t)\geq f(a)+f'(a)|t|-f(a+|t|)\geq f(a_0)+f'(a_0)|t|-f(a_0+|t|)=g_{a_0}(|b-a|).$$

The function f' is strictly decreasing (because f is strictly concave) consequently $\forall h>0, g'_{a_0}(h)=f'(a_0)-f'(a_0+h)>0$ thus, g is strictly increasing. Since $g_{a_0}(0)=0$, we have $\epsilon:=g_{a_0}(\eta)>0$. Finally, if $|b-a|>\eta$ we have

$$f(a)+f'(a)(b-a)-f(b)\geq g_{a_0}(|b-a|)>g_{a_0}(\eta)=\epsilon.$$

□

In the following, we denote $|x|:=(|x_i|)_{1\leq i\leq p}$ with $x\in\mathbb{R}^p$.

Lemma 7 *The sequence $(x^{(k)})_{k\in\mathbb{N}}$ described in (3) satisfies*

$$\lim_{k\to+\infty} d_\infty(|x^{(k+1)}|,|x^{(k)}|)=0$$

Proof : Let us define the sequence $(u_k)_{k\in\mathbb{N}}$ with $u_k:=\sum_{1\leq i\leq p} f_\alpha(|x_i^{(k)}|+\Delta)$. The convergence of this sequence is given in the proof of the theorem 2.

Assume that $d_\infty(|x^{(k+1)}|,|x^{(k)}|)$ does not converge to 0, we have

$$\exists\eta>0,\forall K\geq 0,\exists k_0\geq K \text{ such that } d_\infty(|x^{(k_0+1)}|,|x^{(k_0)}|)\geq\eta.$$

If $d_\infty(|x^{(k_0+1)}|,|x^{(k_0)}|)\geq\eta$ then, there exists $i_0\in[\![1,p]\!]$ such that $\left||x_{i_0}^{(k_0+1)}|-|x_{i_0}^{(k_0)}|\right|\geq\eta$. Because the sequence $(x^{(k)})_{k\in\mathbb{N}}$ is bounded (proof 1 of the theorem 4), there exists $a_0\geq 0$ such that $\forall k\in\mathbb{N},\|x^{(k)}\|_\infty\leq a_0$. By the lemma 6 we have

$$\exists\epsilon>0 \text{ such that } f_\alpha(|x_{i_0}^{(k_0)}|+\Delta)+f'_\alpha(|x_{i_0}^{(k_0)}|+\Delta)(|x_{i_0}^{(k_0+1)}|-|x_{i_0}^{(k_0)}|)-f_\alpha(|x_{i_0}^{(k_0+1)}|+\Delta)\geq\epsilon.$$

Furthermore the concavity of f_α implies that

$$\forall i \neq i_0, f_\alpha(|x_i^{(k_0)}| + \Delta) + f'_\alpha(|x_i^{(k_0)}| + \Delta)(|x_i^{(k_0+1)}| - |x_i^{(k_0)}|) - f_\alpha(|x_i^{(k_0+1)}| + \Delta) \geq 0.$$

These two inequalities imply that

$$u_{k_0+1} + \epsilon = \sum_{i=1}^{p} f_\alpha(|x_i^{(k_0+1)}| + \Delta) + \epsilon \leq \sum_{i=1}^{p} f_\alpha(|x_i^{(k_0)}| + \Delta) + f'_\alpha(|x_i^{(k_0)}| + \Delta)(|x_i^{(k_0+1)}| - |x_i^{(k_0)}|)$$

Furthermore, by definition of $x^{(k_0+1)}$, we have

$$\sum_{i=1}^{p} f_\alpha(|x_i^{(k_0)}| + \Delta) + f'_\alpha(|x_i^{(k_0)}| + \Delta)(|x_i^{(k_0+1)}| - |x_i^{(k_0)}|) \leq \sum_{i=1}^{p} f_\alpha(|x_i^{(k_0)}| + \Delta) = u_{k_0}.$$

The previous inequality implies that

$$\forall K, \exists k_0 \geq K \text{ such that } |u_{k_0+1} - u_{k_0}| \geq \epsilon.$$

The last inequality provides a contradiction for the convergence of the sequence $(u_k)_{k\in\mathbb{N}}$. □

Lemma 8 *Let $x^{(\phi(k))}$ be a subsequence of $(x^{(k)})_{k\in\mathbb{N}}$ that converges toward $\tilde{x}$ then, the sequence $(x^{(\phi(k)+1)})_{k\in\mathbb{N}}$ converges toward $\tilde{x}$.*

Proof : The proof 1) in the theorem 4 shows that the sequence $(x^{(k)})_{k\in\mathbb{N}}$ is bounded. Consequently, $(x^{(\phi(k)+1)})_{k\in\mathbb{N}}$ is bounded too. To prove that the bounded sequence $(x^{(\phi(k)+1)})_{k\in\mathbb{N}}$ converges to $\tilde{x}$, it is sufficient to show that $\tilde{x}$ is the only limit point of this sequence. Let $(x^{(\phi(\psi(k))+1)})_{k\in\mathbb{N}}$ be a converging subsequence such that

$$\lim_{k\to+\infty} x^{(\phi(\psi(k))+1)} = \tilde{x}^1, \text{ with } \tilde{x}_1 \neq \tilde{x}.$$

By the lemma 7, we have $\lim_{k\to+\infty} d_\infty(|x^{(\phi(\psi(k))+1)}|, |x^{(\phi(\psi(k)))}|) = 0$. Since $\lim_{k\to+\infty} x^{(\phi(\psi(k)))} = \tilde{x}$, one deduces that $|\tilde{x}| = |\tilde{x}^1|$. Let us define $\tilde{x}^2$ as $\tilde{x}^2 := (\tilde{x}^1 + \tilde{x})/2$. Because

$$x^{(\phi(\psi(k))+1)} := \text{argmin} \sum_{1\leqslant i\leqslant p} f'_\alpha(|x_i^{(\phi(\psi(k)))}| + \Delta)|x_i| \text{ subject to } \|y - Dx\|^2 \leq \epsilon,$$

we have

$$\sum_{i=1}^{p} f'_\alpha(|x_i^{(\phi(\psi(k)))}| + \Delta)|x_i^{(\phi(\psi(k))+1)}| \leq \sum_{i=1}^{p} f'_\alpha(|x_i^{(\phi(\psi(k)))}| + \Delta)(|\tilde{x}_i^2|).$$

Taking the limit in the previous expression, one obtains

$$\sum_{1\leq i\leq p} f'_\alpha(|\tilde{x}_i| + \Delta)|\tilde{x}_i^1| \leq \sum_{1\leq i\leq p} f'_\alpha(|\tilde{x}_i| + \Delta)|\tilde{x}_i^2|. \tag{11}$$

On the other hand, $\text{supp}(\tilde{x}^2) = \{i \in \text{supp}(\tilde{x}^1) \mid \tilde{x}_i = \tilde{x}_i^1\}$, which implies that $\text{supp}(\tilde{x}^2) \subsetneq \text{supp}(\tilde{x}^1)$ and $\forall i \in \text{supp}(\tilde{x}^2), \tilde{x}_i^2 = \tilde{x}_i^1$. Consequently, we have

$$\sum_{1\le i\le p} f'_\alpha(|\tilde{x}_i|+\Delta)|\tilde{x}_i^1| > \sum_{i\in\text{supp}(\tilde{x}^2)} f'_\alpha(|\tilde{x}_i|+\Delta)|\tilde{x}_i^1| = \sum_{i\in\text{supp}(\tilde{x}^2)} f'_\alpha(|\tilde{x}_i|+\Delta)|\tilde{x}_i^2| = \sum_{1\le i\le p} f'_\alpha(|\tilde{x}_i|+\Delta)|\tilde{x}_i^2|. \quad (12)$$

The inequality (12) provides a contradiction with the inequality (11). Therefore, the only limit point of the bounded sequence $(x^{(\phi(k)+1)})_{k\in\mathbb{N}}$ is $\tilde{x}$. □

Lemma 9 *Let x^ω be a solution of the weighted lasso problem*

$$\text{argmin} \sum_{i=1}^{p} w_i|x_i| \text{ subject to } \|y-Dx\|^2 \leqslant \epsilon, \text{with } \forall i \in [\![1,p]\!], \omega_i > 0. \quad (13)$$

Furthermore, let us assume that $\|y\|^2 > \epsilon$ then, $\|y - Dx^\omega\|^2 = \epsilon$.

Proof : Let us assume that $\|y - Dx^\omega\|^2 < \epsilon$. Consider the points $x(t)$ defined by

$$\forall i \in [\![1,p]\!], x_i(t) = sign(x_i^\omega)(|x_i^\omega| - t)_+, \text{ where } (a)_+ = \max\{a, 0\}.$$

One can check that $\|x(t) - x^\omega\|_\infty \le t$. Because the set $\{x \in \mathbb{R}^p \mid \|y - Dx\|^2 < \epsilon\}$ is an open set, there exists $t_0 > 0$ small enough such that $\|y - Dx(t_0)\|^2 < \epsilon$. Finally, we have

$$\forall i \notin \text{supp}(x^\omega), |x_i(t_0)| = |x_i^\omega| = 0 \text{ and } \forall i \in \text{supp}(x^\omega), |x_i(t_0)| < |x_i^\omega|.$$

Because $\vec{0}$ is not an admissible point, one has $x^\omega \neq \vec{0}$. Consequently, we have the following inequality.

$$\sum_{i=1}^{p} w_i|x_i(t_0)| < \sum_{i=1}^{p} w_i|x_i^\omega|.$$

Such a result provides a contradiction for the minimality of $\sum_{i=1}^p \omega_i|x_i^\omega|$. □

Proof of theorem 4 :

1) For any $k \ge 1$, $x^{(k)}$ is the solution of a weighted lasso. By lemma 3, the family $(d_i)_{i\in\text{supp}(x^{(k)})}$ is linearly independent. Consequently, $\forall k \ge 1, x^{(k)} \in E$, where E is the set given in the lemma 4. Because E is a compact set of $\mathbb{R}^p$, one deduces that $(x^{(k)})_{k\in\mathbb{N}}$ is bounded.

2-i) Because $\lim_{k\to+\infty} x^{(\phi(k))} = \tilde{x}$, there exists k_0 such that

$$\forall k \ge k_0, \text{supp}(\tilde{x}) \subset \text{supp}(x^{(\phi(k))}).$$

Since by lemma 3 $(d_i)_{i\in\text{supp}(x^{(k_0)})}$ is linearly independent, one deduces that $(d_i)_{i\in\text{supp}(\tilde{x})}$ is linearly independent.

2-ii) For any $k \geq 1$, $x^{(k)}$ is the solution of a weighted lasso with positive weights and $\|y\|^2 > \epsilon$. Consequently from the lemma 9, for all $k \geq 1$, $\|y - Dx^{(k)}\|^2 = \epsilon$. Because the set $\{x \in \mathbb{R}^p \mid \|y - Dx\|^2 = \epsilon\}$ is a closed set, one deduces that the limit point $\tilde{x}$ satisfies $\|y - D\tilde{x}\|^2 = \epsilon$.

2-iii) By definition of $x^{(k)}$ we have

$$x^{(\phi(k)+1)} := \text{argmin} \sum_{i=1}^{p} f'_\alpha(|x_i^{(\phi(k))}| + \Delta)|x_i| \text{ subject to } \|y - Dx\|_2^2 \leqslant \epsilon.$$

According to [1] (chapter 5.3), there exists $\lambda \geq 0$ such that

$$x^{(\phi(k)+1)} := \text{argmin } f'_\alpha(|x_i^{(\phi(k))}| + \Delta)|x_i| + \lambda\|y - Dx\|_2^2.$$

Consequently, the subdifferential of the previous expression evaluated in $x^{(\phi(n)+1)}$ contains the null vector

$$\vec{0} \in \partial\text{pen}(x^{(\phi(k)+1)}) - \lambda D^T(y - Dx^{(\phi(k)+1)}), \tag{14}$$

with $\partial\text{pen}(x^{(\phi(k)+1)}) = C_1 \times \cdots \times C_p$, where

$$C_i := \begin{cases} \left[-f'_\alpha(|x_i^{(\phi(k))}| + \Delta), f'_\alpha(|x_i^{(\phi(k))}| + \Delta)\right] \text{ if } x_i^{(\phi(k)+1)} = 0 \\ \text{sign}(x_i^{\phi(k)+1}) f'_\alpha(|x_i^{(\phi(k))}|) \text{ otherwise} \end{cases}.$$

Since $\lim_{n\to+\infty} x^{(\phi(k))} = \lim_{k\to+\infty} x^{(\phi(k)+1)} = \tilde{x}$, the vectors $(x^{(\phi(k)+1)})_{i\in\text{supp}(\tilde{x})}$ and $(\tilde{x})_{i\in\text{supp}(\tilde{x})}$ have the same sign for k large enough. Moreover, since f'_α is continuous, by taking the limit in (14), we see that the vectors $(d_i^T(y - D\tilde{x}))_{i\in\text{supp}(\tilde{x})}$ and $(\text{sign}(\tilde{x}_i) f'_\alpha(\tilde{x}_i))_{i\in\text{supp}(\tilde{x})}$ are collinear.

Proof of proposition 2:

i) The proof of this part is exactly the same as the one provided in lemma 9.

ii) The proof of this part is exactly the same as the one provided in lemma 2.

iii) The vector $x^\alpha_{\text{supp}(x^\alpha)} := (x_i^\alpha)_{i\in\text{supp}(x^\alpha)}$ is a solution of the problem

$$\text{argmin} \sum_{i\in\text{supp}(x^\alpha)} f_\alpha(|x_i| + \Delta) \text{ subject to } \|y - \tilde{D}x\|_2^2 \leqslant \epsilon, \text{ with } \tilde{D} \text{ the matrix with columns } (d_i)_{i\in\text{supp}(x^\alpha)}. \tag{15}$$

Indeed, assume that $x^\alpha_{\text{supp}(x^\alpha)}$ is not a solution of the previous problem, then there exists $\bar{x} \in \mathbb{R}^{\text{Card}(\text{supp}(x^\alpha))}$ such that

$$\|y - \tilde{D}\bar{x}\|_2^2 \leqslant \epsilon \text{ and } \sum_{i\in\text{supp}(x^\alpha)} f_\alpha(|\bar{x}_i| + \Delta) < \sum_{i\in\text{supp}(x^\alpha)} f_\alpha(|x_i^\alpha| + \Delta).$$

Let us set $x' \in \mathbb{R}^p$ such that $x'_i := \bar{x}_i$ if $i \in \text{supp}(x^\alpha)$ and $x'_i := 0$ otherwise. By definition of x' we have $\|x'\|_0 \leq \|x^\alpha\|_0$. On the other hand, since $\tilde{D}\bar{x} = Dx'$ we have $\|y - Dx'\|^2 \leq \epsilon$ therefore $x' \in S_0^\epsilon$. Let us show that $\sum_{i=1}^p f_\alpha(|x'_i|+\Delta) < \sum_{i=1}^p f_\alpha(|x^\alpha_i|+\Delta)$

$$\begin{aligned}\sum_{i=1}^p f_\alpha(|x'_i|+\Delta) &= \sum_{i\notin \mathrm{supp}(x^\alpha)} f_\alpha(\Delta) + \sum_{i\in \mathrm{supp}(x^\alpha)} f_\alpha(|\bar{x}_i|+\Delta),\\ &< \sum_{i\notin \mathrm{supp}(x^\alpha)} f_\alpha(\Delta) + \sum_{i\in \mathrm{supp}(x^\alpha)} f_\alpha(|x^\alpha_i|+\Delta) = \sum_{i=1}^p f_\alpha(|x^\alpha_i|+\Delta).\end{aligned}$$

The previous inequality contradicts that $x^\alpha \in L^\alpha$. According to [1] (chapter 5.3), there exists $\lambda \geq 0$ such that $x^\alpha_{\mathrm{supp}(x^\alpha)}$, the solution of (15), is also the solution of the problem

$$\mathrm{argmin} \sum_{i\in \mathrm{supp}(x^\alpha)} f_\alpha(|x_i|+\Delta) + \lambda\|y-\tilde{D}x\|^2, \text{ where } \lambda \geq 0.$$

Because the partial derivatives of $\sum_{i\in \mathrm{supp}(x^\alpha)} f_\alpha(|x_i|+\Delta) + \lambda\|y-\tilde{D}x\|^2$ at $x^\alpha_{\mathrm{supp}(x^\alpha)}$ are null we have

$$\forall i \in \mathrm{supp}(x^\alpha), \mathrm{sign}(x^\alpha_i)f'_\alpha(|x^\alpha_i|+\Delta) - \lambda d_i^T(y-\tilde{D}x^\alpha_{\mathrm{supp}(x^\alpha)}) = 0.$$

Since $\tilde{D}x^\alpha_{\mathrm{supp}(x^\alpha)} = Dx^\alpha$, one obtains that the vectors $(\mathrm{sign}(x^\alpha_i)f'_\alpha(|x^\alpha_i|+\Delta))_{i\in \mathrm{supp}(x^\alpha)}$ and $\left(d_i^T(y-Dx^\alpha)\right)_{i\in \mathrm{supp}(x^\alpha)}$ are colinear. □

Acknowledgements

The authors are grateful for the real data provided by the following metabolomicians from Toxalim: Cécile Canlet, Laurent Debrauwer and Marie Tremblay-Franco. This work is part of the project GMO90+ supported by the grant CHORUS 2101240982 from the Ministry of Ecology, Sustainable Development and Energy in the national research program RiskOGM. Patrick Tardivel is partly supported by a phD fellowship from GMO90+. We also received a grant for the project from the IDEX of Toulouse "Transversalité 2014".

References

[1] Dimitri P Bertsekas. *Nonlinear programming.* Athena Scientific, 1999.

[2] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[3] Peter Bühlmann and Sara van de Geer. *Statistics for High-Dimensional Data: Methods, Theory and Applications.* Springer, 2011.

[4] Florentina Bunea, Alexandre Tsybakov, and Marten Wegkamp. Sparsity oracle inequalities for the Lasso. *Electronic Journal of Statistics*, 1:169–194, 2007.

[5] T. Tony Cai and Anru Zhang. Sharp RIP bound for sparse signal and low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 35(1):74–93, 2013.

[6] Emmanuel J Candes. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592, 2008.

[7] Emmanuel J Candes, Justin K Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[8] Emmanuel J Candes and Terence Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[9] Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted l_1 minimization. *Journal of Fourier Analysis and Applications*, 14(5-6):877–905, 2008.

[10] Rick Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters*, 14(10):707–710, 2007.

[11] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Compressed sensing and best k-term approximation. *Journal of the American Mathematical Society*, 22(1):211–231, 2009.

[12] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, 39(1):1–38, 1977.

[13] David L Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

[14] David L Donoho, Michael Elad, and Vladimir N Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, 2006.

[15] Charles Dossal. A necessary and sufficient condition for exact sparse recovery by l_1 minimization. *Comptes Rendus Mathematique*, 350(1):117–120, 2012.

[16] Simon Foucart and Ming-Jun Lai. Sparsest solutions of underdetermined linear systems via l_q-minimization for $0 < q \leq 1$. *Applied and Computational Harmonic Analysis*, 26(3):395–407, 2009.

[17] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing.* Springer, 2013.

[18] Rémi Gribonval and Morten Nielsen. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12):3320–3325, 2003.

[19] Rémi Gribonval and Morten Nielsen. Highly sparse representations from dictionaries are unique and independent of the sparseness measure. *Applied and Computational Harmonic Analysis*, 22(3):335–355, 2007.

[20] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer, 2009.

[21] David R Hunter and Kenneth Lange. A tutorial on MM algorithms. *The American Statistician*, 58(1):30–37, 2004.

[22] Ming-Jun Lai. On sparse solutions of underdetermined linear systems. *Journal of Concrete and Applicable Mathematics*, 8(2):296–327, 2010.

[23] Kenneth Lange. Elementary optimization. In *Optimization*, pages 1–17. Springer, 2004.

[24] Miguel Sousa Lobo, Maryam Fazel, and Stephen Boyd. Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research*, 152(1):341–365, 2007.

[25] Nicolai Meinshausen. Group bound: confidence intervals for groups of variables in sparse high dimensional regression without assumptions on the design. *Journal of the Royal Statistical Society: Series B*, 77(5):923–945, 2015.

[26] Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462, 2006.

[27] Saharon Rosset, Ji Zhu, and Trevor Hastie. Boosting as a regularized path to a maximum margin classifier. *Journal of Machine Learning Research*, 5:941–973, 2004.

[28] Qiyu Sun. Recovery of sparsest signals via l_q-minimization. *Applied and Computational Harmonic Analysis*, 32(3):329–341, 2012.

[29] Wei Sun, Junhui Wang, and Yixin Fang. Consistent Selection of Tuning Parameters via Variable Selection Stability. *Journal of Machine Learning Research*, 14:3419–3440, 2013.

[30] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[31] Ryan J Tibshirani. The lasso problem and uniqueness. *Electronic Journal of Statistics*, 7:1456–1490, 2013.

[32] Andreas M Tillmann and Marc E Pfetsch. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *IEEE Transactions on Information Theory*, 60(2):1248–1259, 2014.

[33] Sara A Van de Geer. High-dimensional generalized linear models and the lasso. *The Annals of Statistics*, 36(2):614–645, 2008.

[34] Sara A Van De Geer and Peter Bühlmann. On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, 3:1360–1392, 2009.

[35] Joseph Woodworth and Rick Chartrand. Compressed sensing recovery via nonconvex shrinkage penalties. *Inverse Problems*, 32(7):075004, 2016.

[36] Zheng Zhang, Yong Xu, Jian Yang, Xuelong Li, and David Zhang. A survey of sparse representation: algorithms and applications. *IEEE Access*, 3:490–530, 2015.

[37] Peng Zhao and Bin Yu. On model selection consistency of lasso. *Journal of Machine Learning Research*, 7:2541–2563, 2006.

[38] Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429, 2006.