A Smoothing Method for Sparse Optimization over Polyhedral Sets

Mounir Haddou, Tangi Migot

A Smoothing Method for Sparse Optimization over Polyhedral Sets

M. Haddou* T. Migot†

Abstract

In this paper, we investigate a class of heuristic schemes to solve the NP-hard problem of minimizing ℓ_0-norm over a polyhedral set. A well-known approximation is to consider the convex problem of minimizing ℓ_1-norm. We are interested in finding improved results in cases where the problem in ℓ_1-norm does not provide an optimal solution to the ℓ_0-norm problem. We consider a relaxation technique using a family of smooth concave functions depending on a parameter. Some other relaxations have already been tried in the literature and the aim of this paper is to provide a more general context. This motivation allows deriving new theoretical results that are valid for general constraint set. We use a homotopy algorithm, starting from a solution to the problem in ℓ_1-norm and ending in a solution of the problem in ℓ_0-norm. We show the existence of the solutions of the subproblem, convergence results, a kind of monotonicity of the solutions as well as error estimates leading to an exact penalization theorem. We also provide keys for implementing the algorithm and numerical simulations.

Mathematics Subject Classification. 90-08 and 65K05
Keywords : smoothing functions ; sparse optimization ; concave minimization ; l0-norm

Introduction

Consider a polyhedron F defined by linear inequalities, $F = \{x \in \mathbb{R}^n | \ Ax \leq b\} \cap \mathbb{R}^n_+$ for some $b \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times n}$, which we suppose non-empty and not reduced to a singleton. Although we consider a polyhedron here, most of the results presented in this article can be generalized as F being a closed convex set in $\mathbb{R}^n_+$. One should note that the hypothesis of considering a polyhedron in the non-negative orthant is not restrictive. It is only assumed to simplify the presentation and to avoid the absolute value in the definition of the problem.

*IRMAR-INSA Rennes ; mounir.haddou@insa-rennes.fr
†IRMAR-INSA Rennes ; tangi.migot@insa-rennes.fr

We are interested in finding the sparsest point over this polyhedron, which is equivalent to minimize the ℓ_0-norm, i.e.

$$\min_{x \in F} \|x\|_0, \tag{P_0}$$

where

$$\forall x \in \mathbb{R}^n,\ \|x\|_0 := \sum_{i=1}^{n} s(|x_i|), \text{ where for } t \in \mathbb{R},\ s(t) = \{0 \text{ if } t = 0\ ;\ 1 \text{ otherwise}\}. \tag{1}$$

Note that the ℓ_0-norm is not a norm as it does not have the homogeneity property. (P_0) is an NP-hard problem as shown in [49].

This problem has several applications and received a considerable interest recently. Sparsity is involved in several domains including signal and image processing [52, 34, 21, 43, 10], statistics [30, 61, 57], machine learning [9, 44, 46]. The compressed sensing [17, 22, 26, 23, 11, 12, 14] has been the most popular application involving sparsity and creating cross-disciplinary attention in recent years and stimulates a plethora of new applications of sparsity. For more details about applications in image and signal modelling as well as a review on related questions see [10] or [58].

The problem (P_0) being difficult to solve, a classical approximation consists in solving the convex problem in ℓ_1-norm. The ℓ_1-norm is denoted by

$$\forall x \in \mathbb{R}^n,\ ||x||_1 = \sum_{i=1}^{n} |x_i|. \tag{2}$$

The convex problem in ℓ_1-norm is defined by

$$\min_{x \in F} ||x||_1. \tag{P_1}$$

It can be seen as a convexification of (P_0), because the absolute value of x is the convex envelope of $s(x)$ for $x \in [-1, 1]$. Furthermore, (P_1) has the benefits that it can be reformulated as a linear program.

This approach has been extensively studied in [24, 13, 14, 17, 27, 31, 60] and in particular with inequality constraints. Moreover, several criteria have been found which guarantee that solving (P_1) will also solve (P_0) under various assumptions involving the coefficients of the matrix A. These criteria, denoted mutual coherence [25], restricted isometry property [14], null space property [19], exact recovery condition [59, 31], and the range space property [66], show the efficiency of this convex approximation to solve (P_0).

A more sophisticated version of this convex formulation and computationally efficient approach consider a reweighted-ℓ_1 problem as proposed in [15] and later studied in several recent papers, see [2, 50, 65, 64, 18, 67, 68]. It is clear from this references that the study of the convex problem (P_1) to solve (P_0) is of great importance.

Also formulation (P_1) does not solve all the time the initial problem. Consider for instance the following example in two dimension.

Example 0.1. *Given a matrix $A \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$ such that*

$$A = \begin{pmatrix} -0.1 & -1 \\ -10 & -1 \end{pmatrix} \text{ and } b = \begin{pmatrix} -1 \\ -10 \end{pmatrix}. \tag{3}$$

Geometrical observation allows to conclude that the solution of problem (P_1) is $(\frac{10}{11}, \frac{10}{11})^T$, while solution of problem (P_0) are of the form $(0, 10+\epsilon)^T$ and $(10+\epsilon, 0)^T$ with $\epsilon \geq 0$.

Nonconvex optimization has been one of the main approach to tackle this problem [63, 28, 7, 62, 39]. For instance, in [28, 7], the authors proposed a reformulation of the problem as a mathematical program with complementarity constraints. Thresholding algorithms have also some recent popularity in [62, 20, 8, 51, 5, 42]. A Difference of Convex (DC) decomposition of the ℓ_0-norm combined with DC Algorithm has been used in [39]. We are interested here in nonconvex methods to improve the solution we get by solving (P_1) in the general case where this approach does not solve the initial problem. In this aim, several concave relaxations of $||.||_0$ have been tried in the literature.

An intuitive approach trying to bridge the gap between the ℓ_1-norm and the ℓ_0-norm has been to study homotopy methods based on the ℓ_p-norm for $0 \leq p \leq 1$. This approach has been initiated in [35] and later analyzed in [32, 16, 29, 38, 33], where the authors prove the link between (P_0) and (P_1) as well as conditions involving the coefficients of A to show a sufficient convergence condition, so that p does not have to decrease to 0 but only to some small value. The homotopy method considers non-convex subproblems and solving the problem in ℓ_p is not a trivial task. In [32], the authors study a linearization algorithm, while in [33] the authors consider an interior-point method to solve the subproblems. Besides, the problem of minimizing the ℓ_p-norm might lead to numerical difficulties due to the non-differentiability at the origin, in [38] the authors consider a smoothing of the ℓ_p-norm to circumvent this problem.

Following the progress made during the last decade in the study of reweighted ℓ_1-norm and ℓ_p-norm, we study here smooth regularizations. In [48] and related works the authors present a general family of smoothing function including the gaussian family and propose a homotopy method starting from the ℓ_2-norm solutions.

Approximating the ℓ_0-norm by smooth functions through an homotopy method starting from the ℓ_1-norm has been studied in the PhD thesis [53] and in [55, 54, 41]. In these works, the authors consider a selection of minimization problems using smooth functions such that $(t+r)^p$ with $r > 0$ and $0 < p < 1$, $-(t+r)^{-p}$ with $r > 0$ and $1 < p$, $\log(t+r)$ with $0 < r << 1$ or $1 - e^{-rt}$ with $r > 0$ and $p \in \mathbb{N}$. The subproblems of the homotopy algorithm are solved using a Frank and Wolfe approach [55], also called SLA in [32], and this method is further studied in [40].

The aim of this paper is to pursue the study of smooth concave approximation of the ℓ_0-norm by offering a more general theoretical context for this study. Focusing on concave functions is a logical choice considering that the ℓ_p-norm is itself concave. The motivation here is to keep the good properties of the method from [55] and related work, a homotopy method between the ℓ_1-norm and the ℓ_0-norm problems, and smoothness at the origin. In particular, such a theoretical study has not been done in the literature.

The method considered here is a homotopy method with a parameter r such that the method recovers the ℓ_1-norm problem for r large and the ℓ_0-norm problem for r small. We provide here a complete analysis of the convergence of the algorithm as well as a monotonicity study of the objective function during the iterations of the homotopy scheme. We also prove the existence of the solutions of the subproblems without any boundedness assumption on the constraints.

For the convex problem of minimizing the ℓ_1-norm, we already pointed out that several criteria involving the coefficients of the matrix A guarantee that solving the problem is sufficient to compute a solution to (P_0). Such a result guarantees the good behavior of the method. Considering our homotopy algorithm, we show a similar result independently of the constraints that state that it is not necessary to tend r to zero to compute a solution of (P_0). It can be seen as an exact penalty result. This property is a key to ensure the interest of the method.

Most of the theoretical results presented here are valid for any non-empty closed convex set F, which make them valid for several smoothing functions but also for several formulations of the problem.

In order to validate our approach, we give technical details and some numerical results on a Frank and Wolfe method to solve the subproblems of the homotopy scheme. In particular, these results show that we manage to improve the results given by the ℓ_1 norm, which shows the validity of our approach.

This document is organized as follows. Section 1 introduces a general formulation of the relaxation methods using concave functions. Section 2 discusses convergence and monotonicity results leading to a homotopy method. Section 3 proves error estimates and an exact penalization theorem. Finally, Section 4 presents the algorithm with several remarks concerning its implementation and numerical results can be found in Section 5.

1 A smoothing method

We consider a family of smooth functions designed to approximate the ℓ_0-norm. This family has already been used in the different context of complementarity [37, 1] and image restoration [6]. These functions are smooth non-decreasing concave functions such that

$$\theta : \mathbb{R} \to]-\infty, 1[\text{ with } \theta(t) < 0 \text{ if } t < 0,\ \theta(0) = 0 \text{ and } \lim_{t \to +\infty} \theta(t) = 1. \tag{4}$$

One way to build θ functions is to consider non-increasing probability density functions $f : \mathbb{R}_+ \to \mathbb{R}_+$ and then take the corresponding cumulative distribution function

$$\forall t \geq 0,\ \theta(t) = \int_0^t f(x)dx \text{ and } \forall t < 0,\ \theta(t) < 0. \tag{5}$$

By definition of f we can verify that

$$\lim_{t\to+\infty}\theta(t)=\int_0^{+\infty}f(x)dx=1 \text{ and } \theta(0)=\int_0^0 f(x)dx=0. \tag{6}$$

The non-increasing hypothesis on f gives the concavity of θ.

Examples of this family are $\theta^1(t)=t/(t+1)$ if $t\geq 0$ and $\theta^1(t)=t$ if $t<0$, $\theta^2(t)=1-e^{-t}$ with $t\in\mathbb{R}$.

Then using a scaling technique similar to the perspective functions in convex analysis we define $\theta(t,r):=\theta\left(\frac{t}{r}\right)$ for $r>0$ and we get

$$\theta(0,r)=0\ \forall r>0 \text{ and } \lim_{r\to 0}\theta(t,r)=1\ \forall t>0. \tag{7}$$

For the previous examples of this family and $t\geq 0$ we have $\theta^1(t,r)=t/(t+r)$, $\theta^2(t,r)=1-e^{-t/r}$. The function $\theta^1(t,r)$ will be extensively used in this paper.

Throughout this paper we will consider the concave optimization problem for $r>0$

$$\min_{x\in F}\sum_{i=1}^{n}\theta(x_i,r). \tag{P_r}$$

Before moving to the proofs of convergence, we give a result on the existence of solutions of (P_r). The proof relies on an argument similar to the use of asymptotic cones and directions as introduced in [3].

Theorem 1.1. *Let $F\subset\mathbb{R}^n_+$ be a non-empty closed convex set. The optimal set of P_r for $r>0$ is non-empty.*

Proof. Since $\sum_{i=1}^n\theta(x_i,r)$ is bounded below on the closed set F it admits an infimum. Now, assume by contradiction that there exists an unbounded sequence $\{x_n\}$ such that $x_n\in F,\ \forall n$ and

$$\lim_{n\to\infty}f(x_n)=\inf_{x\in F}f(x)<f(x_0).$$

Let $\{d_n\}$ be the sequence defined for all n by

$$d_n:=\frac{x_n-x_0}{\|x_n\|}.$$

This sequence is bounded, therefore it converges, up to a subsequence, to some limit, $\lim_{n\to\infty}d_n=\lim_{n\to\infty}x_n/\|x_n\|=d\in F^\infty$, where F^∞ denotes the cone of asymptotic directions of F (cf. [3]). Since F is a closed convex set, it holds for all $x\in F$ that

$$x+\alpha d\in F,\ \forall\alpha\geq 0.$$

Then, since $F\subset\mathbb{R}^n_+$, we obtain that $d\geq 0$.

Using component-wise monotonicity and continuity assumption on θ gives

$$\lim_{n\to\infty}f(x+\alpha_n d_n)\geq f(x),\ \forall x\in F$$

as long as $\alpha_n > 0$ for all n and the sequence $\{\alpha_n d_n\}$ admits some limit. Choosing $x = x_0$, $\alpha_n = \|x_n\|$ and d_n as defined above, we obtain

$$\lim_{n\to\infty} f(x_0 + \alpha_n d_n) = \lim_{n\to\infty} f(x_n) \geq f(x_0),$$

which is a contradiction with our initial assumption. This completes the proof. □

2 Convergence

In this section, we will show the link between problems (P_0), (P_1) and (P_r). We denote $S^*_{||.||_0}$ the set of solutions of (P_0), $S^*_{||.||_1}$ the set of solutions of (P_1) and S^*_r the set of solutions of (P_r).

Our aim is to illustrate that for r sufficiently large (P_r) is close to (P_1) (see Theorem 2.2), and for r sufficiently small (P_r) is close to (P_0) (see Theorem 2.1). In this way, we define an homotopy method starting from r large and decreasing r step by step. Thus, we use the convex approximation (P_1) and come closer and closer to the problem we want to solve. A monotonicity-kind result of the sequence computed by the homotopy scheme is proved in Theorem 2.3. Finally, Theorem 2.4 shows that this formulation may also be of interest for more complicated objective function than the one in (P_0).

Theorem 2.1 gives convergence of (P_r) to (P_0) for r decreasing to 0.

Theorem 2.1 (Convergence to ℓ_0-norm). *Let $F \subset \mathbb{R}^n_+$ be a non-empty closed convex set. Every limit point of any sequence $\{x_r\}_r$, such that $x_r \in S^*_r$ and $r \downarrow 0$, is an optimal solution of* (P_0).

Proof. Given $\bar{x}$ the limit of the sequence $\{x_r\}_r$, up to a subsequence, and $x^* \in S^*_{||.||_0}$. Since F is a closed set one has $\bar{x} \in F$. Furthermore we have for any r in the corresponding subsequence

$$\sum_{i\in supp(\bar{x})} \theta(x_{r,i}, r) \leq \sum_{i=1}^{n} \theta(x_{r,i}, r) \leq \sum_{i=1}^{n} \theta(x^*_i, r) \leq \|x^*\|_0. \tag{8}$$

Moreover the definition of $\theta(., r)$ functions, for $r > 0$ and $t \in \mathbb{R}^n$ give

$$\sum_{i=1}^{n} \lim_{r\downarrow 0} \theta(t_i, r) = ||t||_0. \tag{9}$$

Replacing into (8) we get

$$||\bar{x}||_0 \leq ||x^*||_0, \tag{10}$$

and thanks to the definition of $\bar{x}$

$$||\bar{x}||_0 = ||x^*||_0. \tag{11}$$

□

We now give another convergence result from [32], which adds that the convergence appears in a finite number of iteration in the case, where the feasible set of (P_0) is a polyhedron.

Proposition 2.1. *Given a non-empty polyhedron $F \subset \mathbb{R}^n_+$. Then there exists a $\bar{r}$ such that for all $r \leq \bar{r}$ a vertex of F is an optimal solution of (P_0) and $(P_{r \leq \bar{r}})$.*

Proof. (P_r) is a problem of minimizing a concave function over a polyhedron F. We can use Corollary 32.3.4 of [56], since there is no half-line in F such that $\theta(.,r)$ is unbounded below, so the infimum over F is attained and it is attained at one of the extremal points of F.

Given that there is a finite number of extremal point, one vertex, say x', will repeadetly solve (P_r) for some increasing infinite sequence $R = (r_0, r_1, r_2, ...)$. Moreover the objective function of (P_r) is non-increasing and bounded below by the infimum of ℓ_0-norm, so

$$\sum_{i=1}^{n} \theta(x'_i, r_j) = \min_{x \in F} \sum_{i=1}^{n} \theta(x_i, r_j) \leq \inf_{x \in F} ||x||_0. \tag{12}$$

Going through the limit in R for $j \to \infty$ and as the concave function is continuous and $x' \in F$, we have the results. $\square$

Theorem 2.2 and Proposition 2.1 show that the scheme converge to (P_0) as r decreases to zero.

The next theorem shows that for r sufficiently large the solutions of (P_r) are the same than solutions of (P_1). This will be especially useful as an initialization of the homotopy scheme.

Theorem 2.2 (Convergence to ℓ_1-norm)**.** *Let $F \subset \mathbb{R}^n_+$ be a non-empty closed convex set. Every limit point of any sequence $\{x_r\}_r$, such that $x_r \in S^*_r$ for $r \uparrow \infty$, is an optimal solution of (P_1).*

Proof. As $r > 0$, we can use a scaling technique for $S^{*(2)}_r = \arg\min_{x \in F} \sum_{i=1}^n r\theta(x_i, r)$

$$\min_{x \in F} \sum_{i=1}^{n} \theta(x_i, r) \quad \Longleftrightarrow \quad \min_{x \in F} \sum_{i=1}^{n} r\theta(x_i, r), \tag{13}$$

$$S^*_r \quad = \quad S^{*(2)}_r. \tag{14}$$

So, it is sufficient to show that every limit point of any sequence $\{x_r\}_r$, such that $x_r \in S^*_r$ for $r \uparrow \infty$, is an optimal solution of (P_1).

Given $x^r \in S^{*(2)}_r$ and $\bar{x} \in S^*_{||\cdot||_1}$. We use the first order Taylor's theorem for $\theta(t)$ in 0,

$$\theta(t) = t\theta'(0) + g(t), \text{ where } \lim_{t \to 0} \frac{g(t)}{t} = 0. \tag{15}$$

By concavity of the functions θ, it holds that $\theta'(0) > 0$.

By definition of $\bar{x}$, we get

$$\sum_{i=1}^{n} r\theta(x_i^r, r) \leq \sum_{i=1}^{n} r\theta(\bar{x}_i, r). \tag{16}$$

Now, using (15) yields

$$\begin{aligned}
\sum_{i=1}^{n} x_i^r \theta'(0) + rg(\frac{x_i^r}{r}) &\leq \sum_{i=1}^{n} x_i^r \theta'(0) + rg(\frac{\bar{x}_i}{r}), && (17)\\
\sum_{i=1}^{n} x_i^r - \sum_{i=1}^{n} \bar{x}_i &\leq \frac{r}{\theta'(0)} \sum_{i=1}^{n} g(\frac{\bar{x}_i}{r}) - \frac{r}{\theta'(0)} \sum_{i=1}^{n} g(\frac{x_i^r}{r}), && (18)\\
&\leq \frac{1}{\theta'(0)} \left| \sum_{i=1}^{n} \frac{g(\frac{\bar{x}_i}{r})}{\frac{\bar{x}_i}{r}} \bar{x}_i \right| + \frac{1}{\theta'(0)} \left| \sum_{i=1}^{n} \frac{g(\frac{x_i^r}{r})}{\frac{x_i^r}{r}} x_i^r \right|, && (19)\\
&\leq \frac{1}{\theta'(0)} \left(\sum_{i=1}^{n} \left| \frac{g(\frac{\bar{x}_i}{r})}{\frac{\bar{x}_i}{r}} \right| \right) \left(\sum_{i=1}^{n} \bar{x}_i \right), && (20)\\
&+ \frac{1}{\theta'(0)} \left(\sum_{i=1}^{n} \left| \frac{g(\frac{x_i^r}{r})}{\frac{x_i^r}{r}} \right| \right) \left(\sum_{i=1}^{n} x_i^r \right), && (21)
\end{aligned}$$

$$\sum_{i=1}^{n} x_i^r \leq \left(\sum_{i=1}^{n} \bar{x}_i \right) \frac{1 + \frac{1}{\theta'(0)} \left(\sum_{i=1}^{n} \left| \frac{g(\frac{\bar{x}_i}{r})}{\frac{\bar{x}_i}{r}} \right| \right)}{1 - \frac{1}{\theta'(0)} \left(\sum_{i=1}^{n} \left| \frac{g(\frac{x_i^r}{r})}{\frac{x_i^r}{r}} \right| \right)}. \tag{22}$$

Then, we show that the right-hand side in previous equation goes to 1, when passing to the limit.

It holds true that

$$\lim_{r \to +\infty} \frac{\bar{x}}{r} = 0. \tag{23}$$

Besides, by definition of x^r yields

$$\begin{aligned}
\sum_{i=1}^{n} \theta(x_i^r, r) &\leq \sum_{i=1}^{n} \theta(\bar{x}_i, r) && (24)\\
\lim_{r \to +\infty} \sum_{i=1}^{n} \theta(x_i^r, r) &\leq \lim_{r \to +\infty} \sum_{i=1}^{n} \theta(\bar{x}_i, r) \leq 0. && (25)
\end{aligned}$$

so, we get

$$\lim_{r \to +\infty} \sum_{i=1}^{n} \theta(x_i^r, r) = 0. \tag{26}$$

By definition of functions θ, it is true that $\theta(x, r) := \theta(x/r)$ and $\theta^{-1}(0, r) = 0$. Thus, by previous equation we obtain

$$\lim_{r \to +\infty} \frac{x_i^r}{r} = 0 \ \forall i. \tag{27}$$

Using (23) and (27) it follows

$$\lim_{r\to+\infty} \frac{\bar{x}_i}{r} = 0 \implies \lim_{r\to+\infty} \frac{g(\frac{\bar{x}_i}{r})}{\frac{\bar{x}_i}{r}} = 0, \tag{28}$$

$$\lim_{r\to+\infty} \frac{x_i^r}{r} = 0 \implies \lim_{r\to+\infty} \frac{g(\frac{x_i^r}{r})}{\frac{x_i^r}{r}} = 0. \tag{29}$$

Then going to the limit in (22) yields

$$\lim_{r\to+\infty} \sum_{i=1}^{n} x_i^r \leq \sum_{i=1}^{n} \bar{x}_i. \tag{30}$$

However, by definition of $\bar{x}$, it always hold that $\sum_{i=1}^{n} \bar{x}_i \leq \sum_{i=1}^{n} x_i$ for all x feasible for (P_1). Since, this is true for the limit point of the sequence $\{x^r\}_r$, the inequality in (30) is actually an equality. So, the limit point of the sequence $\{x^r\}_r$ is also a solution of (P_1). This proves the result. □

The next theorem gives a monotonicity result, which illustrates the relations between the three problems (P_0), (P_1) and (P_r). By monotonicity, we mean that for a given feasible point we want a relation of monotony in r for the objective function of (P_1), (P_r) and (P_0). As the components of the ℓ_0-norm and the $\theta_r(t)$ are in $[0, 1[$ it is necessary to put the components of ℓ_1-norm in a similar box, which explains the change of variable in the theorem.

Remark 2.1. *In the following theorem we use the hypothesis that θ functions are convex in r. This is not so restrictive as we think several functions verify it. If we take the three examples of θ functions given in the introduction, θ^1 and $\theta^{log} := \log(1 + x)/\log(1 + x + r)$ are convex in r but not θ^2.*

Theorem 2.3 (Monotonicity of solutions)**.** *Given $x \in F$, we define $y := x/(||x||_\infty + \epsilon)$ where $\epsilon > 0$, so that $y \in [0, 1[^n$. Let a function $\Psi(t, r) : [0, 1[\to [0, 1[$ be defined as*

$$\Psi(t, r) = \frac{\theta(t, r)}{\theta(1, r)}. \tag{31}$$

We consider here functions θ that are convex with respect to r. For r and $\bar{r}$ such that $0 < \bar{r} < r < +\infty$, then one has

$$||y||_1 \leq \sum_{i=1}^{n} \Psi(y_i, r) \leq \sum_{i=1}^{n} \Psi(y_i, \bar{r}) \leq ||y||_0. \tag{32}$$

Proof. The proof is divided in three step regarding the three inequalities.

The functions θ are sub-additive functions, since they concave and $\theta(0) = 0$. Then, it follows

$$\theta(y_i, r) \geq y_i \; \theta(1, r). \tag{33}$$

Therefore, we get

$$\sum_{i=1}^{n} \Psi(y_i, r) - \ ||y||_1 \ = \ \sum_{i=1}^{n} (\frac{\theta(y_i, r)}{\theta(1, r)} - \ y_i), \tag{34}$$

$$\geq \ 0, \tag{35}$$

which leads to the first inequality

$$||y||_1 \leq \ \sum_{i=1}^{n} \Psi(y_i, r). \tag{36}$$

We continue with the second inequality showing that $\Psi(y, r)$ functions are non-increasing in r, i.e.

$$\sum_{i=1}^{n} \Psi(y_i, r) \leq \ \sum_{i=1}^{n} \Psi_(y_i, \bar{r}). \tag{37}$$

The functions $\Psi(y, r)$ is non-increasing in r if its derivative with respect to r

$$\frac{\partial}{\partial r}\Psi(y, r) = \frac{(\frac{\partial}{\partial r}\theta(y, r))\theta(1, r) - (\frac{\partial}{\partial r}\theta(1, r))\theta(y, r)}{\theta(1, r)^2}, \tag{38}$$

is negative. Since $\theta(y, r)$ is an non-decreasing function in y we have

$$\frac{\theta(y, r)}{\theta(1, r)} < 1, \tag{39}$$

and

$$\frac{\partial}{\partial r}\theta(y, r) = -\frac{1}{r^2}, \frac{\partial}{\partial y}\theta(y, r) < 0. \tag{40}$$

So, $\theta(y, r)$ is non-increasing function in r. Using convexity of $\theta(y, r)$ in r it follows

$$\frac{\frac{\partial}{\partial r}\theta(y, r)}{\frac{\partial}{\partial r}\theta(1, r)} = \frac{\frac{\partial}{\partial r}\theta(1, r/y)}{\frac{\partial}{\partial r}\theta(1, r)} \geq 1. \tag{41}$$

Then in (38) the derivative with respect to r is negative and we have (37). Finally, since $\theta(y, r)$ is non-decreasing in y and $y \ \in \ [0, 1[^n$ one has

$$||y||_0 - \sum_{i=1}^{n} \Psi(y_i, \bar{r}) = \sum_{i=1; y_i \neq 0}^{n} 1 - \frac{\theta(y_i, \bar{r})}{\theta(1, \bar{r})} \ \geq \ 0, \tag{42}$$

which gives the last inequality and completes the theorem. □

Remark 2.2. *Both choice of scaling parameter in Theorem 2.2 and Theorem 2.3 are linked. In the former, we set that* $\lim_{r \to +\infty} r \sum_{i=1}^{n} \theta(x_i, r) = \sum_{i=1}^{n} x_i$, *so evaluating in one dimension and* $x = 1$ *we have* $\lim_{r \to +\infty} r\theta(1, r) = 1$ *and then we see that* r *and* $1/\theta(1, r)$ *have the same behavior for* r *sufficiently large.*

All these results lead us to the general behavior of the method. First, we start from one solution of (P_1) then by decreasing parameter r the solution of (P_r) becomes closer to a solution of (P_0).

Another approach would be to define a new problem which selects one solution of the possibly many optimal solutions of (P_0). We consider the following problem which is a selective version of (P_r)

$$\min_{x \in F} \sum_{i=1}^{n} \theta(x_i, r) + \sum_{i=1}^{n} r^{\frac{i}{2n+1}} x_i. \tag{P_{r-sel}}$$

We will use a lexicographic norm and we note

$$||y||_{lex} < ||x||_{lex} \iff \exists i \in \{1, ..., n\},\ y_i < x_i \text{ and } \forall\ 1 \leq j < i,\ y_j = x_j. \tag{43}$$

In the next theorem we want to choose the solution of (P_0) which has the smallest lexicographic norm. From the previous equation it is clear that this optimal solution is unique.

Theorem 2.4 (Convergence of the selective concave problem)**.** *We use functions θ such that $\theta \geq \theta^1$. Given $\{x^r\}_r$ the sequence of solutions of (P_{r-sel}) and $\bar{x}$ a limit point of this sequence. Then, $\bar{x}$ is the unique solution of $S^*_{||\cdot||_0}$ such that $\forall y \in S^*_{||\cdot||_0}$, $||\bar{x}||_{lex} \leq ||y||_{lex}$.*

Proof. Let x^* be an optimal solution of (P_0) such that $\forall y \in S^*_{||\cdot||_0}$, $||x^*||_{lex} \leq ||y||_{lex}$ and $\bar{x}$ the limit of a sequence of $\{x^r\}_r$ solution of (P_{r-sel}). So, there exists a $\bar{r}$ such that for every $r < \bar{r}$ we have $\bar{x}$ solution of (P_{r-sel}), up to a subsequence, and

$$\sum_{i=1}^{n} \theta(\bar{x}_i, r) + \sum_{i=1}^{n} r^{\frac{i}{2n+1}} \bar{x}_i \leq \sum_{i=1}^{n} \theta(x^*_i, r) + \sum_{i=1}^{n} r^{\frac{i}{2n+1}} x^*_i \leq \|x^*\|_0 + \sum_{i=1}^{n} r^{\frac{i}{2n+1}} x^*_i. \tag{44}$$

Going to the limit for $r \downarrow 0$ we have

$$||\bar{x}||_0 \leq ||x^*||_0, \tag{45}$$

which is an equality by definition of x^* and prove the first part of the theorem. Now we need to verify the selection of the solution. Using that functions θ are bounded by 1, one has

$$\sum_{i=1}^{n} (\theta(\bar{x}_i, r) - 1) + k + \sum_{i=1}^{n} r^{\frac{i}{2n+1}} \bar{x}_i \leq k + \sum_{i=1}^{n} r^{\frac{i}{2n+1}} x^*_i, \tag{46}$$

using that $\theta \geq \theta^1$

$$\sum_{i=1}^{n} (\theta^1(\bar{x}_i, r) - 1) + \sum_{i=1}^{n} r^{\frac{i}{2n+1}} \bar{x}_i \leq \sum_{i=1}^{n} (\theta(\bar{x}_i, r) - 1) + \sum_{i=1}^{n} r^{\frac{i}{2n+1}} \bar{x}_i \leq \sum_{i=1}^{n} r^{\frac{i}{2n+1}} x^*_i. \tag{47}$$

Now, for r sufficiently small, such that $\min_{\{i|x_i\neq 0\}} \bar{x}_i \geq \sqrt{r}$, we have

$$-k\frac{r}{r+\sqrt{r}} + \sum_{i=1}^{n} r^{\frac{i}{2n+1}} \bar{x}_i \leq \sum_{i=1}^{n} r^{\frac{i}{2n+1}} x_i^*, \tag{48}$$

where k denotes the optimal value of (P_0), i.e. $\|\bar{x}\|_0 = \|x^*\|_0 = k$.

Dividing by $r^{\frac{1}{2n+1}}$ in the previous inequality yields

$$-k\frac{r}{r^{\frac{1}{2n+1}}(r+\sqrt{r})} + x_1 + \sum_{i=2}^{n} r^{\frac{i}{2n+1}} \bar{x}_i \leq x_1^* + \sum_{i=2}^{n} r^{\frac{i}{2n+1}} x_i^*. \tag{49}$$

Therefore, going to the limit for $r \downarrow 0$ one has

$$\bar{x_1} \leq x_1^* \tag{50}$$

which is an equality by hypothesis on x^* being the smallest $||.||_{lex}$ solution of (P_0). So, as $\bar{x_1} = x_1^*$ in (48) one has

$$-k\frac{r}{r+\sqrt{r}} + \sum_{i=2}^{n} r^{\frac{i}{2n+1}} \bar{x}_i \leq \sum_{i=2}^{n} r^{\frac{i}{2n+1}} x_i^*. \tag{51}$$

By induction we get $\bar{x_i} = x_i^*,\ \forall i \in \{1, ..., n\}$ and so $\bar{x} = x^*$, because we have

$$\forall j \in \{1, ..., n\},\ \lim_{r\to 0} \frac{r}{r^{\frac{j}{2n+1}}(r+\sqrt{r})} = 0. \tag{52}$$

Finally we have the results as $\bar{x}$ is the optimal solution which has the smallest lexicographic norm. □

Remark 2.3. *If we try to get an equivalent result as in Theorem 2.2 for this selection problem, it is clear that for r sufficiently large we will solve the ℓ_1-norm problem but with a reversed lexicographical order than the one we are looking for, i.e. for a non-decreasing sequence of r_j*

$$\bar{x} = \lim_{j\to\infty} \{x^{r_j}\}_{r_j} \text{ with } x^{r_j} \in S^*_{r_j - sel} \Longrightarrow \bar{x} \in S^*_{||.||_1} \text{ and } \bar{x} = arg \max_{y \in S^*_{||.||_1}} ||y||_{lex}. \tag{53}$$

This will definitely prevent us of any kind of monotonicity result such as Theorem 2.3. So, unless $S^*_{||.||_0}$ admits only one solution, the initial point as a solution of (P_1) has no chance of being a good initial point. This argument and the fact that this problem looks numerically not advisable lead us not to follow the study of this selective problem.

3 Error estimate

In this section we focus on what happen when r becomes small. We denote card(I) the number of elements in a set I. Note that the following results are given for functions $\theta \geq \theta^1$ with $\theta^1(t, r) = t/(t + r)$ for $t, r \in \mathbb{R}_+$.

Lemma 3.1. *Consider θ functions where $\theta \geq \theta^1$. Let $\mathbb{N} \ni k = ||x^*||_0 < n$ be the optimal value of problem (P_0) and $I(x,r) = \{i | x_i \geq kr\}$. Then one has*

$$x^r \in \arg\min_{x \in F} \sum_{i=1}^{n} \theta(x_i, r) \Rightarrow \mathit{card}(I(x^r, r)) \leq k. \tag{54}$$

Proof. We use a proof by contradiction. Consider that $\text{card}(I(x^r, r)) \geq k+1$ and we have $x^r \in \arg\min_{x \in F} \sum_{i=1}^{n} \theta(x_i, r)$, then

$$\sum_{i=1}^{n} \theta(x_i^r, r) \geq (k+1)\theta(kr, r) \geq (k+1)\theta^1(kr, r) = (k+1)\frac{kr}{kr+r} = k, \tag{55}$$

which is a contradiction with the definition of x^r. □

This lemma gives us a theoretical stopping criterion for the decrease of r, as for $r < \bar{r} = \min_{x_i^r \neq 0} x_i^r / k$, x^r becomes an optimal solution. In the following lemma we look at the consequences in the evaluation of θ.

Lemma 3.2. *Consider θ functions where $\theta \geq \theta^1$. Let $\mathbb{N} \ni k = ||x^*||_0 < n$ be the optimal value of problem (P_0) and*

$$\bar{r} = \min_{x_i^r \neq 0} x_i^r / k.$$

Then one has

$$r \leq \bar{r} \Longleftrightarrow \theta(\min_{x_i^r \neq 0} x_i^r, r) \geq \theta^1(\min_{x_i^r \neq 0} x_i^r, r) \geq \frac{k}{k+1}. \tag{56}$$

Proof. We first show the equivalence in (56) for θ^1. Assume that

$$\theta^1(\min_{x_i^r \neq 0} x_i^r, r) \geq \frac{k}{k+1}. \tag{57}$$

Using the expression of θ^1, it follows

$$\theta^1(\min_{x_i^r \neq 0} x_i^r, r) = \frac{\min_{x_i^r \neq 0} x_i^r}{\min_{x_i^r \neq 0} x_i^r + r} \geq \frac{k}{k+1} \tag{58}$$

$$\Longleftrightarrow \quad \min_{x_i^r \neq 0} x_i^r (k+1) \geq k(\min_{x_i^r \neq 0} x_i^r + r) \tag{59}$$

$$\Longleftrightarrow \quad \min_{x_i^r \neq 0} x_i^r \geq kr \tag{60}$$

$$\Longleftrightarrow \quad \bar{r} = \frac{\min_{x_i^r \neq 0} x_i^r}{k} \geq r. \tag{61}$$

Considering the functions θ such that $\theta \geq \theta^1$, the equivalence follows in the exact same way. This proves the result. □

Both previous lemmas lead us to the following theorem, which is an exact penalization result for our method.

Theorem 3.1 (Exact Penalization Theorem)**.** *Consider θ functions where $\theta \geq \theta^1$. Let $\mathbb{N} \ni k = ||x^*||_0 < n$ be the optimal value of problem (P_0) and $x^r \in S_r^*$. Then one has*

$$\theta(\min_{x_i^r \neq 0} x_i^r, r) \geq \frac{k}{k+1} \Longrightarrow x^r \in S^*_{||.||_0}. \tag{62}$$

Proof. By Lemma 3.2 and with $\bar{r} = \min_{x_i^r \neq 0} x_i^r / k$ one has

$$\theta(\min_{x_i^r \neq 0} x_i^r, r) \geq \frac{k}{k+1} \Longleftrightarrow r \leq \bar{r}. \tag{63}$$

Then by Lemma 3.1 and using $x_r \in S_r^*$ we have

$$x^r \in \arg\min_{x \in F} \sum_{i=1}^{n} \theta(x_i, r) \Rightarrow \mathrm{card}(I(x_r, r)) \leq k. \tag{64}$$

Finally, using $r \leq \bar{r}$ and that k is the optimal value of problem in ℓ_0-norm we have the result. □

We use in the previous result the minimum non-zero component of x^r, which is logical as we expect that for r sufficiently small the sequence of $\{\min_{x_i^r \neq 0} x_i^r\}_r$ should be increasing. The following lemma gives us a clue on this behavior.

Lemma 3.3. *Consider θ functions where $\theta \geq \theta^1$. Let $x^* \in S^*_{||.||_0}$, $||x^*||_0 = k$ and*

$$r^* = \frac{1}{k} \min_{x_i^* \neq 0} x_i^*.$$

Then one has

$$\forall r \leq r^*, x^r \in S_r^* \Longrightarrow \min_{x_i^r \neq 0} \; x_i^r \leq \min_{x_i^* \neq 0} \; x_i^*. \tag{65}$$

Proof. Suppose that $\min\limits_{x_i \neq 0} \; x_i > \min\limits_{x_i^* \neq 0} \; x_i^*$. Since $x^r \in S_r^*$ we have

$$\begin{aligned}
\sum_{i=1}^{n} \theta(x^r, r) &\geq \sum_{i=1}^{n} \theta(x^r, r^*) && (66)\\
&> (k+1)\theta(\min_{x_i^* \neq 0} \; x_i^*, r^*) && (67)\\
&> (k+1)\frac{kr^*}{kr^* + r^*} && (68)\\
&= k, && (69)
\end{aligned}$$

which is in contradiction with the definition of x^r. □

4 Algorithm

The previous results allow us to build a generic algorithm

$$[\text{Thetal0}] \begin{cases} \{r^k\}_{k \in \mathbb{N}}, \; r^0 > 0 \text{ and } \lim\limits_{k \to +\infty} r^k = 0, \\ \text{find } x^k : \; x^k \in \arg\min\limits_{x \in F} \sum_{i=1}^{n} \theta(x_i, r^k). \end{cases} \tag{70}$$

Now, several questions remain to be answered such as initialization, choice of the sequence $\{r^k\}$ and the method used to solve the concave minimization problems. In Section 3, we have shown an exact penalization result, which help us building a stopping criterion. We make a few remarks about these questions. Note that interesting related remarks can be found in [48].

Remark 4.1 (On the behavior of θ functions)**.** *These concave functions are acting as step function for r sufficiently small, i.e.*

$$\theta(t,r) \simeq \begin{cases} 1 \text{ if } t >> r \\ 0 \text{ if } t << r \end{cases} . \tag{71}$$

This gives a strategy to update r. Let x^k be our current iterate and r^k the corresponding parameter. We divide our iterate into two sets, those with indices in $I = \{i \mid x_i^k \geq r^k\}$ and the others with indices in $\bar{I} = \{i \mid x_i^k < r^k\}$. We can see I as the set of indices of the "non-zero" components and $\bar{I}$ as the set of indices of the "zero" components of x^k. So we will choose r^{k+1} around $\max_{i \in \bar{I}} x_i^k$ to ask whether or not it belongs to zeros and we repeat this operation until r is sufficiently small to consider $\bar{I}$ the set of effective zeros. Also this is a general behavior, to be sure to have decrease of r one should add a fixed parameter of minimum decrease.

Remark 4.2 (Initialization)**.** *It is the main purpose of our method to start with the solution x^0 of the problem (P_1), which is a convex problem. So, we need to find the r^0 related to x^0. A natural, but non-trivial, way of doing this would be to find the parameter which minimizes the following problem*

$$\min_{r>0} || \sum_{i=1}^{N} \theta(x_i^0, r) - ||x^0||_1 \ ||_2^2. \tag{72}$$

A simpler idea is to be inspired from last remark and put r^0 as a value which is just beyond the top value of x_i^0.

Remark 4.3 (Stopping criterion)**.** *It has been shown, in Section 3, an exact penalization theorem using the quantity $k/(k+1)$, which depends on the solution we are looking for. Numerically, we can make more iterations but being sure to satisfy this criterion using the fact that $||x^0||_0 \geq k$, which gives the following criterion*

$$\theta(\min_{x_i^r \neq 0} x_i^r, r) \geq \frac{||x^0||_0}{||x^0||_0 + 1} \geq \frac{k}{k+1}. \tag{73}$$

Remark 4.4 (Algorithm for concave minimization)**.** *In the same way as in [32] and [53] we will use a successive linearization algorithm (SLA) algorithm to solve the concave minimization problem at each iteration in r. This algorithm is a finitely timestep Franck & Wolf algorithm, [45].*

Proposition 4.1 (SLA algorithm for concave minimization)**.** *Given ϵ sufficiently small and r^k. We know x^k and we find x^{k+1} as a solution of the linear* *problem*

$$\min_{x\in F}\; x^t \nabla_x \theta(x^k, r^k), \tag{74}$$

with x^0 a solution of the problem (P_1). We stop when

$$x^{k+1} \in F \text{ and } (x^{k+1} - x^k)^t \nabla_x \theta(x^k, r^k) \le \epsilon. \tag{75}$$

This algorithm generates a finite sequence with strictly decreasing objective function values.

Proof. see [[45], Theorem 4.2]. □

We note that this algorithm did not necessarily provide a global optimum as it ends in a local solution, so we do not expect global solutions in our algorithm. Besides, the gradient of functions θ in the objective tends to be very large as $\theta_r'(t) \approx O(1/r)$, so it can be numerically efficient to add a scaling parameter of order r.

5 Numerical Simulations

Thanks to the previous sections, we have keys for an algorithm. We now present some numerical results. These simulations have been done using MATLAB language, [47], with the linear programming solver GUROBI, [36].

The precision in our simulations is $\epsilon = 10^{-8}$. We generate various polyhedron $F = \{x \in \mathbb{R}^n |\; b \in \mathbb{R}^m,\; Ax \le b\} \cap \mathbb{R}^n_+$ with $m < n$. In the same way as in [32] we choose $n = (500, 750, 1000)$ and in each case $m = (40\%, 60\%, 80\%)$. For each pair (n,m), we choose randomly one hundred problems. We take a random matrix A of size $m \times n$ and a default sparse solution x_{init} with 10% of non-zero components. We get b by computing the matrix-vector product $b = Ax_{init}$. Finally, we compare the sparsity of the solution from Thetal0-algorithm using θ^1 ($\#\theta^1$), the default sparse solution ($\#\ell_0$) and the initial iterate ($\#\ell_1$). We get the initial iterate as a solution of problem (P_1). The item # indicates the number of non-zero components in a vector.

Results are sum up in Table 1. The first two columns give the dimensions of the problems. Column 3 gives the number of problems, where the solution of Thetal0-algorithm has at least the same sparsity as the default sparse solution. In the same vein, Column 4 compares the sparsity of the solution in ℓ_1-norm with the default sparse solution. Column 5 gives the number of problems where the solution by Thetal0-algorithm improves strictly the solution by ℓ_1-norm.

These results validate our algorithm, as in the majority of the cases it manages to find at least an equivalent solution to the default sparse solution. One may notice that in many cases the ℓ_1-norm minimization solution solves the problem in ℓ_0-norm, which is not surprising according to [22].

In Figure 1, we show the behavior of the minimum non-zero component of the current iterate along the iterations in r for one example. We can see the increasing behavior that is the general behavior expected in the Remark 4.3.

Table 1: Numerical results with random $F = \{x \in \mathbb{R}^n_+ | \ b \in \mathbb{R}^m, \ Ax \leq b\}$, dimensions of the problem are first 2 columns. Compare a default sparse solution with 10% of non-zero components, $\#l_0$, the initial iterate solution of (P_1), $\#l_1$, and the solution by θ-algorithm with function θ^1, $\#\theta^1$. The item # indicates the number of non-zeros.

n	m	$\#\ell_0 \geq \#\theta^1$	$\#\ell_1 \leq \#\ell_0$	$\#\theta^1 < \#\ell_1$
1000	800	100	100	0
1000	600	100	98	2
1000	400	50	1	99
750	600	100	100	0
750	450	100	98	2
750	300	54	0	100
500	400	100	100	0
500	300	100	94	6
500	200	63	0	100

6 Conclusion and Outlook

We proposed a class of heuristic schemes to solve the NP-hard problem of minimizing the ℓ_0-norm. Our method requires finding a sequence of solutions from concave minimization problems, which we solved with a successive linearization algorithm. These methods have the benefit that they can only improve the solution we get by solving the ℓ_1-norm problem. We gave an existence result, convergence results, an exact penalization theorem, and keys to implement the methods. To confirm the validity of this algorithm we gave numerical results from randomly generated problems.

Further studies can investigate the special case where the ℓ_1-norm solves the ℓ_0-norm problem, to find an improved stopping condition. Thanks to several studies, for instance [22], we have criteria which can help us identify the cases where the solution we get by solving (P_1) is an optimal solution of (P_0). We wonder if there exists a better sufficient condition than the one presented here in the case where $x^r \in S^*_{||.||_1} \cap S^*_r$

We can also study a very similar problem which is the one of minimizing ℓ_0-norm with noise, see for instance [26] or [4], that is

$$(P_{0,\delta}) \ \min ||x||_0 \text{ s.t. } Ax \leq b + \delta. \tag{76}$$

As a first step in this direction we run our heuristic schemes on some perturbed problems. We generate polyhedron in a similar way as in the previous section with noise in $b = \ Ax_{init} + \vartheta$, where ϑ follows $\mathcal{N}(0, \sigma^2 I_{n\times n})$. We build a signal to noise ratio (SNR) for several values of σ^2 from 0.5 to 0,

$$SNR = 20 \log(\frac{||x^*||_2}{||x^* - x_b||_2}), \tag{77}$$

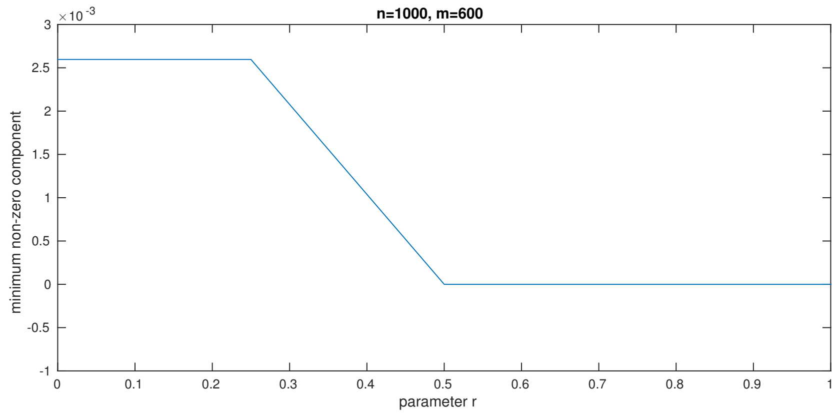

Figure 1: Evolution of the minimum non-zero component of x_r in function of the parameter $r \downarrow 0$.

where x^* and x_b are generated by our algorithm. The former comes from the problem without noise and the later from the perturbed problem. We choose dimensions $n = 500$ and $m = 200$. Then, for one hundred randomly selected problems we compute the mean of the SNR. Results in Figure 2 show very

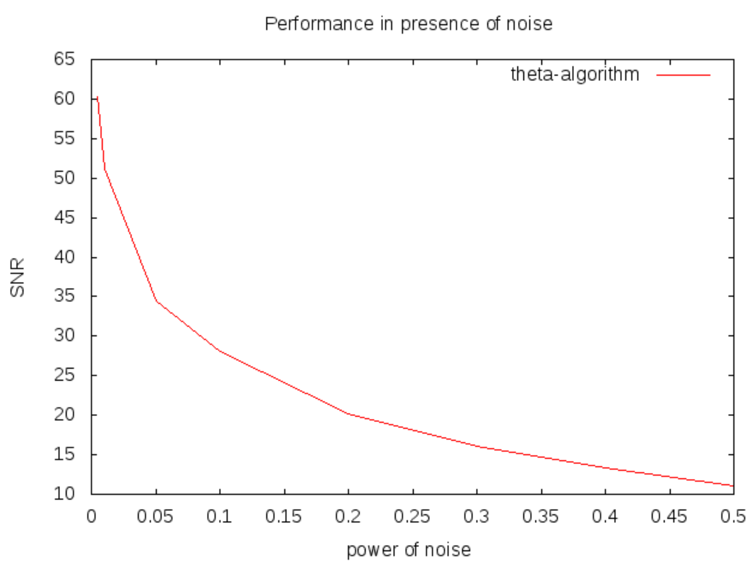

Figure 2: Performance of θ-algorithm in presence of noise, using function $\theta^1(t, r) = t/(t + r)$. $n = 500$, $m = 200$. Mean of SNR for 100 random problems in function of σ^2.

logical behavior as more noises is present more informations are lost. Further work could compare these results with existing methods and shows theoretical study, which could help building an improved algorithm.

References

[1] L. Abdallah, M. Haddou, and T. Migot. Solving absolute value equation using complementarity and smoothing functions. *Journal of Computational and Applied Mathematics*, 327:196–207, 2018.

[2] M. S. Asif and J. Romberg. Sparse recovery of streaming signals using ℓ_1-homotopy. *IEEE Transactions on Signal Processing*, 62(16):4209–4223, 2014.

[3] A. Auslender and M. Teboulle. *Asymptotic cones and functions in optimization and variational inequalities*. Springer Science & Business Media, 2006.

[4] M. Babaie-Zadeh and C. Jutten. On the stable recovery of the sparsest overcomplete representations in presence of noise. *Signal Processing, IEEE Transactions on*, 58(10):5396–5400, 2010.

[5] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[6] M. Bergounioux and M. Haddou. A new relaxation method for a discrete image restoration problem. *Journal of Convex Analysis*, 17(3):421–435, 2010.

[7] S. Bi, X. Liu, and S. Pan. Exact penalty decomposition method for zero-norm minimization based on mpec formulation. *SIAM Journal on Scientific Computing*, 36(4):A1451–A1477, 2014.

[8] T. Blumensath and M. E. Davies. Gradient pursuits. *IEEE Transactions on Signal Processing*, 56(6):2370–2382, 2008.

[9] P. S. Bradley and O. L. Mangasarian. Feature selection via concave minimization and support vector machines. In *ICML*, volume 98, pages 82–90, 1998.

[10] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM review*, 51(1):34–81, 2009.

[11] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, Feb 2006.

[12] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[13] E. J. Candes, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on pure and applied mathematics*, 59(8):1207–1223, 2006.

[14] E. J. Candes and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, Dec 2005.

[15] E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier analysis and applications*, 14(5):877–905, 2008.

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

[17] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.

[18] X. Chen and W. Zhou. Convergence of the reweighted ℓ_1 minimization algorithm for $\ell_2 - \ell_p$ minimization. *Computational Optimization and Applications*, 59(1-2):47–61, 2014.

[19] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *Journal of the American mathematical society*, 22(1):211–231, 2009.

[20] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.

[21] D. L. Donoho. De-noising by soft-thresholding. *IEEE transactions on information theory*, 41(3):613–627, 1995.

[22] D. L. Donoho. Neighborly polytopes and sparse solutions of underdetermined linear equations. Technical report, Department of Statistics, Stanford University, 2004.

[23] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[24] D. L. Donoho. For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. *Communications on pure and applied mathematics*, 59(6):797–829, 2006.

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

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

[27] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001.

[28] Mingbin Feng, John E Mitchell, Jong-Shi Pang, Xin Shen, and Andreas Wächter. Complementarity formulations of l0-norm optimization problems. Technical report, Technical report, Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY, 2013.

[29] S. Foucart and M.-J. 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.

[30] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

[31] J.-J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE transactions on Information theory*, 50(6):1341–1344, 2004.

[32] G. M. Fung and O. L. Mangasarian. Equivalence of minimal ℓ_0 and ℓ_p norm solutions of linear equalities, inequalities and linear programs for sufficiently small p. *Journal of Optimization Theory and Applications*, 151(1):1–10, 2011.

[33] D. Ge, X. Jiang, and Y. Ye. A note on the complexity of l p minimization. *Mathematical programming*, 129(2):285–299, 2011.

[34] I. F. Gorodnitsky, J. S. George, and B. D. Rao. Neuromagnetic source imaging with focuss: a recursive weighted minimum norm algorithm. *Electroencephalography and clinical Neurophysiology*, 95(4):231–251, 1995.

[35] R. Gribonval and M. Nielsen. Sparse decomposition in unions of bases. *IEEE Trans. Inf. Theory*, 49(12):3320–3325, 2003.

[36] Inc. Gurobi Optimization. Gurobi optimizer reference manual, 2015.

[37] M. Haddou and P. Maheux. Smoothing methods for nonlinear complementarity problems. *Journal of Optimization Theory and Applications*, 160(3):711–729, 2014.

[38] M.-J. Lai and J. Wang. An unconstrained ℓ_q minimization with $0 \leq q \leq 1$ for sparse solution of underdetermined linear systems. *SIAM Journal on Optimization*, 21(1):82–101, 2011.

[39] H. A. Le Thi, T. P. Dinh, H. M. Le, and X. T. Vo. Dc approximation approaches for sparse optimization. *European Journal of Operational Research*, 244(1):26–46, 2015.

[40] G. Liuzzi and F. Rinaldi. Solving \ℓ_0-penalized problems with simple constraints via the frank–wolfe reduced dimension method. *Optimization Letters*, 9(1):57–74, Jan 2015.

[41] D. Di Lorenzo, G. Liuzzi, F. Rinaldi, F. Schoen, and M. Sciandrone. A concave optimization-based approach for sparse portfolio selection. *Optimization Methods and Software*, 27(6):983–1000, 2012.

[42] D. Malioutov and A. Aravkin. Iterative log thresholding. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 7198–7202. IEEE, 2014.

[43] S. Mallat. *A wavelet tour of signal processing*. Academic press, 1999.

[44] O. L. Mangasarian. Machine learning via polyhedral concave minimization. In *Applied Mathematics and Parallel Computing*, pages 175–188. Springer, 1996.

[45] O. L. Mangasarian. Machine learning via polyhedral concave minimization. In *Applied Mathematics and Parallel Computing*, pages 175–188. Springer, 1996.

[46] O. L. Mangasarian. Minimum-support solutions of polyhedral concave programs. *Optimization*, 45(1-4):149–162, 1999.

[47] MATLAB. *version 8.4.0.150421 (R2014b)*. The MathWorks Inc., Natick, Massachusetts, 2014.

[48] H. Mohimani, M. Babaie-Zadeh, and C. Jutten. A fast approach for overcomplete sparse decomposition based on smoothed norm. *Signal Processing, IEEE Transactions on*, 57(1):289–301, 2009.

[49] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234, 1995.

[50] D. Needell. Noisy signal recovery via iterative reweighted l1-minimization. In *Signals, Systems and Computers, 2009 Conference Record of the Forty-Third Asilomar Conference on*, pages 113–117. IEEE, 2009.

[51] D. Needell and J. A. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[52] W. B. Pennebaker and J. L. Mitchell. *JPEG: Still image data compression standard*. Springer Science & Business Media, 1992.

[53] F. Rinaldi. *Mathematical programming methods for minimizing the zero-norm over polyhedral sets*. PhD thesis, PhD thesis, Sapienza, University of Rome, 2009.

[54] F. Rinaldi. Concave programming for finding sparse solutions to problems with convex constraints. *Optimization Methods and Software*, 26(6):971–992, 2011.

[55] Francesco Rinaldi, Fabio Schoen, and Marco Sciandrone. Concave programming for minimizing the zero-norm over polyhedral sets. *Computational Optimization and Applications*, 46(3):467–486, 2010.

[56] R. T. Rockafellar. *Convex analysis*. Princeton university press, 1970.

[57] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[58] J. Tropp, S. J. Wright, et al. Computational methods for sparse solution of linear inverse problems. *Proceedings of the IEEE*, 98(6):948–958, 2010.

[59] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information theory*, 50(10):2231–2242, 2004.

[60] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE transactions on information theory*, 52(3):1030–1051, 2006.

[61] V. Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[62] S. Voronin and R. Chartrand. A new generalized thresholding algorithm for inverse problems with sparsity constraints. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 1636–1640. IEEE, 2013.

[63] J. Weston, A. Elisseeff, B. Schölkopf, and M. Tipping. Use of the zero-norm with linear models and kernel methods. *Journal of machine learning research*, 3(Mar):1439–1461, 2003.

[64] D. Wipf and S. Nagarajan. Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):317–329, 2010.

[65] Z. Xie and J. Hu. Rewighted l1-minimization for sparse solutions to underdetermined linear systems. In *Image and Signal Processing (CISP), 2013 6th International Congress on*, volume 3, pages 1660–1664. IEEE, 2013.

[66] Y.-B. Zhao. Rsp-based analysis for sparsest and least ℓ_1-norm solutions to underdetermined linear systems. *IEEE Transactions on Signal Processing*, 61(22):5777–5788, 2013.

[67] Y.-B. Zhao and M. Kočvara. A new computational method for the sparsest solutions to systems of linear equations. *SIAM Journal on Optimization*, 25(2):1110–1134, 2015.

[68] Y.-B. Zhao and Z.-Q. Luo. Constructing new weighted ℓ_1-algorithms for the sparsest points of polyhedral sets. *Mathematics of Operations Research*, 42(1):57–76, 2016.