Journal of Optimization Theory and Applications
https://doi.org/10.1007/s10957-018-1385-3

Maximal Solutions of Sparse Analysis Regularization

Abdessamad Barbara[1] · Abderrahim Jourani[1] · Samuel Vaiter[2]

Received: 30 January 2018 / Accepted: 24 August 2018

Abstract
This paper deals with the non-uniqueness of the solutions of an analysis—Lasso regularization. Most previous works in this area are concerned with the case, where the solution set is a singleton, or to derive guarantees to enforce uniqueness. Our main contribution consists in providing a geometrical interpretation of a solution with a maximal analysis support: such a solution abides in the relative interior of the solution set. Our result allows us to provide a way to exhibit a maximal solution using a primal-dual interior point algorithm.

Keywords Lasso · Analysis sparsity · Uniqueness · Inverse problem · Support identification · Barrier penalization

Mathematics Subject Classification 90C25 · 49J52

1 Introduction

This paper is concerned with solving linear inverse problems with a generalized sparsity constraint. More specifically, we provide a refined study of the solution set of the analysis Lasso.

The linear model is widely used in imaging for degradation such as entry-wise masking, convolution, etc, or in statistics under the name of linear regression. Typically,

---

Communicated by Asen L. Dontchev.

✉ Abderrahim Jourani
jourani@u-bourgogne.fr

Abdessamad Barbara
barbara@u-bourgogne.fr

Samuel Vaiter
vaiter@u-bourgogne.fr

1 Université de Bourgogne Franche-Comté, Dijon, France

2 CNRS, Université de Bourgogne Franche-Comté, Dijon, France

this inverse problem is ill-posed, and one should add additional informations in order to recover at least an approximation of the ground truth. During the last decade, sparse regularization in orthogonal basis has become a classical tool in the analysis of such inverse problems, in particular in imaging [1,2] or in statistics and machine learning [3]. In this work, we consider the more general framework, known as the sparse analysis prior [4,5], cosparse prior [6] or generalized Lasso. The idea is to not measure the sparsity of the coefficients in an orthogonal basis only, but in any dictionary.

Probably the most popular example of analysis sparsity-inducing regularizer is the total variation, which was introduced in [7] in a continuous setting for denoising. In the discrete setting, it corresponds to taking the dictionary as a discretization of a derivative operator. In the context of one-dimensional signals, a popular choice is to take a forward finite difference operator. Other popular choices of dictionary include translation invariant wavelets (which can be viewed as a higher-order total variation following [8]) or the concatenation of a derivative operator with the identity, known under the name of Fused Lasso [9] in statistics.

It is important to keep in mind that the solution set is typically not a singleton. Most previous works in this area are concerned with the case, where the solution set is a singleton, or aim at deriving guarantees to enforce uniqueness. Necessary and sufficient conditions have been derived in [10,11] and also in [12] for the constrained case. In this paper, we tackle the case, where the solution set is not a singleton, and we want to better understand the structure of the solution set in this case. Some insights are given in [13], but their results are limited to the synthesis case. In that work, the authors give a bound on the size of the support and prove that the LARS algorithm converges to a solution with a maximal support. To our knowledge, our work is the first to address this structure in the context of analysis sparsity.

2 Problem Setting

We consider the problem of estimating an unknown vector $x_0 \in \mathbb{R}^n$ from noisy observations

$$y = \Phi x_0 + w \in \mathbb{R}^q, \tag{1}$$

where Φ is a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^q$ and w is the realization of a noise.

The sparsity of some coefficients $x \in \mathbb{R}^n$ is measured by using the counting function, or abusively ℓ^0 norm, which reads

$$\|x\|_0 := \mathrm{Card}(\mathrm{supp}(x)), \text{ where } \mathrm{supp}(x) := \{i \in \{1, \dots, n\} \ : \ x_i \neq 0\}$$

is coined the support of the vector x and Card E denotes the cardinality of the set E. The associated regularization problem to (1) is given by

$$\operatorname*{Argmin}_{x \in \mathbb{R}^n} \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|x\|_0 ,$$

which is known to be NP-hard [14]. One way to alleviate this issue consists in adopting greedy methods, such as the Matching Pursuit [15] or derivation from it such as the

OMP [16], CoSAMP [17]. This will not be the concern of this paper, which focuses on one of its most popular convex relaxation through the ℓ^1-norm. More precisely, we consider the (synthesis) Lasso optimization problem [3]

$$\underset{x \in \mathbb{R}^n}{\operatorname{Argmin}} \ \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|x\|_1 , \tag{2}$$

where the ℓ^1-norm is defined as $\|x\|_1 = \sum_{i=1}^{n} |x_i|$.

In the more general framework of analysis regularization, a dictionary is used to analyze the sparsity pattern of the solution set. Formally, a dictionary D is a linear operator from $\mathbb{R}^p$ to $\mathbb{R}^n$ which is defined through p n-dimensional *atoms* d_i which may be redundant. Using this dictionary, one can build an analysis regularization $\|D^* \cdot\|_1$ associated to the variational framework defined by

$$\mathbf{X}_\lambda := \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}} \ h(x) = \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \left\|D^* x\right\|_1 . \tag{3}$$

When there is no noise, i.e., $y = \Phi x_0$, it is common to use a constrained version of (3) given by

$$\mathbf{X}_0 := \underset{x \in \mathbb{R}^n}{\operatorname{Argmin}} \ \left\|D^* x\right\|_1 \ \text{subject to } \Phi x = y. \tag{4}$$

It has been first introduced in [1] under the name Basis Pursuit for $D = \mathrm{Id}$, and one can easily see that (4) can be recasted as a linear program (LP).

Suppose that someone aims to solve a problem[1] of the form (3), and that running two different algorithm, he obtains two solutions

$$\begin{aligned} x^1 &= \begin{pmatrix} 0 & \frac{1}{4} & \frac{1}{4} & 1 & \dots & 1 \end{pmatrix} \quad \text{such that} \quad D^* x^1 = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \dots 0 \end{pmatrix}, \\ x^2 &= \begin{pmatrix} 0 & 0 & \frac{1}{2} & 1 & \dots & 1 \end{pmatrix} \quad \text{such that} \quad D^* x^2 = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 \dots 0 \end{pmatrix}. \end{aligned}$$

He may be confused by the fact that the first component of $D^* x^2$ is active, but not for $D^* x^1$. Indeed, the sparsity level (with respect to D^*) is different, and without more information on the problem, it is impossible to know whether the first component is relevant or not. This paper takes a worst-case approach:

– How to characterize a solution $x^\star$ such that the support of $D^* x^\star$ is maximal?
– Is it possible to give an algorithm to determine it?

We will provide a geometrical characterization of this specific notion of maximality, and we will show that an interiorpoint method gives a maximal solution, without any constraint on the initialization.

[1] We come back in Sect. 6 to this example.

3 Contributions

In all this paper, we consider the following hypothesis of restricted injectivity

$$\text{Ker } D^* \cap \text{Ker } \Phi = \{0\}, \tag{5}$$

It is not difficult to show that this condition is equivalent to the well-definedness and boundedness of $\mathbf{X}_\lambda$ (it is equivalent to the coercivity of the objective function). In Sect. 4, we review some properties of the solution set. We prove in particular that $\mathbf{X}_\lambda$ is a polytope, i.e., a bounded polyhedron.

Our main contribution is proved in Sect. 5. It consists in providing a geometrical interpretation of a solution with a maximal D-support, namely the fact that such a solution lives in the relative interior of the solution set.

Definition 3.1 A vector $x^+ \in \mathbb{R}^n$ is *a solution of maximal D-support* if x^+ is a solution, i.e., $x^+ \in \mathbf{X}_\lambda$, and for all $x \in \mathbf{X}_\lambda$, $\|D^*x\|_0 \leq \left\|D^*x^+\right\|_0$.

The set of solution of (3) which have maximal D-support will be denoted by $\mathbf{S}_\lambda$.

Relation (5) ensures that this set is well-defined and is contained in $\mathbf{X}_\lambda$.

Now, we are able to give a characterization of solutions with maximal D-support.

Theorem 3.1 *Let $\bar{x} \in \mathbf{X}_\lambda$. Then $\bar{x}$ is a maximally D-supported solution if, and only if, $\bar{x} \in$ ri $\mathbf{X}_\lambda$ (or equivalently if $\bar{x} \in$ ri $\mathbf{S}_\lambda$). In other words,*

$$\mathbf{S}_\lambda = \text{ri } \mathbf{S}_\lambda = \text{ri } \mathbf{X}_\lambda.$$

Thus, our theorem gives not only a geometrical characterization of $\mathbf{S}_\lambda$ but also a topological one namely the relative openness of $\mathbf{S}_\lambda$. We recall that for any set S, the relative interior ri S of S is defined as its interior with respect to the topology of the affine hull of S.

Using this theorem, we provide a way to construct such maximal solutions. In Sect. 6, we show that with the help of the classical barrier penalization, we build a path which converges to a point in the relative interior of $\mathbf{X}_\lambda$. We defer the precise statement of this construction to Sect. 6.

4 The Solution Set

This section reviews some properties of the solution set $\mathbf{X}_\lambda$. The following proposition shows that even if $\mathbf{X}_\lambda$ is not reduced to a singleton, its image by Φ or the analysis-ℓ^1-norm is single-valued.

Proposition 4.1 (Unique image) *Let $x^1, x^2 \in \mathbf{X}_\lambda$. Then,*

1. *They share the same image by Φ, i.e., $\Phi x^1 = \Phi x^2$;*
2. *They have the same analysis-ℓ^1-norm, i.e., $\left\|D^*x^1\right\|_1 = \left\|D^*x^2\right\|_1$.*

A proof of this statement can be found for instance in [5] and Corollary A.2 in "Appendix".

It is known that the standard ℓ^2-regularization suffers from sign inconsistencies, i.e., two different solutions can be of opposite signs at some index. The following proposition gives another important information: the cosign of two solutions cannot be opposite.

Proposition 4.2 (Consistency of the sign) *Let $x^1, x^2 \in \mathbf{X}_\lambda$. Then,*

$$\forall i \in \{1, \dots, p\}, \quad u_i^1 u_i^2 \geq 0,$$

where $u^k = D^* x^k$ for $k = 1, 2$.

Proof The proof of this statement follows closely the proof found in [18] for ℓ^1. Suppose there exists i such that u_i^1 and u_i^2 have opposite signs. Then, one has

$$|u_i^1 + u_i^2| < |u_i^1| + |u_i^2|. \tag{6}$$

Let $z = \frac{1}{2}\left(x^1 + x^2\right)$. Using the convexity of $x \mapsto \|y - \Phi x\|_2^2$ and inequality (6), we get

$$\begin{aligned}
\frac{1}{2}\|y - \Phi z\|_2^2 + \lambda \left\|D^* z\right\|_1 &\leq \frac{1}{2}\left[\frac{1}{2}\left\|y - \Phi x^1\right\|_2^2 + \frac{1}{2}\left\|y - \Phi x^2\right\|_2^2\right] \\
&\quad + \frac{\lambda}{2}\sum_{j=1}^{n}|u_j^1 + u_j^2| \\
&< \frac{1}{2}\left[\frac{1}{2}\left\|y - \Phi x^1\right\|_2^2 + \frac{1}{2}\left\|y - \Phi x^2\right\|_2^2\right] \\
&\quad + \frac{\lambda}{2}\sum_{j=1}^{n}|u_j^1| + \frac{\lambda}{2}\sum_{j=1}^{n}|u_j^2| \\
&= \frac{1}{2}\left(\frac{1}{2}\left\|y - \Phi x^1\right\|_2^2 + \lambda\left\|D^* x^1\right\|_1\right) \\
&\quad + \frac{1}{2}\left(\frac{1}{2}\left\|y - \Phi x^2\right\|_2^2 + \lambda\left\|D^* x^2\right\|_1\right) \\
&= \min_{x \in \mathbb{R}^n} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda\left\|D^* x\right\|_1,
\end{aligned}$$

which is a contradiction. □

As it is said in the beginning of this section, condition (5) ensures non-emptiness, convexity and compactness of $\mathbf{X}_\lambda$. In fact, as stated in the following proposition, the solution set $\mathbf{X}_\lambda$ is a polytope.

Proposition 4.3 $\mathbf{X}_\lambda$ *is a polytope (i.e., a bounded polyhedron).*

Proof It is a consequence of Proposition A.1 in "Appendix". □

Owing to Proposition 4.3, we can rewrite the set $\mathbf{X}_\lambda$ as the convex hull of k points in $\mathbb{R}^n$ as

$$\mathbf{X}_\lambda = \operatorname{conv}\{a_1, \ldots, a_k\},$$

where a_i are the extremal points of $\mathbf{X}_\lambda$. Observe that each a_i lives on the boundary of the analysis-ℓ^1-ball of radius $\|D^*\bar{x}\|_1$ where $\bar{x}$ is any element of $\mathbf{X}_\lambda$. Naturally, we can even rewrite the solution as

$$\mathbf{X}_\lambda = A\Delta_k = \{Az \ : \ z \in \Delta_k\},$$

where A is a matrix $n \times k$ such that its columns are the vectors a_i and the n-simplex Δ_n of $\mathbb{R}^n$ is defined as

$$\Delta_n := \left\{ x \in \mathbb{R}^n \ : \ \sum_{i=1}^n x_i = 1 \text{ and } \forall i, x_i \geq 0 \right\} = \operatorname{conv}\{e_1, \ldots, e_n\},$$

where $(e_1, \ldots, e_n)$ is the canonical basis of $\mathbb{R}^n$. Since a_i are the extremal points of $\mathbf{X}_\lambda$, the matrix A has a maximal rank. Observe in particular that the lines of the matrix D^*A have same signs according to Proposition 4.2.

5 Maximal Support and Proof of Theorem 3.1

The following proposition proves that *the* D-maximal support is indeed uniquely defined.

Proposition 5.1 *Let $x \in \mathbf{X}_\lambda$. Then the two following assertions are equivalent.*

1. *x is a solution of maximal D-support, i.e., $x \in \mathbf{S}_\lambda$.*
2. *For any $\bar{x} \in \mathbf{X}_\lambda$,* $\operatorname{supp}(D^*\bar{x}) \subseteq \operatorname{supp}(D^*x)$.

In particular, two solutions of maximal support share the same D-support.

Proof The two directions are proved separately.

1. $\Rightarrow$ 2.: Suppose there exists $i_0 \in \{1, \ldots, p\}$ such that $i_0 \in \operatorname{supp}(D^*\bar{x})$ and $i_0 \notin \operatorname{supp}(D^*x)$. Observe that $\tilde{x} = \frac{1}{2}(\bar{x} + x)$ is also an element of $\mathbf{X}_\lambda$ by convexity of $\mathbf{X}_\lambda$. Using Proposition 4.2, we get

$$\operatorname{supp}(D^*\tilde{x}) \supseteq \operatorname{supp}(D^*\bar{x}) \cup \operatorname{supp}(D^*x).$$

In particular, $\operatorname{supp}(D^*\tilde{x}) \supseteq \operatorname{supp}(D^*x) \cup \{i_0\} \supsetneq \operatorname{supp}(D^*x)$. Hence,

$$\operatorname{Card}(\operatorname{supp}(D^*\tilde{x})) > \operatorname{Card}(\operatorname{supp}(D^*x)),$$

which contradicts the fact that x has maximal D-support.

2. $\Rightarrow$ 1.: This implication is obvious. □

For $x \in \mathbf{S}_\lambda$, we set

$$m = \mathrm{Card}(\mathrm{supp}(D^* x)). \tag{7}$$

Note that, by definition of $\mathbf{S}_\lambda$, m does not depend on the choice of $x \in \mathbf{S}_\lambda$ but only on $\mathbf{S}_\lambda$.

We start by a technical corollary of Proposition 4.2.

Corollary 5.1 *Let x^+ be an element of $\mathbf{S}_\lambda$. Let $\tilde{\Sigma}$ be the permutation matrix associated to a permutation $\tilde{\sigma}$ which maps $\mathrm{supp}(D^* x^+)$ to $\{1, \dots, m\}$. Define the permutation $\sigma : \{1, \dots, p\} \to \{1, \dots, p\}$ by*

$$\sigma(i) := \begin{cases} \tilde{\sigma}(i), & \text{if } i \in \{1, \dots, m\}, \\ i, & \text{if } i \in \{m+1, \dots, p\}. \end{cases}$$

Let Σ be the permutation matrix associated to σ. Then there exists a matrix $\Lambda = \mathrm{diag}(\lambda_i)_{i=1,\dots,p}$ with $\lambda_i \in \{-1, 1\}$ for $i \in \{1, \dots, m\}$ and $\lambda_i = 0$ for $i \in \{m+1, \dots, p\}$ such that for $\Gamma = \Lambda\Sigma$, one has

$$\Gamma D^* \mathbf{X}_\lambda \subset (\mathbb{R}_+)^m \times \{0\}^{p-m}.$$

Moreover, for all $x \in \mathrm{aff}\, \mathbf{X}_\lambda$, $\|\Gamma D^* x\|_1 = \|D^* x\|_1$.

Proof Define the matrix Λ by its diagonal by

$$\forall i \in \{1, \dots, m\}, \quad \lambda_i := \begin{cases} 1, & \text{if } (D^* x^+)_{\sigma^{-1}(i)} > 0, \\ -1, & \text{if } (D^* x^+)_{\sigma^{-1}(i)} < 0, \end{cases}$$

and $\lambda_i = 0$ for $i \in \{m+1, \dots, p\}$.

Now, take any solution $x \in \mathbf{X}_\lambda$ and consider the vector $u = \Gamma D^* x$. Let $i \in \{1, \dots, m\}$, then

$$u_i = \left\langle e_i,\ \Lambda\Sigma D^* x \right\rangle.$$

Since Λ is self-adjoint, one has

$$u_i = \left\langle \Lambda e_i,\ \Sigma D^* x \right\rangle.$$

As Λ is diagonal, we get

$$u_i = \lambda_i \left\langle e_i,\ \Sigma D^* x \right\rangle.$$

Using the fact that Σ is a permutation matrix, we have $\Sigma^* = \Sigma^{-1}$, that is,

$$u_i = \lambda_i \left\langle \Sigma^{-1} e_i,\ D^* x \right\rangle.$$

The permutation σ associated to Σ leads to

$$u_i = \lambda_i \left\langle e_{\sigma^{-1}(i)},\ D^* x \right\rangle = \lambda_i (D^* x)_{\sigma^{-1}(i)}.$$

According to Proposition 4.2, one has $(D^*x)_{\sigma^{-1}(i)}(D^*x^+)_{\sigma^{-1}(i)} \geq 0$. Moreover, λ_i has the same sign as $(D^*x^+)_{\sigma^{-1}(i)}$. Thus, $u_i = \lambda_i(D^*x)_{\sigma^{-1}(i)} \geq 0$. For $i \in \{m+1, \dots, p\}$, $\lambda_i = 0$ and hence $u_i = 0$.

The equality $\|\Gamma D^*x\|_1 = \|D^*x\|_1$, with $x \in \mathbf{X}_\lambda$, follows from the inclusion $\operatorname{supp}(D^*x) \subseteq \operatorname{supp}(D^*x^+)$ and the fact that the ℓ^1-norm is invariant by permutation and change of signs. □

Note that the matrix Λ and Σ are not uniquely defined. Corollary 5.1 allows us to work only on $m-$dimensional positive vectors.

We will also need to exclude at some point the case where a solution x lives in the kernel of D^*. The following lemma shows that if this is the case, then the solution set $\mathbf{X}_\lambda$ is a singleton.

Lemma 5.1 *If* Ker $D^* \cap \mathbf{X}_\lambda \neq \emptyset$*, then* $\mathbf{X}_\lambda$ *is a singleton.*

Proof Let $x \in \operatorname{Ker} D^* \cap \mathbf{X}_\lambda$. We recall that $\mathbf{X}_\lambda \subset x + \operatorname{Ker} \Phi$. Pick $\bar{x} \in \mathbf{X}_\lambda$, and rewrite it as $\bar{x} = x + h$ where $h \in \operatorname{Ker} \Phi$. Then, according to Proposition 4.1, one has $\|D^*\bar{x}\|_1 = \|D^*x\|_1 = 0$. In particular, $\|D^*x + D^*h\|_1 = \|D^*h\|_1 = 0$. Using hypothesis (5), we get $h = 0$. □

We can now provide the proof of Theorem 3.1.

Proof of Theorem 3.1 We exclude here the case where $\mathbf{X}_\lambda$ is a singleton, since the result is then trivially verified. Let us prove both direction separately.

($\Leftarrow$: ri $\mathbf{X}_\lambda \subseteq \mathbf{S}_\lambda$). First, we recall that ri $\mathbf{X}_\lambda = \operatorname{ri}(A\Delta_k) = A \operatorname{ri} \Delta_k$. Let $\bar{x} \in \operatorname{ri} \mathbf{X}_\lambda$. We have

$$\bar{x} = A\bar{z} \text{ with } \sum_{i=1}^{k} \bar{z}_i = 1 \text{ and } \bar{z}_i > 0.$$

Given $i \in \{1, \dots, m\}$ we shall prove $(\Gamma D^*\bar{x})_i \neq 0$. So suppose the contrary, i.e., $(\Gamma D^*\bar{x})_i = 0$. Then we have

$$(\Gamma D^*\bar{x})_i = (\Gamma D^*A\bar{z})_i = \left\langle e_i,\ \Gamma D^*A\bar{z}\right\rangle = \left\langle e_i,\ \Lambda\Sigma D^*A\bar{z}\right\rangle,$$

where x^+, m, Γ and Σ are given by Corollary 5.1. Using the fact that Λ is a diagonal matrix and Σ is a permutation matrix, we obtain

$$(\Gamma D^*\bar{x})_i = \lambda_i \left\langle D\Sigma^{-1}e_i,\ A\bar{z}\right\rangle.$$

As $\Sigma^{-1}e_i = e_{\sigma^{-1}(i)}$, where σ is the permutation associated to Σ, we get

$$0 = (\Gamma D^*\bar{x})_i = \lambda_i \left\langle (D^*A)^* e_{\sigma^{-1}(i)},\ \bar{z}\right\rangle.$$

By construction of Λ, we have $\lambda_i \in \{-1, 1\}$ and hence

$$\left\langle (D^*A)^* e_{\sigma^{-1}(i)},\ \bar{z}\right\rangle = 0.$$

By definition of $\bar{z}$, $\bar{z}_j > 0, \forall j \in \{1, \ldots, k\}$, and according to Proposition 4.2, we necessarily have

$$0 = \left((D^*A)^* e_{\sigma^{-1}(i)}\right)_j = \left\langle (D^*A)^* e_{\sigma^{-1}(i)},\ e_j \right\rangle = \left\langle e_{\sigma^{-1}(i)},\ D^* a_j \right\rangle,$$

for any extremal point a_j of $\mathbf{X}_\lambda$. Thus $\left\langle e_{\sigma^{-1}(i)},\ D^*x \right\rangle = 0, \forall x \in \mathbf{S}_\lambda (\subset \mathbf{X}_\lambda)$, which contradicts the fact that i belongs to $\{1, \ldots, m\}$. Thus $(\varGamma D^*\bar{x})_i \neq 0$ and this asserts that $\sigma^{-1}(i) \in \operatorname{supp}(D^*\bar{x})$ because $(\varGamma D^*\bar{x})_i = \lambda_i (D^*\bar{x})_{\sigma^{-1}(i)}$.

Hence, $\operatorname{supp}(D^*\bar{x}) \supseteq \operatorname{supp}(D^*x^+)$ and then $\bar{x} \in \mathbf{S}_\lambda$.

($\Rightarrow$: $\mathbf{S}_\lambda \subseteq \operatorname{ri} \mathbf{X}_\lambda$). We are going to prove that $\mathbf{S}_\lambda = \operatorname{ri} \mathbf{S}_\lambda$. Indeed, according to ($\Leftarrow$), $\operatorname{ri} \mathbf{X}_\lambda \subseteq \mathbf{S}_\lambda$. Moreover, since every element of $\mathbf{S}_\lambda$ is also an element of $\mathbf{X}_\lambda$, we have $\operatorname{ri} \mathbf{X}_\lambda \subseteq \mathbf{S}_\lambda \subseteq \mathbf{X}_\lambda$. In particular, $\operatorname{aff} \mathbf{X}_\lambda = \operatorname{aff} \mathbf{S}_\lambda$.
Let $x^+ \in \mathbf{S}_\lambda$ and let $\varGamma$ be its associated matrix (see Corollary 5.1). Put

$$\alpha = \min_{i \in \operatorname{supp}(D^*x^+)} |(D^*x^+)_i| = \min_{i \in \{1,\ldots,m\}} (\varGamma D^*x^+)_i.$$

Since $\mathbf{X}_\lambda$ is not reduced to a singleton and non-empty, then $\operatorname{supp}(D^*x^+)$ has cardinal greater than 1, hence $\alpha > 0$.

Now take any $u \in B_\infty(x^+, r) \cap \operatorname{aff} \mathbf{X}_\lambda$ where $B_\infty(x^+, r)$ is the ℓ^∞-ball centered at x^+ with radius r defined by

$$r = \frac{\alpha - \epsilon}{\|\varGamma D^*\|_{\infty,\infty}} \text{ and } \left\|\varGamma D^*\right\|_{\infty,\infty} = \max_{\|z\|_\infty \le 1} \left\|\varGamma D^* z\right\|_\infty,$$

for $\epsilon \in]0, \alpha[$.

Let's prove first that $\varGamma D^*u \in (\mathbb{R}_+^*)^m \times \{0\}^{p-m}$. From the definition of u, we get

$$\left\|\varGamma D^*u - \varGamma D^*x^+\right\|_\infty \le \left\|\varGamma D^*\right\|_{\infty,\infty} \left\|u - x^+\right\|_\infty \le \alpha - \epsilon.$$

For $i \in \{1, \ldots, m\}$, one has $|(\varGamma D^*u)_i - (\varGamma D^*x^+)_i| \le \alpha - \epsilon$, in particular

$$(\varGamma D^*u)_i - (\varGamma D^*x^+)_i \ge -\alpha + \epsilon \Leftrightarrow (\varGamma D^*u)_i \ge (\varGamma D^*x^+)_i - \alpha + \epsilon.$$

Since $(\varGamma D^*x^+)_i - \alpha \ge 0$ and $\epsilon > 0$, we conclude that $(\varGamma D^*u)_i > 0$. Thus, $(\varGamma D^*u)_i > 0$ for $i \in \{1, \ldots, m\}$ and $(\varGamma D^*u)_i = 0$ for $i \notin \{1, \ldots, m\}$.

It remains to prove that u is a solution of (3), i.e., $u \in \mathbf{X}_\lambda$. Since $u \in \operatorname{aff} \mathbf{X}_\lambda$, there exist $t \in \mathbb{R}$ and $x \in \mathbf{X}_\lambda$ such that

$$u = x^+ + t(x - x^+).$$

From this equality, we get

$$\left\|D^*u\right\|_1 = \left\|\varGamma D^*u\right\|_1 = \sum_{i=1}^{p} (\varGamma D^*u)_i \qquad \text{according to Corollary 5.1}$$

$$\begin{aligned}
&= \sum_{i=1}^{p} (1-t)(\Gamma D^* x^+)_i + t(\Gamma D^* x)_i \\
&= (1-t) \left\| \Gamma D^* x^+ \right\|_1 + t \left\| \Gamma D^* x \right\|_1 \\
&= \left\| D^* x^+ \right\|_1 \qquad\qquad \text{because } \left\| D^* x^+ \right\|_1 = \left\| D^* x \right\|_1 .
\end{aligned}$$

Moreover, $\Phi u = \Phi x^+ + t(\Phi x - \Phi x^+) = \Phi x^+$. Thus, u is a solution and the proof is completed. □

6 Finding a Maximal Solution

Using the classical barrier function, in this section we show how to get a path that converges to a relative interior point of $\mathbf{X}_\lambda$, which turns out to be the analytic center of $\mathbf{X}_\lambda$.

6.1 A Barrier Approach

Setting $Q = \Phi^* \Phi$ being the Gram matrix and $c = \Phi^* y$, we start by rewriting our initial problem Equation (3) as an augmented quadratic program under constraints, i.e.,

$$\min_{x \in \mathbb{R}^n, t \in \mathbb{R}^p} \frac{1}{2} \langle Qx,\ x \rangle - \langle c,\ x \rangle + \lambda \sum_{i=1}^{p} t_i \text{ subject to } \begin{cases} -t \leq D^* x \leq t \\ t \geq 0 \end{cases},$$

witch also can be rewritten as

$$\min_{x \in \mathbb{R}^n, t \in \mathbb{R}^p} \frac{1}{2} \langle Qx,\ x \rangle - \langle c,\ x \rangle + \lambda \sum_{i=1}^{p} t_i \text{ subject to } \begin{cases} -t + s = D^* x \\ t - s' = D^* x \\ t \geq 0,\ s \geq 0,\ s' \geq 0 \end{cases}.$$

Now observe that $t = \frac{1}{2}(s + s')$. Then setting $z = \frac{1}{2}\begin{pmatrix} s \\ s' \end{pmatrix}$, I_p the p by p identity matrix, $\tilde{I} = \begin{pmatrix} I_p & -I_p \end{pmatrix}$ and $e = (1, \cdots, 1) \in \mathbb{R}^{2p}$, we come to the following equivalent formulation of the problem

$$\min_{x \in \mathbb{R}^n, z \in \mathbb{R}^{2p}} f(x, z) \text{ subject to } z \in [0, +\infty[^{2p} \tag{8}$$

where

$$f(x, z) = \begin{cases} \frac{1}{2} \langle Qx,\ x \rangle - \langle c,\ x \rangle + \lambda \langle e,\ z \rangle, & \text{if } D^* x + \tilde{I} z = 0, \\ +\infty, & \text{elsewhere,} \end{cases}$$

or equivalently

$$f(x,z) = \begin{cases} \frac{1}{2}\|\Phi x - y\|^2 - \frac{1}{2}\|y\|^2 + \lambda \langle e, z\rangle, & \text{if } D^* x + \tilde{I} z = 0, \\ +\infty, & \text{elsewhere.} \end{cases}$$

Its classical dual is

$$\max_{x\in\mathbb{R}^n, s\in\mathbb{R}^{2p}, u\in\mathbb{R}^p} g(x,s,u) \text{ subject to } s \in [0,+\infty[^{2p} \tag{9}$$

where

$$g(x,s,u) = \begin{cases} -\frac{1}{2}\langle Qx, x\rangle, & \text{if } Du + c - Qx = 0,\ s = \lambda e - \tilde{I}^* u, \\ -\infty, & \text{elsewhere.} \end{cases}$$

We set $S_{(P)}$ [resp. $S_{(D)}$] the optimal solution set of problem (8) [resp. problem (9)]. We know that $\mathbf{X}_\lambda$ is non-empty and so $S_{(P)}$. Since, in addition (8) is a convex problem with polyhedral constraints, $S_{(D)}$ is non-empty and there is no duality gap. We denote by α the optimal value of the two problems and recall for all the following that given a lower semicontinuous real-valued extended convex function h on $\mathbb{R}^l$, its recession function can be defined by (Theorem 8.5 of [19])

$$h_\infty(d) := \lim_{\lambda\uparrow+\infty} \frac{h(z+\lambda d) - h(z)}{\lambda}, \ \forall (z,d) \in \mathrm{dom}(h) \times \mathbb{R}^l.$$

Proposition 6.1 1. *The optimal solution $S_{(P)}$ of the problem (8) is bounded or equivalently the set $\{(d_x, d_z) :\ f_\infty(d_x, d_z) \le 0,\ d_z \ge 0\} = \{0\}$,*
2. *$S(., (D)) = \{(s,u) :\ \exists x \in \mathbb{R}^n$ such that $(x,s,u) \in S_{(D)}\}$ is bounded, in other words, the dual feasible solution set is bounded in (s,u).*

Proof 1. Because of relation (5), it is not difficult to show that the optimal solution $S_{(P)}$ of the problem (8) is bounded.

2. Let (x^k, s^k, u^k) be a sequence of the dual feasible solution set. We have $s^k = \lambda e - \tilde{I}^* u = \begin{pmatrix} \lambda e^p \\ \lambda e^p \end{pmatrix} - \begin{pmatrix} u^k \\ -u^k \end{pmatrix} \ge 0$, where $e^p = (1,\ldots 1) \in \mathbb{R}^p$. It follows that $-\lambda e^p \le u^k \le \lambda e^p$. Hence (u^k) and then (s^k) are bounded. □

Using the classical logarithmic barrier function introduced by Frish [20], we deal with the family of problems $(P_\mu)_{\mu>0}$ given by

$$\theta(\mu) = \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^{2p}} F_\mu(x,z) = f(x,z) + \zeta(z,\mu),$$

where

$$\zeta(z,\mu)=\begin{cases}\mu\xi\left(z/\mu\right), & \text{if } \mu>0,\\ \xi_\infty(z), & \text{if } \mu=0,\\ +\infty, & \text{elsewhere,}\end{cases}$$

$$\xi(z)=\begin{cases}-\ln\varphi(z), & \text{if } \varphi(z)>0,\\ +\infty, & \text{elsewhere,}\end{cases} \quad \text{and } \varphi(z)=\begin{cases}\left(\prod_{i=1}^{2p} z_i\right)^{\frac{1}{2p}}, & \text{if } z\ge 0,\\ -\infty, & \text{elsewhere.}\end{cases}$$

Note that the function φ is strictly quasiconcave and then according to Lemma 1 of [21], for every $\mu>0$, the function $\zeta_\mu : z\mapsto \zeta(z,\mu)$ is strictly convex on $]0,+\infty[^{2p}$.

Proposition 6.2 *For every $\mu>0$, the function F_μ is inf-compact on $\mathbb{R}^n\times\mathbb{R}^{2p}$ and strictly convex on $\mathbb{R}^n\times]0,+\infty[^{2p}$.*

Proof Let us show that

$$\xi_\infty(d)=\begin{cases}0 & \text{if } d\ge 0,\\ +\infty & \text{elsewhere.}\end{cases} \tag{10}$$

Let $(z,d)\in \mathrm{dom}(\xi)\times\mathbb{R}^{2p}$. We have necessarily $z>0$. First, we observe that when $d\notin[0,+\infty[^{2p}$, $z+\lambda d\notin[0,+\infty[^{2p}$ for λ large enough and then $\xi_\infty(d)=+\infty$. Now, consider the case $d\ge 0$. Since $z>0$ we have necessarily $z+d>0$. The concave gauge function φ is monotone with respect to its domain, the positive orthant. Then, Proposition 2.1 of [22],

$$0<\varphi(z+d)\le\varphi(z+\lambda d)\le\varphi(\lambda z+\lambda d)=\lambda\varphi(z+d)$$

for λ large enough. It follows that

$$\begin{aligned}0&=\lim_{\lambda\uparrow+\infty}\frac{\ln\varphi(z+d)-\ln\varphi(z)}{\lambda}\le\lim_{\lambda\uparrow+\infty}\frac{\ln\varphi(z+\lambda d)-\ln\varphi(z)}{\lambda}\\&\le\lim_{\lambda\uparrow+\infty}\frac{\ln\lambda\varphi(z+d)-\ln\varphi(z)}{\lambda}=0,\end{aligned}$$

and hence $\lim_{\lambda\uparrow+\infty}\dfrac{\ln\varphi(z+\lambda d)-\ln\varphi(z)}{\lambda}=0$. Consequently $\xi_\infty(d)=0$.

By Proposition 6.1, we have $\{(d_x,d_z):\ f_\infty(d_x,d_z)\le 0,\ d_z\ge 0\}=\{(0,0)\}$. Thus $\{(d_x,d_z):\ F_{\mu_\infty}(d_x,d_z)\le 0,\ d_z\ge 0\}=\{(0,0)\}$, or equivalently, F_μ is inf-compact.

Now let us proceed to prove the strict convexity of F_μ. Take $(x,z)\ne(x',z')$ in $\mathbb{R}^n\times]0,+\infty[^{2p}$ and $t\in]0,1[$. In the case where $z\ne z'$, by strict convexity of ζ_μ on $]0,+\infty[^{2p}$ we have necessarily $F_\mu(t(x,z)+(1-t)(x',z'))<tF_\mu(x,z)+(1-t)F_\mu(x',z')$. Assume that $z=z'$. Using (5) and the definition of f, we obtain $\Phi x\ne\Phi x'$ and the result follows by using the strict convexity of $\|.\|_2^2$. $\square$

Propositions 6.2 and 6.1 assert that for every $\mu > 0$ there is a unique optimal solution $(x(\mu), z(\mu))$ to (P_μ). Moreover using the fact that $F_\mu(x, \cdot)$ is a barrier function for every $x \in \mathbb{R}^n$, $z(\mu) > 0$. Consider the function $\gamma : \mathbb{R}^n \times [0, +\infty[^{2p} \times [0, +\infty[\to \mathbb{R} \cup \{+\infty\}$ defined by

$$\gamma(x, z, \mu) = F_\mu(x, z).$$

Then we have the following proposition.

Proposition 6.3 *The function γ is convex and lsc on $\mathbb{R}^n \times \mathbb{R}^{2p} \times [0, +\infty[$. It is inf-compact on $\mathbb{R}^n \times \mathbb{R}^{2p} \times [0, \overline{\mu}]$, for all $\overline{\mu} > 0$ being fixed. Moreover, θ is convex and continuous on $[0, +\infty[$, $\theta(0) = \alpha$ and $f(x, z) = \gamma(x, z, 0)$, for all $(x, z) \in \mathbb{R}^n \times]0, +\infty[^{2p}$.*

Proof It is known that the function ζ is convex on $\mathbb{R}^{2p} \times [0, +\infty[$ and so is γ. The function θ is then convex on $[0, +\infty[$ as the infimum over (x, z) of a convex function in (x, z, μ). Now, the function $\zeta(z, .)$ is continuous on $[0, +\infty[$ and, because of (10), $\zeta(z, 0) = 0$ for all $z \in]0, +\infty[^{2p}$. Thus $f(x, z) = \gamma(x, z, 0)$ for all $(x, z) \in \mathbb{R}^n \times]0, +\infty[^{2p}$ and therefore $\theta(0) = \alpha$ [the optimal value of problem (8)]. Set $\tilde{\gamma} = \gamma_{|\mathbb{R}^n \times \mathbb{R}^{2p} \times [0, \overline{\mu}]}$ the restriction of γ to the set $\mathbb{R}^n \times \mathbb{R}^{2p} \times [0, \overline{\mu}]$. Then, by Proposition 6.1,

$$\begin{aligned} &\{(d_x, d_z, \mu) : \ \tilde{\gamma}_\infty(d_x, d_z, \mu) \le 0, \ d_z \ge 0, \ \mu = 0\} \\ &\quad = \{(d_x, d_z, 0) : \ f_\infty(d_x, d_z) \le 0, \ d_z \ge 0\} \\ &\quad = \{(0, 0, 0)\}. \end{aligned}$$

The function γ is then inf-compact on $\mathbb{R}^n \times \mathbb{R}^{2p} \times [0, \overline{\mu}]$. Consequently, there is a compact $\tilde{S}$ such that $(x(\mu), z(\mu)) \in \tilde{S}$, $\forall \mu \in]0, \overline{\mu}]$, i.e., $(x(\mu), z(\mu))_{\mu \in (0, \overline{\mu})}$ is bounded. We established that θ is convex on $[0, +\infty[$. It is then continuous on $]0, +\infty[$. Let us show now that $\lim\limits_{\mu \downarrow 0} \theta(\mu) = \theta(0) = \alpha$. In this respect, we shall prove that $\lim\limits_{\mu \downarrow 0} \mu \ln \left(\dfrac{\varphi(z(\mu))}{\mu} \right) = 0$. Let $(\mu^k)_{k \in \mathbb{N}}$ be a positive sequence such that $\lim\limits_{k \uparrow +\infty} \mu^k = 0$. We established that $(x(\mu), z(\mu))_{\mu \in]0, \overline{\mu}]}$ is bounded. It follows that the set $\{(x(\mu^k), z(\mu^k))\}$ contains a subsequence converging to a point $(\tilde{x}, \tilde{z})$. In the case, where $\tilde{z} > 0$ the result is obvious. Assume that $\varphi(\tilde{z}) = 0$. Then, for k sufficiently large, one has

$$\begin{aligned} \alpha - \mu^k \ln \left(\frac{\varphi(z)}{\mu^k} \right) \le \theta(\mu^k) &= f\left(x(\mu^k), z(\mu^k) \right) - \mu^k \ln \left(\frac{\varphi(z(\mu^k))}{\mu^k} \right) \\ &\le f(x, z) - \mu^k \ln \left(\frac{\varphi(z)}{\mu^k} \right) \end{aligned}$$

for every (x, z) satisfying $z > 0$. Since $\lim_{k\uparrow 0} \mu^k \ln\left(\frac{\varphi(z)}{\mu^k}\right) = 0$, we have

$$\alpha \leq \lim_{k\uparrow+\infty} \inf \theta(\mu^k) \leq f(x, z)$$

and then

$$\alpha \leq \lim_{k\uparrow+\infty} \sup \theta(\mu^k) \leq \inf_{x,z}\{f(x, z) : z > 0\} = \inf_{x,z}\{f(x, z) : z \geq 0\} = \alpha.$$

Consequently, $\lim_{k\uparrow+\infty} \theta(\mu^k) = \alpha$. □

Given $\mu > 0$, the KKT optimality conditions for the problem (P_μ) can be formulated, for some $u \in \mathbb{R}^p$, as

$$\begin{aligned}
&Qx(\mu) - c - Du = 0,\\
&\lambda e - \frac{\mu}{2p}(Z(\mu))^{-1}e - \tilde{I}^* u = 0,\\
&D^* x(\mu) + \tilde{I} z(\mu) = 0,
\end{aligned}$$

where $Z(\mu) = diag(z(\mu))$. Observe that u is necessarily unique. Put

$$u = u(\mu) \text{ and } s(\mu) = \frac{\mu}{2p} Z^{-1}(\mu)e.$$

We rewrite the KKT conditions as

$$\begin{aligned}
&Qx(\mu) - c - Du(\mu) = 0, && (E1)\\
&\lambda e - s(\mu) - \tilde{I}^* u(\mu) = 0, && (E2)\\
&Z(\mu)s(\mu) = \frac{\mu}{2p} e, && (E3)\\
&D^* x(\mu) + \tilde{I} z(\mu) = 0. && (E4)
\end{aligned}$$

Proposition 6.4 *For every $\mu > 0$, $(s(\mu), u(\mu))$ is a feasible solution to (9) and $\big((s(\mu), u(\mu)\big)_{\mu\in]0,\overline{\mu}]}$ is bounded.*

Proof By $(E1)$, $(E2)$ and the fact that $s(\mu) = \frac{\mu}{2p}(Z(\mu))^{-1}e > 0$, $(u(\mu), s(\mu))$ is a feasible solution to (9). The boundedness of $(s(\mu), u(\mu))_{\mu\in(0,\overline{\mu}]}$ is due to Proposition 6.1. □

Set $\overline{I} = \bigcup_{z\in S(.,(P))} I(z)$ and $\overline{J} = \bigcup_{s\in S(.,(D))} J(s)$, where

$$\begin{aligned}
S(., (P)) &= \left\{z : \exists x \in \mathbb{R}^n \text{ such that } (x, z) \in S_{(P)}\right\},\\
S(., (D)) &= \left\{s : \exists u \in \mathbb{R}^p \text{ such that } (s, u) \in S_{(D)}\right\},
\end{aligned}$$

and $I(z) = \{i : z_i > 0\}$ and $J(s) = \{i : s_i > 0\}$ are the supports of z and s, respectively.

Lemma 6.1 *There is at least one couple* $(\hat{z}, \hat{s}) \in S(., (P)) \times S(., (D))$ *such that* $\overline{I} = I(\hat{z})$ *and* $\overline{J} = J(\hat{s})$.

Proof We have $\overline{I}$ a subset of a finite set $\{1, \ldots, 2p\}$. Let us consider the k-tuple $(z^1, z^2, \ldots, z^k) \in S(., (P))^k$, for some $k \in \{1, 2, \ldots, 2p\}$ satisfying $\overline{I} = I\left(z^1\right) \cup I\left(z^2\right) \cup \cdots \cup I\left(z^k\right)$. Set $\hat{z} = \dfrac{1}{k}\left(z^1 + z^2 + \cdots + z^k\right)$. Since $S(., (P))$ is convex $\hat{z} \in S(., (P))$. So it is easy to see that $I(z^i) \subset I(\hat{z})$, $\forall i \in \{1, 2, \ldots, k\}$. The result then follows. The vector $\hat{s}$ is constructed in a similar way. $\square$

Observe that every optimal solution (x, z) of the problem (8) satisfying $I(z) = \overline{I}$ is in the relative interior of $S_{(P)}$. Similarly, every optimal solution (x, s, u) of the problem (9) satisfying $J(s) = \overline{J}$ is in the relative interior of $S_{(D)}$.

Set

$$(\overline{x}, \overline{z}) = \arg\max\left\{\varphi_{\overline{I}}(z_{\overline{I}}) : \frac{1}{2}\langle Qx, x\rangle - \langle c, x\rangle + \lambda\langle e, z\rangle = \alpha,\ D^*x + \tilde{I}z = 0,\ z_{\overline{J}} = 0\right\},$$

where

$$\varphi_{\overline{I}}(z_{\overline{I}}) = \begin{cases} \left(\prod\limits_{i\in\overline{I}} z_i\right)^{\frac{1}{\mathrm{Card}(\overline{I})}}, & \text{if } z_J \in]0, +\infty[^{\mathrm{Card}(J)}, \\ -\infty, & \text{elsewhere.} \end{cases}$$

Symmetrically, we set

$$(\overline{s}, \overline{u}) = \arg\max\left\{\varphi_{\overline{J}}(s_{\overline{J}}) :\ s = \lambda e - \tilde{I}^*u,\ Du + c - Q\overline{x} = 0, s_{\overline{I}} = 0\right\},$$

where

$$\varphi_{\overline{J}}(s_{\overline{J}}) = \begin{cases} \left(\prod\limits_{i\in\overline{J}} s_i\right)^{\frac{1}{\mathrm{Card}(\overline{J})}}, & \text{if } s_{\overline{J}} \in]0, +\infty[^{\mathrm{Card}(\overline{J})}, \\ -\infty, & \text{elsewhere.} \end{cases}$$

$(\overline{x}, \overline{z})$ is called the analytic center[2] of (8) and $(\overline{x}, \overline{s}, \overline{u})$ the analytic center of (9). The uniqueness is ensured by the strict quasiconcavity of functions $\varphi_{\overline{I}}$ and $\varphi_{\overline{J}}$ on the interior of their respective domain and the assumption (5).

[2] A generalization of the central path and the analytic center is proposed in [21] by using the so-called concave gauge functions.

6.2 Convergence to a Maximal Solution

We now give an important result concerning the convergence of the path toward an element of the relative interior of the solution set. Its proof is inspired in part by those of Theorems I.7 and I.9 in [23].

Theorem 6.1 *Under assumption* (5)*, we have*

$$\lim_{\mu\downarrow 0}(x(\mu), z(\mu), s(\mu), u(\mu)) = (\overline{x}, \overline{z}, \overline{s}, \overline{u}).$$

Moreover, $(\overline{x}, \overline{z})$ *and* $(\overline{x}, \overline{s}, \overline{u})$ *belong to the relative interior of* $S_{(P)}$ *and* $S_{(D)}$*, respectively.*

Proof According to Propositions 6.3 and 6.4, the nets $\big((x(\mu), z(\mu)\big)_{\mu\in]0,\overline{\mu}]}$ and $\big((s(\mu), u(\mu)\big)_{\mu\in]0,\overline{\mu}]}$ are bounded. Let $(\mu^k)_{k\in\mathbb{N}}$ a positive increasing sequence satisfying

$$\lim_{k\uparrow+\infty}\mu^k = 0 \text{ and } \lim_{k\uparrow+\infty}\Big(x(\mu^k), z(\mu^k), s(\mu^k), u(\mu^k)\Big) = (\tilde{x}, \tilde{z}, \tilde{s}, \tilde{u}).$$

Then replacing μ by μ^k in $(E1)-(E4)$ and letting k tend to $+\infty$, we observe that the pair $\{(\tilde{x}, \tilde{z}), (\tilde{x}, \tilde{s}, \tilde{u})\}$ satisfies the KKT optimality conditions of (8) and then it is a primal-dual optimal solution pair of (8). Let us show now that $I(\tilde{z}) = \overline{I}$ and $J(\tilde{s}) = \overline{J}$. Now by $(E1)$, $(E2)$ and $(E4)$ we have

$$\begin{pmatrix} x(\mu^k) - \overline{x} \\ z(\mu^k) - \overline{z} \end{pmatrix} \in \text{Ker}\,\big(D^*\ \tilde{I}\big) \text{ and } \begin{pmatrix} Q(x(\mu^k) - \overline{x}) \\ -(s(\mu^k) - \overline{s}) \end{pmatrix} \in \Im\begin{pmatrix} D \\ \tilde{I}^* \end{pmatrix}.$$

Then using the following orthogonality property

$$\text{Ker}\,\big(D^*\ \tilde{I}\big) = \left[\Im\begin{pmatrix} D \\ \tilde{I}^* \end{pmatrix}\right]^{\perp}, \tag{11}$$

$(E3)$ and the fact that $\langle\overline{z}, \overline{s}\rangle = \langle\tilde{z}, \tilde{s}\rangle = 0$ we have

$$\langle\overline{z}, s(\mu^k)\rangle + \langle\overline{s}, z(\mu^k)\rangle = \mu^k - \langle Q(x(\mu^k) - \overline{x}), x(\mu^k) - \overline{x}\rangle.$$

Since in addition $I(\overline{z}) = \overline{I}$, $J(\overline{s}) = \overline{J}$ and Q is positive semi-definite we get

$$\sum_{i\in\overline{I}}\overline{z}_i s(\mu^k)_i + \sum_{i\in\overline{J}}\overline{s}_i z(\mu^k)_i = \mu^k - \langle Q(x(\mu^k) - \tilde{x}), x(\mu^k) - \tilde{x}\rangle \le \mu^k.$$

But from $(E3)$, $z(\mu^k)_i s(\mu^k)_i = \dfrac{\mu^k}{2p}$, for all i. It follows that

$$\sum_{i\in\overline{J}}\frac{\overline{s}_i}{s(\mu^k)_i} + \sum_{i\in\overline{I}}\frac{\overline{z}_i}{z(\mu^k)_i} \le 2p.$$

Now, let k tending to $+\infty$, we get on the one hand

$$0 < \sum_{i \in \overline{J}} \frac{\overline{s}_i}{\tilde{s}_i} + \sum_{i \in \overline{I}} \frac{\overline{z}_i}{\tilde{z}_i} \leq 2p < +\infty$$

and then, by construction of $\overline{I}$ and $\overline{J}$, we have necessarily $I(\tilde{z}) = \overline{I}$ and $J(\tilde{s}) = \overline{J}$. On the other hand, using the arithmetic-geometric mean inequality we get

$$\left(\prod_{i \in \overline{J}} \frac{\overline{s}}{\tilde{s}_i} \prod_{i \in \overline{I}} \frac{\overline{z}}{\tilde{z}_i} \right)^{\frac{1}{2p}} \leq \frac{1}{2p} \left(\sum_{i \in \overline{J}} \frac{\overline{s}}{\tilde{s}_i} + \sum_{i \in \overline{I}} \frac{\overline{z}}{\tilde{z}_i} \right) \leq 1$$

and then

$$\varphi_{\overline{J}}(\overline{s}_{\overline{J}}) \varphi_{\overline{I}}(\overline{z}_{\overline{I}}) \leq \varphi_{\overline{J}}(\tilde{s}_{\overline{J}}) \varphi_{\overline{I}}(\tilde{z}_{\overline{I}}).$$

But, by definition of $(\overline{x}, \overline{z}, \overline{s}, \overline{u})$, $\varphi_{\overline{J}}(\tilde{s}_{\overline{J}}) \leq \varphi_{\overline{J}}(\overline{s}_{\overline{J}})$ and $\varphi_{\overline{I}}(\tilde{z}_{\overline{I}}) \leq \varphi_{\overline{I}}(\overline{z}_{\overline{I}})$. The result then follows. $\square$

Consequently, the following corollary holds

Corollary 6.1 *Under assumption* (5), *we have* $\lim_{\mu \downarrow 0} x(\mu) = \bar{x} \in \operatorname{ri} \mathbf{X}_\lambda$.

Proof By Theorem 6.1, $(\overline{x}, \overline{z})$ belongs to the relative interior of $S_{(P)}$ and hence $\overline{x}$ belongs to the linear projection of the relative interior of $S_{(P)}$, which is equal to $\operatorname{ri} \mathbf{X}_\lambda$. $\square$

Using this analysis, we propose an algorithm directly adapted from the Predictor-corrector Mehrotra's algorithm [24]. The pseudo-code is given in Algorithm 1. The user is expected to give an initialization point (x^0, z^0, u^0, s^0) satisfying $z^0 > 0$ and $s^0 > 0$, the scenario Φ, D^*, y, a stopping criterion $\epsilon > 0$, and a relaxation parameter $\eta \in]0, 1[$.

6.3 Non-uniqueness Examples

We now provide examples of non-uniqueness and discuss the behavior of the proposed algorithm.

Example for the Lasso. To illustrate our theoretical results, we consider at first a very simple scenario in $\mathbb{R}^2$ to $\mathbb{R}$. Let $D = \mathrm{Id}_2$, $\Phi = (1 \quad 1)$, $y = 1$ and $\lambda > 0$. The first-order conditions read as follow

$$x_1 + x_2 - 1 + \lambda s_1 = 0 \text{ and } x_1 + x_2 - 1 + \lambda s_2 = 0$$

where $s \in \partial \| \cdot \|_1$. We can rewrite it as

$$s_1 = -\frac{x_1 + x_2 - 1}{\lambda} \text{ and } s_2 = -\frac{x_1 + x_2 - 1}{\lambda}.$$

Algorithm 1 Adapted predictor–corrector Mehrotra's algorithm

Input: (x^0, z^0, u^0, s^0), Φ, D^*, y, $\epsilon > 0$, $\eta \in]0, 1[$

$Q \leftarrow \Phi^*\Phi$, $c \leftarrow \Phi^* y$

Set complementarity measure

$$r_1 \leftarrow Qx - c - Du, r_2 \leftarrow \lambda e - s - \tilde{I}^* u, r_3 \leftarrow Zs, r_4 \leftarrow D^* x + \tilde{I} z.$$

$\mu \leftarrow \dfrac{\langle z, s\rangle}{2p}$

while $\max\{\|r_1\|_2, \|r_2\|_2, \|r_3\|_2, \|r_4\|_2\} > \epsilon$ **do**

Compute the affine scaling direction $(d_x^a, d_z^a, d_u^a, d_s^a)$ by solving the system

$$\begin{aligned} Qd_x^a - Dd_u^a &= -r_1 \\ -d_s^a - \tilde{I}^* d_u^a &= -r_2 \\ Sd_z^a + Zd_s^a &= -r_3 \\ D^* d_x^a + \tilde{I} d_z^a &= -r_4, \end{aligned}$$

$t_{\max}^a \leftarrow \max\{t \geq 0 : \ z + td_z^a \geq 0, \ s + d_s^a \geq 0\}$

$\mu^a \leftarrow \dfrac{\langle z + t_{\max}^a d_z^a, s + t_{\max}^a d_s\rangle}{2p}$

$\sigma \leftarrow \left(\dfrac{\mu^a}{\mu}\right)^3$ ▷ centering parameter

Compute corrector and centering direction $(d_x^c, d_z^c, d_u^c, d_s^c)$ by solving

$$\begin{aligned} Qd_x^c - Dd_u^c &= 0 \\ -d_s^c - \tilde{I}^* d_u^c &= 0 \\ Sd_z^c + Zd_s^c &= -D_z^a d_s^a + \sigma\mu e \\ D^* d_x^a + \tilde{I} d_z^a &= 0, \end{aligned} \quad \text{where } D_z^a = \operatorname{diag}(d_z^a)$$

$(d_x, d_z, d_u, d_s) \leftarrow (d_x^a, d_z^a, d_u^a, d_s^a) + (d_x^c, d_z^c, d_u^c, d_s^c)$ ▷ predictor direction

$t_{\max} \leftarrow \max\{t \geq 0 : \ z + td_z \geq 0, \ s + d_s \geq 0\}$

$(x, z, u, s) \leftarrow (x, z, u, s) + \eta t_{\max}(d_x, d_z, d_u, d_s)$

Update complementarity measure

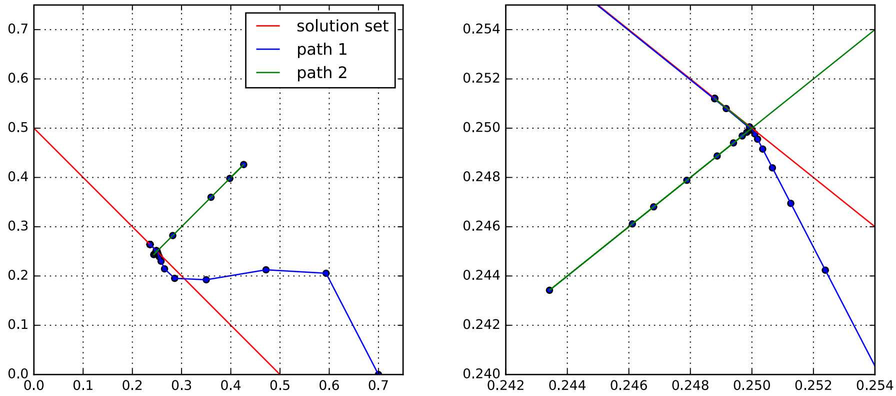
$$r_1 \leftarrow Qx - c - Du, r_2 \leftarrow \lambda e - s - \tilde{I}^* u, r_3 \leftarrow Zs, r_4 \leftarrow D^* x + \tilde{I} z.$$

end while

Using the classical criterion, we know that for $\lambda \geq \|\Phi^* y\|_\infty$, the unique solution of the problem is equal to zero. Here, $\Phi^* y = (1 \quad 1)^*$. Hence, for any $\lambda > 1$, the solution is $(0, 0)$. Thus, we restrict our attention to $\lambda \in]0, 1[$. Observe that for any $\lambda \in]0, 1[$, the vectors $(1 - \lambda \quad 0)^*$ and $(0 \quad 1 - \lambda)^*$ are solutions. Now, since any solution shares the same image by Φ and the same ℓ_1-norm, we have that for any $x \in X_\lambda$,

$$x_1 + x_2 = 1 - \lambda \text{ and } |x_1| + |x_2| = x_1 + x_2 = 1 - \lambda.$$

It means that the solution set is the segment defined by $(1 - \lambda \quad 0)^*$ and $(0 \quad 1 - \lambda)^*$. Figure 1 represents the evolution of the primal iterate on the plane $\mathbb{R}^2$ for $\lambda = \frac{1}{2}$. We remark that contrary to other algorithms which are sensitive to the initialization (such as an iterative soft–thresholding), the proposed barrier method converges to a solution in the relative interior, independently of the initialization.



(a) Path **(b)** Zoom around the analytical center

Fig. 1 Algorithm path. The red line corresponds to the solution set $\mathbf{X}_\lambda$, the blue line is the algorithm path for $x^0 = (0.7\,0)^*$ and the green line for x^0 obtained by a least-square approximation

When D is not the identity Let us consider now the following scenario in $\mathbb{R}^3$. Let

$$D^* = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{pmatrix}, \quad \Phi = \begin{pmatrix} 1 & 1 & 1 \\ 3 & 1 & 1 \\ \sqrt{2} & 0 & 0 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \text{ and } \lambda = \frac{1}{2}. \tag{12}$$

We prove that a necessary and sufficient condition that x is an optimal solution is that $x_1 = 0$, $x_2 + x_3 = \frac{1}{2}$, $x_2 \geq 0$ and $x_3 \geq 0$. So we get

$$\mathbf{X}_\lambda = \operatorname{conv}\left\{ \begin{pmatrix} 0 & \frac{1}{2} & 0 \end{pmatrix}^*, \begin{pmatrix} 0 & 0 & \frac{1}{2} \end{pmatrix}^* \right\}.$$

Obtaining Examples in Higher Dimensions We know explain how the example in the introduction was constructed. It is easy to see that if

$$\Phi = \begin{pmatrix} \Phi_1 & 0 \\ 0 & \Phi_2 \end{pmatrix}, \quad D^* = \begin{pmatrix} D_1^* & 0 \\ 0 & D_2^* \end{pmatrix}, \text{ and } y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

then the solution set $\mathbf{X}_\lambda$ is given by

$$\mathbf{X}_\lambda^1 \otimes \mathbf{X}_\lambda^2 = \left\{ (x_1, x_2) \in \mathbb{R}^{n_1+n_2} \ : \ x_1 \in \mathbf{X}_\lambda^1, x_2 \in \mathbf{X}_\lambda^2 \right\},$$

where $\mathbf{X}_\lambda^i \subseteq \mathbb{R}^{n_i}$ is the solution of $\min\limits_x \dfrac{1}{2} \|y_i - \Phi_i x\|_2^2 + \lambda \left\| D_i^* x \right\|_1$. Now, take $\lambda = \frac{1}{2}$, Φ_1, D_1^* and y_1 as explicited in (12), and take $k \geq 2$,

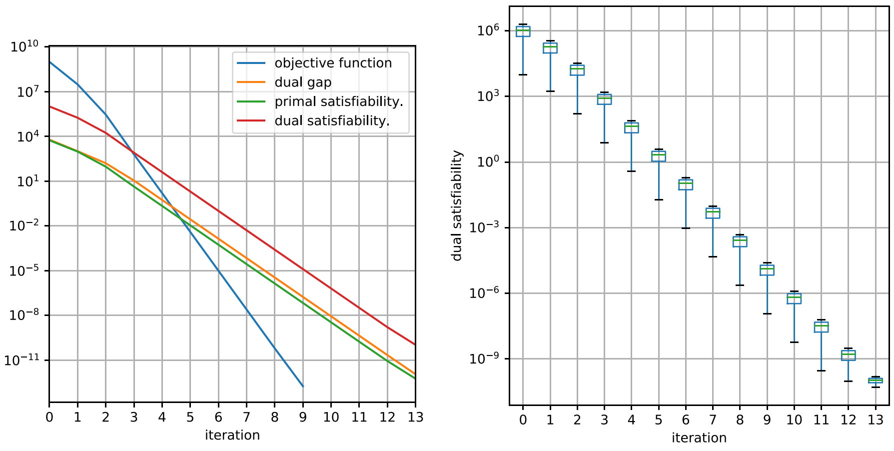


Fig. 2 Behavior of Algorithm 1 for $p = 3000$ with 100 realizations of random initializations. **a** Mean values of the objective function (blue), dual gap (orange), primal satisfiability norm (green) and dual satisfiability (red) toward convergence in log scale. **b** Box plot of the dual satisfiability in log scale. Green line indicates the mean value, black line min and max value, and the box indicates the empirical standard deviation

$$\Phi_2 = \mathrm{Id}_k, \quad D_2^* = \begin{pmatrix} 1 & -1 & 0 & \dots & \dots & 0 \\ 0 & 1 & -1 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 & -1 & 0 \\ 0 & \dots & \dots & 0 & 1 & -1 \end{pmatrix} \text{ and } y_2 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{13}$$

Clearly, $\mathbf{X}_\lambda^2 = \{y_2\}$. Thus, taking the product between $\mathbf{X}_\lambda^1$ and $\mathbf{X}_\lambda^2$, we obtain solutions such as described in the introduction.

The behavior of the Interiorpoint Method Thanks to the previous remark, we build an artificial example where the scenario is 1000 partial copy of (13), where we change the value $\Phi_{3,1}$, i.e.,

$$D^* = \bigotimes_{i=1}^{1000} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{pmatrix}, \quad \Phi = \bigotimes_{i=1}^{1000} \begin{pmatrix} 1 & 1 & 1 \\ 3 & 1 & 1 \\ \sqrt{1+i} & 0 & 0 \end{pmatrix}, \quad y = \bigotimes_{i=1}^{1000} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},$$

hence $y \in \mathbb{R}^{3000}$, $D, \Phi \in \mathbb{R}^{3000 \times 3000}$. Analytically, we show that the solution $\mathbf{X}_\lambda$ is given by

$$\mathbf{X}_\lambda = \bigotimes_{i=1}^{1000} \operatorname{conv}\left\{ \begin{pmatrix} 0 & \frac{1}{2} & 0 \end{pmatrix}^*, \begin{pmatrix} 0 & 0 & \frac{1}{2} \end{pmatrix}^* \right\}$$

$$= \operatorname{conv}\left\{ \bigotimes_{i=1}^{1000} \begin{pmatrix} 0 & \frac{1}{2} & 0 \end{pmatrix}^*, \bigotimes_{i=1}^{1000} \begin{pmatrix} 0 & 0 & \frac{1}{2} \end{pmatrix}^* \right\}.$$

During the execution of the algorithm, we monitor several values: value of the objective function $h(x) = \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|D^* x\|_1$, value of the dual gap, feasibility of the primal problem ($\|r_4\|$) and feasibility of the dual problem ($\max(\|r_1\|, \|r_2\|)$). We show in Fig. 2 the decay of these values with respect to 100 realizations of a Gaussian random vector as initialization. We observe an empirical fast convergence rate, and the algorithm satisfies the stopping criterion around iteration 13.

7 Conclusions

As we can see, the geometric characterization of $\mathbf{X}_\lambda$, the optimal solution set of (3), played a central role in the topological characterization of $\mathbf{S}_\lambda$, the maximal D-support optimal solution set of (3). The topological characterization of $\mathbf{S}_\lambda$ via Theorem 3.1 is particularly interesting in practice. Indeed, to determine an element of $\mathbf{S}_\lambda$, the fact that $\mathbf{S}_\lambda = \operatorname{ri} \mathbf{X}_\lambda$, our main result, has naturally suggested turning to a method of the interior points type, well known to be effective and robust in linear and quadratic programming.

Appendix

In this section, we propose to express some of our results in a general framework. More precisely, we consider the following optimization problem

$$\min_{x \in E} \{f(x) + g(x)\}. \tag{14}$$

where E is a Banach space, $f, g : E \mapsto \mathbb{R} \cup \{+\infty\}$ are convex lower semicontinuous functions. The dual space of E and the pairing between E and E^* will be denoted by E^* and $\langle \cdot, \cdot \rangle$, respectively. The Fenchel subdifferential of f at $\bar{x}$ is defined by

$$\partial f(\bar{x}) := \{x^* \in E^* : \langle x^*, x - \bar{x} \rangle \leq f(x) - f(\bar{x}) \,\forall x \in E\}.$$

The aim of the following proposition is to give a characterization of solutions of the problem (14).

Proposition A.1 *Let $\bar{x} \in E$ be a fixed solution of the problem (14) and $x^* \in \partial g(\bar{x})$ be such that $-x^* \in \partial f(\bar{x})$. Then the following assertions are equivalent:*

(1) *u is a solution of the problem (14),*
(2) *$g(u) \leq g(\bar{x}) + \langle x^*, u - \bar{x} \rangle$ and u is a solution of the problem*

$$\min_{x \in E} \{f(x) + \langle x^*, x \rangle\}. \tag{15}$$

Consequently, if $\{x \in E : g(x) \leq g(\bar{x}) + \langle x^*, x - \bar{x} \rangle\}$ *is a polyhedral set and the function* f *is polyhedral (supremum of a finite affine family), then so is* $\operatorname*{Argmin}_{x \in \mathbb{R}^n} \{f(x) + g(x)\}$.

Proof Since the implication (2) $\Longrightarrow$ (1) is obvious, we will establish only the implication (1) $\Longrightarrow$ (2). First note that, because of our assumptions, assertion (2) is equivalent to say that $x^* \in \partial g(u)$ and $-x^* \in \partial f(u)$. So if u is a solution of the problem (14), we have

$$f(u) + g(u) = f(\bar{x}) + g(\bar{x}). \tag{16}$$

Since $x^* \in \partial g(\bar{x})$ and $-x^* \in \partial f(\bar{x})$, we easily obtain, by using relation (16), that $x^* \in \partial g(u)$ and $-x^* \in \partial f(u)$ and the proof is completed. □

A particular and interesting case is the Hilbert setting with a special form of g.

Corollary A.1 *Suppose that* E *(resp.* F*) is a Hilbert endowed with a scalar product denoted by* $\langle \cdot, \cdot \rangle$ *and the associated norm* $\|\cdot\|$. *Let* $\Phi : E \mapsto F$ *be a linear continuous operator and* $y \in F$. *Define the function* $g : E \mapsto \mathbb{R}$ *by*

$$g(x) = \frac{1}{2}\|\Phi x - y\|^2.$$

Let $\bar{x} \in E$ *be a fixed solution of the problem* (14) *and put* $x^* = \Phi^*(\Phi\bar{x} - y)$. *Then the following assertions are equivalent:*

(1) u *is a solution of the problem* (14)*,*
(2) $\Phi u = \Phi\bar{x}$ *and* u *is a solution of the problem*

$$\min_{x \in E} \{f(x) + \langle x^*, x \rangle\}. \tag{17}$$

Consequently, each solution u *of the problem* (14) *satisfies* $\Phi u = \Phi\bar{x}$ *and* $f(u) = f(\bar{x})$.

Proof It suffices to see that the (in)equality $g(u) \leq g(\bar{x}) + \langle x^*, u - \bar{x} \rangle$ is equivalent to $\Phi u = \Phi\bar{x}$ and to apply Proposition A.1. □

The following corollary asserts that knowing one solution of (14), we can determine all the other ones.

Corollary A.2 *Let the assumptions of Corollary* A.1 *be satisfied. Then*

$$\operatorname*{Argmin}_{x \in \mathbb{R}^n} \{f(x) + g(x)\} = \{x \in E : \Phi x = \Phi\bar{x},\ f(x) = f(\bar{x})\}.$$

References

1. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM J. Sci. Comput. **20**(1), 33–61 (1999)
2. Mallat, S.G.: A wavelet tour of signal processing, 3rd edn. Academic Press, Amsterdam (2009)
3. Tibshirani, R.: Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. Ser. B. Methodol. **58**(1), 267–288 (1996)
4. Elad, M., Milanfar, P., Rubinstein, R.: Analysis versus synthesis in signal priors. Inverse Probl. **23**(3), 947 (2007)
5. Vaiter, S., Peyré, G., Dossal, C., Fadili, M.J.: Robust sparse analysis regularization. IEEE Trans. Inf. Theory **59**(4), 2001–2016 (2013)
6. Nam, S., Davies, M.E., Elad, M., Gribonval, R.: The cosparse analysis model and algorithms. Appl. Comput. Harmon. Anal. **34**(1), 30–56 (2013)
7. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D **60**(1), 259–268 (1992)
8. Steidl, G., Weickert, J., Brox, T., Mrázek, P., Welk, M.: On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides. SIAM J. Numer. Anal. **42**(2), 686–713 (2004)
9. Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K.: Sparsity and smoothness via the fused Lasso. J. R. Stat. Soc. Ser. B. Stat. Methodol. **67**(1), 91–108 (2005)
10. Zhang, H., Yin, W., Cheng, L.: Necessary and sufficient conditions of solution uniqueness in 1-norm minimization. J. Optim. Theory Appl. **164**(1), 109–122 (2015)
11. Zhang, H., Yan, M., Yin, W.: One condition for solution uniqueness and robustness of both l1-synthesis and l1-analysis minimizations. arXiv preprint arXiv:1304.5038 (2013)
12. Gilbert, J.C.: On the solution uniqueness characterization in the ℓ^1 norm and polyhedral gauge recovery. Technical report, INRIA Paris-Rocquencourt (2015)
13. Tibshirani, R.J.: The lasso problem and uniqueness. Electron. J. Stat. **7**, 1456–1490 (2013)
14. Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM J. Comput. **24**(2), 227–234 (1995)
15. Mallat, S.G., Zhang, Z.: Matching pursuits with time–frequency dictionaries. IEEE Trans. Signal Process. **41**(12), 3397–3415 (1993)
16. Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.S.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: Conference on Signals, Systems and Computers, pp. 40–44. IEEE (1993)
17. Needell, D., Tropp, J.A.: CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harmon. Anal. **26**(3), 301–321 (2009)
18. Attouch, H., Cominetti, R.: l^p approximation of variational problems in l^1 and l^∞. Nonlinear Anal. TMA **36**(3), 373–399 (1999)
19. Rockafellar, R.: Convex Analysis, vol. 28. Princeton University Press, Princeton (1996)
20. Frisch, K.R.: The logarithmic potential method of convex programming. Technical report, University Institute of Economics, Oslo, Norway (1955)
21. Barbara, A.: Strict quasi-concavity and the differential barrier property of gauges in linear programming. Optimization **64**(12), 2649–2677 (2015)
22. Barbara, A., Crouzeix, J.P.: Concave gauge functions and applications. Math. Methods OR **40**(1), 43–74 (1994)
23. Roos, C., Terlaky, T., Vial, J.P.: Interior Point Methods for Mathematical Programming. Wiley, New York (2009)
24. Mehrotra, S.: On the implementation of a primal-dual interior point method. SIAM J. Optim. **2**(4), 575–601 (1992)