NP-hardness of $\ell0$ minimization problems: revision and extension to the non-negative setting

Thi Thanh Nguyen, Charles Soussen, Jérôme Idier, El-Hadi Djermoune

NP-hardness of ℓ_0 minimization problems: revision and extension to the non-negative setting

Thanh T. NGUYEN[1], Charles SOUSSEN[2], Jérôme IDIER[3], El-Hadi DJERMOUNE[1]
[1]CRAN, Université de Lorraine, CNRS, F-54506 Vandœuvre-lès-Nancy, France
[2]L2S, CentraleSupélec-CNRS-Université Paris-Sud, Université Paris-Saclay, F-91192 Gif-sur-Yvette, France
[3]LS2N, UMR 6004, F-44321 Nantes, France
{thi-thanh.nguyen,el-hadi.djermoune}@univ-lorraine.fr; charles.soussen@centralesupelec.fr; jerome.idier@ls2n.fr

***Abstract*—Sparse approximation arises in many applications and often leads to a constrained or penalized ℓ_0 minimization problem, which was proved to be NP-hard. This paper proposes a revision of existing analyses of NP-hardness of the *penalized* ℓ_0 problem and it introduces a new proof adapted from Natarajan's construction (1995). Moreover, we prove that ℓ_0 minimization problems with non-negativity constraints are also NP-hard.**

I. INTRODUCTION

Sparse approximation appears in a wide range of applications, especially in signal processing, image processing and compressed sensing [1]. Given a signal data $\boldsymbol{y} \in \mathbb{R}^m$ and a dictionary A of size $m \times n$, the aim is to find a signal $\boldsymbol{x} \in \mathbb{R}^n$ that gives the best approximation $\boldsymbol{y} \approx A\boldsymbol{x}$ and has the fewest non-zero coefficients (*i.e.*, sparsest solution). This task leads to solving one of the following constrained or penalized ℓ_0 minimization problems:

$$\min_{\|\boldsymbol{y}-A\boldsymbol{x}\|_2 \leq \epsilon} \|\boldsymbol{x}\|_0 \qquad (\ell_0 C)$$

$$\min_{\|\boldsymbol{x}\|_0 \leq K} \|\boldsymbol{y}-A\boldsymbol{x}\|_2^2 \qquad (\ell_0 C')$$

$$\min_{\boldsymbol{x}} \|\boldsymbol{y}-A\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_0 \qquad (\ell_0 P)$$

in which ϵ, K and λ are positive quantities related to the noise standard deviation, the sparsity level and regularization strength, respectively. Letters C and P respectively indicate that the problem is constrained or penalized. Depending on application, the appropriate statement will be addressed. It is noteworthy that n and K often depend on m when one considers the size of problem. ($\ell_0 C$) and ($\ell_0 C'$) are well known to be NP-hard [2, 3]. The NP-hardness of ($\ell_0 P$) was claimed to be a particular case of more general complexity analyses in [4, 5]. However, we point out that these complexity analyses do not rigorously apply to ($\ell_0 P$) as claimed. In this paper, we justify the complexity analyses in [4, 5] do not apply to problem ($\ell_0 P$), and we provide a new proof for the NP-hardness of ($\ell_0 P$) adapted from Natarajan's construction [2].

In several applications such as geoscience and remote sensing [6, 7], audio [8], chemometrics [9] and computed tomography [10], the signal or image of interest is non-negative. In such contexts, one often addresses a minimization problem with both sparsity and non-negativity constraints [10–12]. Adding non-negativity constraints to ℓ_0 minimization problems yields the following problems:

$$\min_{\boldsymbol{x}\geq 0, \|\boldsymbol{y}-A\boldsymbol{x}\|_2 \leq \epsilon} \|\boldsymbol{x}\|_0 \qquad (\ell_0 C+)$$

$$\min_{\boldsymbol{x}\geq 0, \|\boldsymbol{x}\|_0 \leq K} \|\boldsymbol{y}-A\boldsymbol{x}\|_2^2 \qquad (\ell_0 C'+)$$

$$\min_{\boldsymbol{x}\geq 0} \|\boldsymbol{y}-A\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_0 \qquad (\ell_0 P+)$$

Several papers address non-negative ℓ_0 minimization problems in the literature (see, *e.g.*, [13–16]). However, to the best of our knowledge, the complexity of these problems has not been addressed yet, the question of their NP-hardness being still open. Here we show that these problems are NP-hard and the proof can be derived from the NP-hardness of ℓ_0 minimization problems.

The rest of paper is organized as follows. In Section II, we discuss the issues related to NP-hardness of ($\ell_0 P$) in existing analyses and we present our proof. In Section III, we discuss about the NP-hardness of non-negative ℓ_0 minimization problems. We draw some conclusions in Section IV.

This work was supported by the project ANR BECOSE (ANR-15-CE23-0021).

This work was carried out in part while T. T. Nguyen was visiting L2S during the academic years 2017-2019.

II. HARDNESS OF ℓ_0 MINIMIZATION PROBLEMS

A. Background on constrained ℓ_0 minimization problems

Let us recall that an NP-complete problem is a problem in NP to which any other problem in NP can be reduced in polynomial time. Thus NP-complete problems are identified as the hardest problems in NP. An NP-complete problem is strongly NP-complete if it remains NP-complete when all of its numerical parameters are bounded by a polynomial in the length of the input. NP-hard problems are at least as hard as NP-complete problems. However, NP-hard problems do not need to be in NP and do not need to be decision problems. Formally, a problem is NP-hard (respectively, strongly NP-hard) if a NP-complete (respectively, strongly NP-complete) problem can be reduced in polynomial time to it. The reader is referred to [17, 18] for more information on this topic.

In the literature, problem ($\ell_0 C$), called SAS in [2], is well known to be NP-hard [2, Theorem 1]. The NP-hardness of ($\ell_0 C$) is a valuable extension of an earlier result: the problem of minimum weight solution to linear equations (equivalent to ($\ell_0 C$) with $\epsilon = 0$) is NP-hard [17, p. 246]. Davis *et al.* proved that ($\ell_0 C'$), called M-optimal approximation in [3], is NP-hard for any $K < m$ [3, Theorem 2.1]. Both analyses of Natarajan and Davis were made by a polynomial time reduction from the "exact cover by 3-sets" problem[1] which is known to be NP-complete [17, p. 221].

B. Existing analyses on penalized ℓ_0 minimization

In [4, 5], the NP-hardness of ($\ell_0 P$) is deduced as a particular case of more general complexity analyses. However, it turns out that the latter do not apply to ($\ell_0 P$), as explained hereafter. Chen *et al.* [4] address the unconstrained ℓ_q-ℓ_p minimization problem, defined by:

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|_q^q + \lambda \|\boldsymbol{x}\|_p^p \qquad (\ell_q\text{-}\ell_p)$$

where $\lambda > 0$, $q \geq 1$ and $0 \leq p < 1$. The authors showed that problem (ℓ_q-ℓ_p) is NP-hard with any $\lambda > 0$, $q \geq 1$ and $0 \leq p < 1$ [4, Theorem 3]. Obviously, ($\ell_0 P$) is the case where $q = 2$ and $p = 0$. The proof was done by i) introducing an invertible transformation which scales any instance of problem (ℓ_q-ℓ_p) to the problem (ℓ_q-ℓ_p) with $\lambda = 1/2$, and ii) establishing a polynomial time reduction from the partition problem which is known to be NP-complete [17] to the problem (ℓ_q-ℓ_p) with $\lambda = 1/2$. In other words, they showed that problem (ℓ_q-ℓ_p) with $\lambda = 1/2$ is NP-hard and, because there exists an invertible transformation from any problem (ℓ_q-ℓ_p) to the one with $\lambda = 1/2$, every problem (ℓ_q-ℓ_p) is NP-hard. Similarly, they showed that (ℓ_q-ℓ_p) is strongly NP-hard [4, Theorem 5] by a reduction from the 3-partition problem which is known to be strongly NP-hard [17]. The invertible transform used in [4] is defined by:

$$\tilde{\boldsymbol{x}} = (2\lambda)^{1/p}\boldsymbol{x}, \quad \tilde{A} = (2\lambda)^{-1/p}A. \qquad (1)$$

Unfortunately, (1) is not well-defined when $p = 0$. Therefore, [4, Theorems 3 and 5] do not apply to ($\ell_0 P$) when $\lambda \neq 1/2$.

Using a different approach, Huo and Chen's paper [5] addresses the penalized least-squares problem defined by:

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \sum_{i=1}^{n} \phi(|x_i|), \qquad \text{(PLS)}$$

where ϕ is a penalty function mapping non-negative values to non-negative values. The authors showed that (PLS) is NP-hard if the penalty function ϕ satisfies the following four conditions [5, Theorem 3.1]:

C1. $\phi(0) = 0$ and $\forall\, 0 \leq \tau_1 < \tau_2$, $\phi(\tau_1) \leq \phi(\tau_2)$.

C2. There exists $\tau_0 > 0$ and a constant $d > 0$ such that

$$\phi(\tau) \geq \phi(\tau_0) - d(\tau_0 - \tau)^2$$

for every $0 \leq \tau < \tau_0$.

C3. For the aforementioned τ_0, if $\tau_1, \tau_2 < \tau_0$ then

$$\phi(\tau_1) + \phi(\tau_2) \geq \phi(\tau_1 + \tau_2).$$

C4. For every $0 \leq \tau < \tau_0$,

$$\phi(\tau) + \phi(\tau_0 - \tau) > \phi(\tau_0). \qquad (2)$$

The proof of [5, Theorem 3.1] is by a reduction from the NP-complete problem X3C to the decision version of (PLS); this leads to the NP-completeness of the decision version of (PLS) and so the NP-hardness of (PLS) [5, Appendix 1]. The authors claimed that the ℓ_0 penalty function satisfies conditions C1-C4 for $\tau_0 = d = 1$. Therefore, the (PLS) problem with the ℓ_0 penalty function is NP-hard [5, Corollary 3.2]. Unfortunately, it turns out that the ℓ_0 penalty does not fulfill condition C4 as claimed. Indeed, for $\tau = 0$ the strict inequality (2) becomes $\phi(0) > 0$. Besides, in the proof [5, Appendix 1], the inputs of the decision problem are not guaranteed to have rational values. This might also violate the polynomiality of the reduction. Therefore, [5, Theorem 3.1] does not apply to ($\ell_0 P$).

In [5], the authors also mention an alternate proof of NP-hardness of ($\ell_0 P$) from Huo and Ni's earlier paper [19] as a special case of their results. In this proof [19, Appendix A.1], the relation between ($\ell_0 P$) and ($\ell_0 C$) is established using the principle of Lagrange multiplier. More precisely, the authors introduce an instance of ($\ell_0 C$) in which ϵ is defined from the minimizer of ($\ell_0 P$) and argue that solving ($\ell_0 P$) is equivalent to solving the mentioned instance of ($\ell_0 C$), which is known to be NP-hard [2]. There are a number of issues in the NP-hardness proof in [19]. For instance, the proposed transformation between ($\ell_0 P$) and ($\ell_0 C$) is not a polynomial time reduction. Besides, it is well known that ($\ell_0 P$) and ($\ell_0 C$) are not equivalent [20].

C. New analysis on penalized ℓ_0 minimization problems

To prove that a problem T is NP-hard, one must establish a polynomial time reduction (briefly called reduction hereafter) from some known NP-hard or NP-complete problem to T [18]. Roughly speaking, the reduction from a problem T1 to another problem T2 implies that T1 is not harder than T2. Therefore, if there exists a reduction from T1 to T2 and if T1 is NP-hard, T2 must be NP-hard too. The NP-hardness proofs in [2] and [3] use this principle. As an adaptation of Natarajan's construction, we prove the NP-hardness of ($\ell_0 P$) using the same principle as follows.

Theorem II.1. *Problem ($\ell_0 P$) is NP-hard for $0 < \lambda < 3$.*

The proof is by a reduction from the known NP-complete problem X3C to ($\ell_0 P$). The proof contains three steps: (1) Construct an instance of ($\ell_0 P$) from a given instance of X3C; (2) Construct a solution of ($\ell_0 P$) from a solution of X3C; (3) Construct a solution of X3C from a solution of ($\ell_0 P$).

1) Construction of an instance of ($\ell_0 P$) from a given instance of X3C: Given an instance of X3C: $S = \{s_1, s_2, ..., s_m\}$ is a set of m elements. C is a collection of n triplets c_j, $1 \leq j \leq n$. Without loss of generality we

[1]The latter problem, denoted by X3C in [2, 17], is stated as follows: Given a set S and a collection C of 3-element subsets of S (called triplets), is there a subcollection of disjoint triplets that exactly covers S?

can assume that m is a multiple of 3 since otherwise there is trivially no exact cover so no solution of X3C.

We now construct an instance of $(\ell_0 P)$. Let $\boldsymbol{y} = [1, 1, ..., 1]^T \in \mathbb{R}^m$. Let $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq n}$ where $a_{ij} = 1$ if $s_i \in c_j$ and $a_{ij} = 0$ otherwise. Let $\lambda \in \mathbb{Q}$, $0 < \lambda < 3$. Let

$$F(\boldsymbol{x}) := \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_0. \tag{3}$$

2) Construction of a solution of $(\ell_0 P)$ from a solution of X3C: Assume that there is a subcollection of disjoint triplets $\hat{C}$ which exactly covers S. Let $\boldsymbol{x}^* = [x_1^*, x_2^*, ..., x_n^*]^T$ where $x_j^* = 1$ if $c_j \in \hat{C}$ and $x_j^* = 0$ otherwise. We will prove that $\boldsymbol{x}^*$ is a solution of $(\ell_0 P)$.

Since $\hat{C}$ exactly covers S, $|\hat{C}| = m/3$ and $\boldsymbol{y} = A\boldsymbol{x}^*$. Thus, $\|\boldsymbol{x}^*\|_0 = m/3$ and

$$F(\boldsymbol{x}^*) = 0 + \lambda \frac{m}{3} = \lambda \frac{m}{3}.$$

Suppose that there exists $\bar{\boldsymbol{x}}$ such that

$$F(\bar{\boldsymbol{x}}) < F(\boldsymbol{x}^*) = \lambda \frac{m}{3}. \tag{4}$$

Let us show that this leads to a contradiction.

Since $F(\bar{\boldsymbol{x}}) \geq \lambda \|\bar{\boldsymbol{x}}\|_0$, from (4) we have $\|\bar{\boldsymbol{x}}\|_0 < m/3$. Therefore, we can rewrite $\|\bar{\boldsymbol{x}}\|_0 = m/3 - q$ for some $q \in \mathbb{N}$, $1 \leq q < m/3$. Note that $A\bar{\boldsymbol{x}}$ has m entries. Since the number of non-zero entries of $A\bar{\boldsymbol{x}}$ identifies with the number of elements s_i recovered by the subcollection corresponding to $\bar{\boldsymbol{x}}$, this number cannot exceed $3\|\bar{\boldsymbol{x}}\|_0 = m - 3q$. As a result, the number of zero entries of $A\bar{\boldsymbol{x}}$ must be between $3q$ and m. Since $\boldsymbol{y}$ is the all-one vector, $\boldsymbol{y} - A\bar{\boldsymbol{x}}$ has at least $3q$ entries valued 1, which implies

$$\|\boldsymbol{y} - A\bar{\boldsymbol{x}}\|_2^2 \geq 3q. \tag{5}$$

Hence,

$$F(\bar{\boldsymbol{x}}) \geq 3q + \lambda \left(\frac{m}{3} - q \right) = \lambda \frac{m}{3} + (3 - \lambda) q > \lambda \frac{m}{3}, \tag{6}$$

which contradicts (4). Therefore, $\boldsymbol{x}^*$ is a solution of $(\ell_0 P)$.

3) Construction of a solution of X3C from a solution of $(\ell_0 P)$: Assume that $\boldsymbol{x}^*$ is a solution of $(\ell_0 P)$. We will consider four cases as follows.

a) Case $\|\boldsymbol{x}^*\|_0 > m/3$: We deduce that X3C has no solution. Indeed, assume that $\hat{C}$ is an exact cover for S. Define $\boldsymbol{x} = [x_1, x_2, ..., x_n]^T$ where $x_j = 1$ if $c_j \in \hat{C}$ and $x_i = 0$ otherwise. Then we have

$$F(\boldsymbol{x}) = \lambda \frac{m}{3} < \lambda \|\boldsymbol{x}^*\|_0 \leq F(\boldsymbol{x}^*)$$

which contradicts the fact that $\boldsymbol{x}^*$ is a solution of $(\ell_0 P)$.

b) Case $\|\boldsymbol{x}^*\|_0 < m/3$: We deduce that X3C has no solution. Indeed, assume that $\hat{C}$ is an exact cover for S. Let $\boldsymbol{x} = [x_1, x_2, ..., x_n]^T$ where $x_j = 1$ if $c_j \in \hat{C}$ and $x_i = 0$ otherwise. Then we have $F(\boldsymbol{x}) = \lambda \frac{m}{3}$. Since $\|\boldsymbol{x}^*\|_0 < m/3$, we can write $\|\boldsymbol{x}^*\|_0 = m/3 - q$ for some $q \in \mathbb{N}$ and $1 \leq q < m/3$. Similar to (6), we have $F(\boldsymbol{x}^*) > \lambda \frac{m}{3}$. Since $F(\boldsymbol{x}) = \lambda \frac{m}{3}$, we obtain $F(\boldsymbol{x}^*) > F(\boldsymbol{x})$ which contradicts the fact that $\boldsymbol{x}^*$ is a solution of $(\ell_0 P)$.

c) Case where $\|\boldsymbol{x}^*\|_0 = m/3$ and $\boldsymbol{y} \neq A\boldsymbol{x}^*$: We deduce that X3C has no solution. Indeed, assume that $\hat{C}$ is an exact cover for S. Define $\boldsymbol{x} = [x_1, x_2, ..., x_n]^T$ where $x_j = 1$ if $c_j \in \hat{C}$ and $x_i = 0$ otherwise. Then we have

$$F(\boldsymbol{x}) = \lambda \frac{m}{3} < \|\boldsymbol{y} - A\boldsymbol{x}^*\|_2^2 + \lambda \|\boldsymbol{x}^*\|_0 = F(\boldsymbol{x}^*)$$

which contradicts the fact that $\boldsymbol{x}^*$ is a solution of $(\ell_0 P)$.

d) Case where $\|\boldsymbol{x}^*\|_0 = m/3$ and $\boldsymbol{y} = A\boldsymbol{x}^*$: Let $\hat{C}$ be the collection of triplets c_j such that the j^{th} entry of $\boldsymbol{x}^*$ is non-zero. Obviously, $\hat{C}$ is an exact cover for S so a solution of X3C.

Thus Theorem II.1 is proved.

It is notable that the proof above is also valid when $F(\boldsymbol{x}) := \|\boldsymbol{y} - A\boldsymbol{x}\|_p^p + \lambda \|\boldsymbol{x}\|_0$ for any $p \geq 1$. Indeed, one only needs to check whether (5) still holds when the ℓ_2 norm is replaced by the ℓ_p norm with $p \geq 1$. This is the case since $\boldsymbol{y} - A\bar{\boldsymbol{x}}$ has at least $3q$ entries equal to 1. Therefore, we have the following generalization of Theorem II.1.

Theorem II.2. *Problem* $\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|_p^p + \lambda \|\boldsymbol{x}\|_0$ *is NP-hard for* $p \geq 1$ *and* $0 < \lambda < 3$.

III. Hardness of non-negative ℓ_0 minimization problems

The NP-hardness of non-negative ℓ_0 minimization problems is a consequence of NP-hard proofs of $(\ell_0 C)$ [2], $(\ell_0 C')$ [3] and $(\ell_0 P)$ (Theorem II.1). Indeed, all these proofs consist in a reduction from X3C and the solution that established equivalence is binary. Therefore, the additional non-negativity constraints do not change the validity of these proofs. In other words, one can repeat the same proofs as in [2, 3] and that of Theorem II.1 for the corresponding non-negative ℓ_0 minimization problems $(\ell_0 C+)$, $(\ell_0 C'+)$ and $(\ell_0 P+)$. Another way to prove the NP-hardness of non-negative ℓ_0 minimization problems is by a reduction from the corresponding ℓ_0 minimization problems which are known to be NP-hard. In this reduction, the instance of non-negative problems is defined by

$$\tilde{\boldsymbol{y}} = \boldsymbol{y}, \quad \tilde{A} = [A, -A], \quad \tilde{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{x}^+ \\ \boldsymbol{x}^- \end{bmatrix}$$

where $\boldsymbol{x}^+ = \max\{\boldsymbol{x}, \boldsymbol{0}\}$, $\boldsymbol{x}^- = \max\{-\boldsymbol{x}, \boldsymbol{0}\}$. Naturally, by this construction, one gets $\tilde{\boldsymbol{x}} \geq \boldsymbol{0}$, $\|\tilde{\boldsymbol{x}}\|_0 = \|\boldsymbol{x}\|_0$ and $\tilde{A}\tilde{\boldsymbol{x}} = A\boldsymbol{x}$. The proofs (skipped for brevity) contain three steps similar to that of Theorem II.1.

Therefore, we can state the following theorem without proof.

Theorem III.1. $(\ell_0 C+)$, $(\ell_0 C'+)$ *are NP-hard. The same for* $(\ell_0 P+)$ *with* $0 < \lambda < 3$.

In the same spirit and using the same argument as at the end of Section II-C one can directly extend Theorem II.2 to the non-negative setting.

Theorem III.2. *Problem* $\min_{\boldsymbol{x} \geq \boldsymbol{0}} \|\boldsymbol{y} - A\boldsymbol{x}\|_p^p + \lambda \|\boldsymbol{x}\|_0$ *is NP-hard for* $p \geq 1$ *and* $0 < \lambda < 3$.

IV. Conclusion

NP-hardness of penalized ℓ_0 minimization problems cannot be deduced from previous complexity analyses, as stated in [4, 5]. Here, we introduced a new proof of NP-hardness of penalized ℓ_0 minimization problems when the regularization parameter λ is smaller than 3, by an adaptation of Natarajan's construction [2], while the case $\lambda \geq 3$ is still open. Besides, we showed that the ℓ_0 minimization problems with non-negative constraints are also NP-hard.

This work can be extended in several directions. For instance, researchers interested in what makes NP-hard problems even harder might be interested in the strong NP-hardness of the aforementioned optimization problems. As it is widely believed that X3C is strongly NP-complete, one might easily deduce the strong NP-hardness of $(\ell_0 C)$, $(\ell_0 C')$ and other problems which are reduced from X3C. However, to the best of our knowledge, X3C is only proved to be NP-complete [17, pp. 53, 221] and the strong NP-completeness has not been rigorously shown yet. Therefore, we believe that the question of strong NP-hardness of (non-negative) ℓ_0 minimization problems is not trivial and needs more work in future.

Besides, as (non-negative) ℓ_0 minimization problems are NP-hard, it would be interesting to know if the associated decision problems are in NP (so being NP-complete). Let us consider the decision problem associated with $(\ell_0 C)$: given $\boldsymbol{y} \in \mathbb{Q}^m$, $A \in \mathbb{Q}^{m\times n}$, a positive rational number ϵ and a positive integer K, does there exist $\boldsymbol{x} \in \mathbb{R}^n$ such that $\|\boldsymbol{y} - A\boldsymbol{x}\|_2 \leq \epsilon$ and $\|\boldsymbol{x}\|_0 \leq K$? This decision problem should be in NP since if one can guess a rational solution $\boldsymbol{x}$, it can be verified in polynomial time if $\|\boldsymbol{y} - A\boldsymbol{x}\|_2 \leq \epsilon$ and $\|\boldsymbol{x}\|_0 \leq K$. Similarly, we conjecture that the decision version of other optimization problems mentioned in the paper are in NP as well.

Another perspective is the approximability of aforementioned NP-hard problems. The hardness of approximating $(\ell_0 C)$ was discussed in [21, 22]. It was shown that approximating $(\ell_0 C)$ to within a factor of $(1-\alpha)\ln(n)$, $0 < \alpha < 1$ is NP-hard [22]. Examining whether similar results can be obtained on other NP-hard problems presented in the paper would require more involved theoretical analysis, which is left for future work.

V. Acknowledgment

The authors would like to thank M. Mongeau (ENAC, Toulouse) for helpful discussions and two anonymous reviewers for constructive and valuable comments.

References

[1] M. Elad, *Sparse and redundant representations: From theory to applications in signal and image processing*, Springer, New York, Aug. 2010.

[2] B. K. Natarajan, "Sparse approximate solutions to linear systems", *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.

[3] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations", *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, Mar. 1997.

[4] X. Chen, D. Ge, Z. Wang, and Y. Ye, "Complexity of unconstrained l2-lp minimization", *Mathematical Programming*, , no. 143, pp. 371–383, 2014.

[5] X. Huo and J. Chen, "Complexity of penalized likelihood estimation", *J. Statist. Comput. Simul.*, , no. 80:7, pp. 747–759, 2010.

[6] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data", *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, June 2011.

[7] E. Esser, Y. Lou, and J. Xin, "A method for finding structured sparse solutions to nonnegative least squares problems with applications", *SIAM J. Imaging Sci.*, vol. 6, no. 4, pp. 2010–2046, Oct. 2013.

[8] T. Virtanen, J. F. Gemmeke, and B. Raj, "Active-set Newton algorithm for overcomplete non-negative representations of audio", *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 11, pp. 2277–2289, Nov. 2013.

[9] R. Bro and S. De Jond, "A fast non-negativity-constrained least squares algorithm", *J. Chemometrics*, vol. 11, no. 5, pp. 393–401, 1997.

[10] S. Petra and C. Schnörr, "Average case recovery analysis of tomographic compressive sensing", *Linear Alg. Appl.*, vol. 441, pp. 168–198, 2014.

[11] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints", *J. Mach. Learning Res.*, vol. 5, pp. 1457–1469, 2004.

[12] J. Rapin, J. Bobin, A. Larue, and J.-L. Starck, "Sparse and non-negative BSS for noisy data", *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5620–5632, Nov. 2013.

[13] A. M. Bruckstein, M. Elad, and M. Zibulevsky, "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equation", *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4813–4820, Nov. 2008.

[14] R. Peharz and F. Pernkopf, "Sparse nonnegative matrix factorization with l0 constraints", *Neurocomputing*, vol. 80, pp. 38–46, 2012.

[15] M. Yaghoobi, D. Wu, and M. E. Davies, "Fast non-negative orthogonal matching pursuit", *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1229–1233, Sept. 2015.

[16] Z. Wang, R. Zhu, K. Fukui, and J. Xue, "Cone-based joint sparse modelling for hyperspectral image classification", *Signal Processing*, vol. 144, pp. 417–429, 2018.

[17] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, New York, 1979.

[18] J. v. Leeuwen, *Handbook of theoretical computer science (vol. A): algorithms and complexity*, Elsevier Science Publishers, 1990.

[19] X. Huo and X. Ni, "When do stepwise algorithms meet subset selection criteria?", *Annals Statist.*, vol. 35, no. 2, pp. 870–887, 2007.

[20] M. Nikolova, "Description of the minimizers of least squares regularized with ℓ_0 norm. Uniqueness of the global minimizer", *SIAM J. Imaging Sci.*, vol. 6, no. 2, pp. 904–937, May 2013.

[21] E. Amaldi and V. Kann, "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems", *Theoretical Computer Science*, vol. 209, no. 1, pp. 237 – 260, 1998.

[22] A. M. Tillmann, "On the computational intractability of exact and approximate dictionary learning", *IEEE Signal Processing Letters*, vol. 22, no. 1, pp. 45–49, Jan 2015.