



**HAL**  
open science

## Sign preservation analysis of orthogonal greedy algorithms

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

► **To cite this version:**

Thi Thanh Nguyen, Charles Soussen, Jérôme Idier, El-Hadi Djermoune. Sign preservation analysis of orthogonal greedy algorithms. 2020. hal-01971697v2

**HAL Id: hal-01971697**

**<https://hal.science/hal-01971697v2>**

Preprint submitted on 30 Jul 2020 (v2), last revised 20 Jul 2021 (v5)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sign preservation analysis of orthogonal greedy algorithms

Thanh T. Nguyen, Charles Soussen, Jérôme Idier, *Member, IEEE*,  
and El-Hadi Djermoune, *Member, IEEE*

## Abstract

We bring a contribution to the exact recovery theory of a non-negative  $K$ -sparse vector from noisy linear measurements under the condition  $\mu < 1/(2K - 1)$ , where  $\mu$  denotes the mutual coherence. While it is known that Orthogonal Matching Pursuit (OMP) and Orthogonal Least Squares (OLS) identify the true support in  $K$  iterations, we prove that the weights of the selected atoms have the correct sign in the best current approximation at any of the  $K$  iterations. Therefore, OMP and OLS identify with their sign-aware versions, which allows us to establish for the first time an exact support recovery property based on mutual coherence for non-negative versions of OMP and OLS.

## Index Terms

Greedy Algorithms; Orthogonal Matching Pursuit; Orthogonal Least Squares; Non-Negativity; Mutual Coherence; Sparse Non-Negative Least Squares.

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.

T. T. Nguyen and E.-H. Djermoune are with the Université de Lorraine and CNRS at the Centre de Recherche en Automatique de Nancy, F-54506 Vandœuvre-lès-Nancy, France (e-mail: thanh59hnue@gmail.com, el-hadi.djermoune@univ-lorraine.fr).

C. Soussen was with CRAN. He is now with Université Paris-Saclay, CNRS, and CentraleSupélec at the Laboratoire des Signaux et Systèmes, F-91192 Gif-sur-Yvette, France (e-mail: charles.soussen@centralesupelec.fr).

J. Idier is with CNRS at the Laboratoire des Sciences du Numérique de Nantes (UMR 6004), F-44321 Nantes, France (e-mail: jerome.idier@ls2n.fr).

## I. INTRODUCTION

Orthogonal Matching Pursuit (OMP) [1] and Orthogonal Least Squares (OLS) [2] are well known greedy, iterative algorithms for sparse decomposition<sup>1</sup>. Their common principle is to sequentially select atoms from a given dictionary, that produce a maximal decrease of the residual error in the least square sense. Finding the  $K$  atoms that jointly minimize the residual error at a given sparsity level  $K$  is an NP-hard subset selection problem [6]. Greedy algorithms such as OMP and OLS form a family of approximate schemes, with a relatively low computing cost compared to exact methods [7], while convex relaxation yields another important branch of approximate methods [8], [9].

OMP and OLS only differ by their atom selection rule. For OMP, it simply consists in maximizing the magnitude of the inner product between the residual error and the candidate atoms, assumed normalized. The OLS rule can be interpreted in similar terms, but the involved atoms are renormalized, projected<sup>2</sup> versions of the candidate atoms [10], [11].

Orthogonal greedy algorithms achieve exact support recovery under certain conditions. The Mutual Incoherence Property (MIP) states that OMP is guaranteed to recover the  $K$ -sparse representation of a data signal in the noiseless case if the mutual coherence of the dictionary is less than  $\frac{1}{2K-1}$  [12], [13]. Much attention was paid to variations over such a condition. In particular, extensions to noisy cases were considered, see, *e.g.*, [14], [15], and relaxations given partial information on the decomposition of the data signal were proposed [13], [16]. Other studies about the theoretical assessment of OMP also consider the Restricted Isometry Property (RIP) [17]–[21]. Exact support recovery properties of OLS have been studied more recently. In [13], the MIP condition was extended to OLS in the noiseless case, while [22]–[24] are based on RIP assumptions. In [11], [13], [16], OMP and OLS were treated as two instances of a unique orthogonal greedy scheme, and the generic acronym Oxx was used to refer to both. We will adopt the same convention in this paper.

In the traditional  $K$ -step analysis, exact support recovery holds when each iteration of Oxx necessarily selects an atom in the true support  $S^*$ , so that  $S^*$  is recovered after  $K = \text{card}(S^*)$  iterations. Our main contribution is to unveil a sign preservation property regarding the weights of selected atoms in the best current approximation. Under the MIP, we show that at any iteration  $k \in \{1, \dots, K\}$ , the  $k$  coefficients found by Oxx have the same sign as the corresponding ones in the true decomposition of the data signal.

<sup>1</sup>In the literature, OLS is also known as Order Recursive Matching Pursuit (ORMP) [3], Optimized Orthogonal Matching Pursuit (OOMP) [4], and Pure Orthogonal Matching Pursuit [5].

<sup>2</sup>onto the orthogonal complement of the already selected ones.

As an immediate consequence of this sign preservation property, we can theoretically assess exact support recovery properties for the non-negative versions of OMP and other non-negative greedy schemes. Non-negative OMP (NNOMP) was first introduced by Bruckstein *et al.* [25] under the name OMP, and then renamed NNOMP in [26] (see also [27], [28]). It relies on the repeated maximization of the positive inner product between the residual error and the dictionary atoms, followed by the resolution of a non-negative least square problem. For such sign constrained version of OMP, existing exact support recovery analyses are rare, and somewhat discordant. On the one hand, Bruckstein *et al.* claimed that a MIP type property holds for NNOMP [25], and that the proof should be similar to the one given in [12], [29] for OMP. Specifically, [25, Theorem 3] states without proof that  $\mu < \frac{1}{2K-1}$  is a sufficient condition for exact recovery of any  $K$ -sparse representation using NNOMP. On the other hand, Kim *et al.* elaborated a unified MIP analysis of NNOMP and its generalized version in the multiple measurement vector setting [27, Theorem 1]. In the specific case of NNOMP, that is, for single measurement vectors, the authors acknowledged that their MIP condition turns out to be very restrictive. Indeed,  $\mu$  is required to be lower than  $\frac{1}{K-1} - \frac{1}{2}$ , which can occur only when  $K \in \{1, 2\}$ .

Indeed, we think that it is impossible to prove [25, Theorem 3] as a direct extension of [12], [29], as claimed by Bruckstein *et al.* The major obstacle is that the NNOMP selection rule performs comparisons between *signed* inner products, whereas a small mutual coherence condition yields a bound on the unsigned magnitude of inner products (see Section II-D for further details). This is precisely why Kim *et al.*'s analysis, closely following Donoho *et al.*'s approach [29], yields over-pessimistic guarantees, instead of the expected result. Fortunately, we found that the sign preservation property of OMP implies that OMP and NNOMP coincide in the exact support recovery regime, so [25, Theorem 3] becomes a trivial byproduct of our sign preservation analysis. Under the same conditions, OLS and Non-negative OLS (NNOLS) [30] coincide, which allows us to extend [25, Theorem 3] to NNOLS; NNOLS and Suboptimal NNOLS (SNNOLS) [30] also coincide, so the same applies to SNNOLS. Furthermore, our analysis is not restricted to the noiseless case but also applies to noisy cases in which the noise is bounded.

The article is organized as follows. Section II introduces useful notations and briefly recalls known results about orthogonal greedy algorithms and their non-negative versions. Section III contains our main results on sign preservation of Oxx, as well as their consequences on sign-constrained greedy algorithms. The key result, stated as Theorem III.1, is proved in the same section, most technical steps being postponed in Appendix. Section IV contains a limited set of simulations to illustrate the average behavior of algorithms outside the exact support recovery regime. Section V includes concluding remarks and extensive discussions.

## II. NOTATIONS AND BACKGROUND

## A. Notations

Let us denote by  $\mathbf{y} \in \mathbb{R}^m$  the data signal and by  $A \in \mathbb{R}^{m \times n}$  the dictionary of elementary atoms  $\mathbf{a}_i \in \mathbb{R}^m$ ,  $1 \leq i \leq n$ . We are interested in the representation  $\mathbf{y} = A\mathbf{x}$  in which the vector solution  $\mathbf{x} \in \mathbb{R}^n$  has  $K$  non-zero elements. We call this a  $K$ -sparse representation. We will further state our results for noisy  $K$ -sparse representations  $\mathbf{y} = A\mathbf{x} + \mathbf{n}$ . Without loss of generality, the elementary atoms are assumed to be normalized, that is  $\|\mathbf{a}_i\|_2 = 1$ . Hereafter, the  $\ell_2$  norm will also be denoted by  $\|\cdot\|$ . Notations  $^t$  and  $^\dagger$  stand for the transpose and the Moore-Penrose pseudo-inverse, respectively. For any set of indices  $S \subset \{1, \dots, n\}$ , the subdictionary and subvector indexed by  $S$  are respectively denoted by  $A_S$  and  $\mathbf{x}_S$ . We denote by  $\text{span}(A_S)$  the subspace spanned by the dictionary atoms indexed by  $S$ , by  $\text{span}(A_S)^\perp$  the orthogonal complement of  $\text{span}(A_S)$ , and by  $P_S^\perp = I_m - A_S A_S^\dagger$  the orthogonal projection onto  $\text{span}(A_S)^\perp$ , where  $I_m$  stands for the identity matrix of size  $m \times m$ . We place the convention that  $P_S^\perp = I_m$  when  $S = \emptyset$ . When  $A_S$  is full column rank, one has a further explicit formulation  $A_S^\dagger = (A_S^t A_S)^{-1} A_S^t$ . Then we denote by  $\tilde{\mathbf{a}}_i^S = P_S^\perp \mathbf{a}_i$  the projected atoms. Clearly,  $\tilde{\mathbf{a}}_i^S = \mathbf{0}$  whenever  $i \in S$ . Let  $\tilde{\mathbf{b}}_i^S = \tilde{\mathbf{a}}_i^S / \|\tilde{\mathbf{a}}_i^S\|$ , with the convention that  $\tilde{\mathbf{b}}_i^S = \mathbf{0}$  when  $\tilde{\mathbf{a}}_i^S = \mathbf{0}$ . Later, we also use the generic notation

$$\tilde{\mathbf{c}}_i^S = \begin{cases} \tilde{\mathbf{a}}_i^S & \text{(OMP case),} \\ \tilde{\mathbf{b}}_i^S & \text{(OLS case)} \end{cases} \quad (1)$$

to refer to either  $\tilde{\mathbf{a}}_i^S$  or  $\tilde{\mathbf{b}}_i^S$  depending on the context. We will denote by  $\tilde{A}^S$  (resp.,  $\tilde{B}^S$  and  $\tilde{C}^S$ ) the matrix of size  $m \times n$  formed by gathering all projected atoms  $\tilde{\mathbf{a}}_i^S$  (resp.,  $\tilde{\mathbf{b}}_i^S$  and  $\tilde{\mathbf{c}}_i^S$ ).

The residual vector and the support found by orthogonal greedy algorithms at iteration  $k$  will be denoted  $\mathbf{r}_k$  and  $S_k$ , respectively, with  $\text{card}(S_k) = k$ . By extension, we will denote  $\mathbf{r}_0 = \mathbf{y}$  and  $S_0 = \emptyset$ . Whenever unambiguous, we will use the simpler notations  $\mathbf{r}$  and  $S$ .

In our analysis, we make use of the mutual coherence of the dictionary, defined by  $\mu(A) = \max_{i \neq j} |\mathbf{a}_i^t \mathbf{a}_j|$ . This quantity tells us how much the dictionary atoms look alike.

## B. OMP and OLS

OMP and OLS address the following minimization problem:

$$\min_{\|\mathbf{x}\|_0 \leq K} \|\mathbf{y} - A\mathbf{x}\|^2 \quad (2)$$

where the  $\ell_0$ -“norm”  $\|\cdot\|_0$  counts the number of nonzero elements. Both algorithms start with a zero vector solution corresponding to the empty support. At each iteration, a new atom  $\ell$  is selected and added

---

**Algorithm 1:** Oxx in  $K$  steps.

---

**Input:**  $\mathbf{y}$ ,  $A$ **Output:**  $\hat{\mathbf{x}}$  with  $\|\hat{\mathbf{x}}\|_0 \leq K$ 1 Initialization:  $\hat{\mathbf{x}} \leftarrow \mathbf{0}$ ;  $S \leftarrow \emptyset$ ;2 **repeat**3      $\mathbf{r} \leftarrow \mathbf{y} - A_S \hat{\mathbf{x}}_S$ ;4     Choose an atom  $\ell$  according to either (3) or (4);5      $S \leftarrow S \cup \{\ell\}$ ;6      $\hat{\mathbf{x}}_S \leftarrow A_S^\dagger \mathbf{y}$ ;7 **until**  $\text{card}(S) = K$ ;

to the support. This process is repeated until the support of cardinality  $K$  is reached (see Algorithm 1). OMP and OLS share the same coefficient update and only differ by their selection rule:

$$\text{OMP: } \ell \in \arg \max_{i \notin S} |\mathbf{r}^\dagger \mathbf{a}_i|, \quad (3)$$

$$\text{OLS: } \ell \in \arg \max_{i \notin S} |\mathbf{r}^\dagger \tilde{\mathbf{b}}_i^S| \quad (4)$$

where  $\mathbf{r} = P_S^\perp \mathbf{y}$  denotes the current residual vector. Note that (3) can also be written as

$$\ell \in \arg \max_{i \notin S} |\mathbf{r}^\dagger \tilde{\mathbf{a}}_i^S| \quad (5)$$

since  $\mathbf{r}$  is orthogonal to  $\text{span}(A_S)$ . These inner product expressions come from the geometrical interpretation of OMP and OLS [10]. Besides, from an optimization viewpoint, the selection rule of OMP is based on the minimization of  $\|\mathbf{r} - z_i \mathbf{a}_i\|^2$  w.r.t. scalar  $z_i$ , whereas OLS relies on the minimization of  $\|\mathbf{y} - A_{S \cup \{i\}} \mathbf{z}\|^2$  w.r.t. vector  $\mathbf{z}$  of length  $\text{card}(S) + 1$  [10]. From this point of view, it is obvious that the OLS selection rule is more costly than that of OMP, although recursive update schemes are available [2], [4].

Let us state the MIP condition that holds for both OMP and OLS when  $\mu(A) < \frac{1}{2K-1}$ . Note that under this assumption, any  $K$ -sparse representation is unique [31, Theorem 7]. In [12, Theorem 3.1-3.5] and [13, Theorem 2], it was shown that when  $\mu(A) < \frac{1}{2K-1}$ , the support of any  $K$ -sparse representation  $\mathbf{y} = A\mathbf{x}^*$  can be exactly recovered by Oxx algorithms in  $K$  iterations. The latter noiseless analyses were then extended to noisy settings for different types of additive noise including bounded noise and Gaussian noise, see, *e.g.*, [15] and [16] for OMP and OLS analyses, respectively. Let us recall the results in [15],

---

**Algorithm 2:** Non-negative Oxx for  $K$ -sparse non-negative recovery.

---

**Input:**  $\mathbf{y}$ ,  $A$

**Output:**  $\hat{\mathbf{x}} \geq \mathbf{0}$  with  $\|\hat{\mathbf{x}}\|_0 \leq K$

1 Initialization:  $S \leftarrow \emptyset$ ,  $\hat{\mathbf{x}} \leftarrow \mathbf{0}$ ;

2 **repeat**

3      $\mathbf{r} \leftarrow \mathbf{y} - A_S \hat{\mathbf{x}}_S$  ;

4     Choose an atom  $\ell$  according to prescribed selection rule (SR) in Table I;

5      $S \leftarrow S \cup \{\ell\}$ ;

6      $\hat{\mathbf{x}}_S \leftarrow \arg \min_{\mathbf{z} \geq \mathbf{0}} \|\mathbf{y} - A_S \mathbf{z}\|^2$ ;

7 **until**  $\text{card}(S) = K$  or prescribed stopping condition (SC) in Table I;

---

[16] in the case of a bounded noise satisfying

$$\|\mathbf{n}\| < \frac{1 - (2K - 1)\mu(A)}{2} \min\{|x_i^*|, x_i^* \neq 0\}. \quad (6)$$

**Lemma II.1.** [15, Theorem 1], [16, Theorem 4]. Assume that  $\mu(A) < \frac{1}{2K-1}$ . Let  $\mathbf{y} = A\mathbf{x}^* + \mathbf{n}$  where  $\mathbf{x}^*$  is  $K$ -sparse and the noise term  $\mathbf{n}$  satisfies (6). Then Oxx recover the support of  $\mathbf{x}^*$  in  $K$  iterations.

Obviously, Lemma II.1 implies that in the noise-free case, exact support recovery is guaranteed whenever  $\mu(A) < \frac{1}{2K-1}$ . Note that the condition given in [16, Theorem 4] is slightly weaker than the one of Lemma II.1. Condition (6) imposes that the minimum nonzero coefficient in  $\mathbf{x}^*$  is bounded, which can be interpreted as a flat vector analysis. In [16, Theorem 4], it was shown that the latter conditions can be improved for non-flat vectors, *e.g.*, vectors with decaying amplitudes.

Here and throughout the paper, we consider that in the special case where the Oxx selection rules (3)-(4) yield several solutions (*i.e.*, two atoms  $i_1$  and  $i_2$  for which  $|\mathbf{r}^t \tilde{\mathbf{c}}_{i_1}^S| = |\mathbf{r}^t \tilde{\mathbf{c}}_{i_2}^S|$ ) including an atom that does not belong to the support of  $\mathbf{x}^*$ , Oxx makes the wrong decision and hence  $K$ -step exact recovery does not occur.

### C. Non-negative setting

In many fruitful applications of sparse decomposition such as multispectral unmixing [32], microarray data analysis [33], mass spectroscopy [34] and fluid mechanics [35], to name a few, the sought solution is required to be non-negative. Several non-negative extensions of OMP and OLS have been studied so far, the main being NNOMP [25], [26], NNOLS and SNNOLS [30]. All three solve non-negative least squares

TABLE I

NON-NEGATIVE GREEDY SCHEMES EXISTING IN THE LITERATURE, WITH RELATED SELECTION RULES (SR) AND STOPPING CONDITIONS (SC).

NNOMP [25]	(SR) $\arg \max_{i \notin S} \mathbf{r}^t \mathbf{a}_i$ (SC) $\max_{i \notin S} \mathbf{r}^t \mathbf{a}_i \leq 0$
SNNOLS [30]	(SR) $\arg \max_{i \notin S} \mathbf{r}^t \tilde{\mathbf{b}}_i^S$ (SC) $\max_{i \notin S} \mathbf{r}^t \tilde{\mathbf{b}}_i^S \leq 0$
NNOLS [30]	(SR) $\arg \min_{i \notin S} \left( \min_{\mathbf{z} \geq \mathbf{0}} \ \mathbf{y} - A_{S \cup \{i\}} \mathbf{z}\ ^2 \right)$ (SC) $\min_{i \notin S} \left( \min_{\mathbf{z} \geq \mathbf{0}} \ \mathbf{y} - A_{S \cup \{i\}} \mathbf{z}\ ^2 \right) = \ \mathbf{r}\ ^2$

(NNLS) subproblems to update the selected coefficients at each iteration according to Algorithm 2, while they differ at the atom selection stage, see Table I. NNOMP adopts the selection rule of OMP by simply removing the modulus, *i.e.*, ignoring negative inner products. Similarly, the selection rule of SNNOLS is the same as OLS without modulus. On the contrary, NNOLS solves as many NNLS subproblems as candidate atoms to find the largest decrease of the residual error. In the case of OLS, the inner product formulation (4) is equivalent to the residual error minimization approach since the residual error is orthogonal to the column space of selected atoms, see [4], [10]. In contrast, when solving an NNLS problem, the orthogonality is not guaranteed anymore (because of the possible activation of non-negativity constraints), so the selection rules of SNNOLS and NNOLS are not equivalent in general. In [28], the reader can find a detailed study about the design and practical implementation of non-negative orthogonal greedy algorithms. As compared with basic implementations, huge possibilities of accelerations can be made possible for NNOLS using a fully recursive implementation, assorted with a dynamic pruning strategy. To a lesser extent, this also applies to NNOMP and to SNNOLS.

#### D. Exact recovery analysis in the non-negative setting

1) *Extension of K-step exact recovery analysis of OMP:* Bruckstein *et al.* [25] claimed that the  $K$ -step exact recovery analysis of NNOMP can be carried out as a straightforward extension of the classical  $K$ -step exact recovery analyses of OMP [12], [29]. Here, we argue that this extension does not actually appear to be trivial.

Tropp's reasoning in [12] consists in lower bounding the modulus of the inner product  $|\mathbf{r}^t \mathbf{a}_i|$  between the residual and the correct dictionary atoms based on induced matrix norm identities. Unfortunately,

similar lower bounds cannot be directly extended when the modulus is dropped. Donoho *et al.*'s recursive proof [29] exploits that for noiseless sparse inputs, the data residual lays in the subspace spanned by the true atoms. So the result at the first iteration (guaranteed selection of a true atom) can be exploited again in the next iterations by replacing the original input signal by the residual. To generalize this reasoning to the non-negative case, one would need to prove that for non-negative sparse decompositions  $\mathbf{y} = A\mathbf{x}^*$  (*i.e.*, inputs laying in the positive span of the true atoms), the residual lays in the positive span of the dictionary atoms as well. This conjecture turns out to be false; if true atoms are selected until iteration  $k$ , the residual  $\mathbf{r} = A(\mathbf{x}^* - \hat{\mathbf{x}}^{(k)})$  lays in the subspace spanned by the true atoms, where  $\hat{\mathbf{x}}^{(k)}$  denotes the NNOMP iterate. Provided that the latter atoms are linearly independent, it is clear that  $\mathbf{r}$  lays in their positive span if and only if  $\hat{\mathbf{x}}^{(k)} \leq \mathbf{x}^*$ . Simple numerical tests show that even for toy problems,  $\hat{\mathbf{x}}^{(k)} \leq \mathbf{x}^*$  may not hold when  $\mathbf{x}^* \geq \mathbf{0}$  and  $\mu(A) < \frac{1}{2K-1}$ : see § IV-B and Fig. 2(a). Therefore, sign-preservation does not apply to the residual vector, hence Donoho *et al.*'s reasoning cannot be directly extended to the non-negative case.

2)  $\ell_1$  analysis with non-negativity constraints: Unlike the analysis of orthogonal greedy algorithms, the theoretical analysis of non-negative versions of Basis Pursuit (BP:  $\min_{\mathbf{x}} \|\mathbf{x}\|_1$  s.t.  $\mathbf{y} = A\mathbf{x}$ ) and Basis Pursuit Denoising (BPDN:  $\min_{\mathbf{x}} \|\mathbf{y} - A\mathbf{x}\|^2 + \lambda\|\mathbf{x}\|_1$ ) is closely linked to the standard analysis of BP and BPDN. Indeed, it is well-known that contrary to greedy algorithms, the exact recovery analysis of BP heavily depends on the sign pattern. Fuchs [36] proved that when  $\mathbf{y} = A\mathbf{x}^*$ , BP and BPDN (for sufficiently small  $\lambda$ ) both have a unique solution under the MIP assumption, and that this solution identifies with  $\mathbf{x}^*$  as long as  $A_{S^*}$  is full rank and

$$\forall j \notin S^*, |\langle \boldsymbol{\sigma}^*, A_{S^*}^\dagger \mathbf{a}_j \rangle| < 1 \quad (7)$$

where  $S^* := \text{supp}(\mathbf{x}^*)$  and  $\boldsymbol{\sigma}^* := \text{sign}(\mathbf{x}_{S^*}^*) \in \{-1, 1\}^K$  denotes the sign pattern. It is noticeable that the latter condition<sup>3</sup> only depends on the sign pattern, but not on the magnitudes of coefficients  $x_i^*$ . Let  $\boldsymbol{\sigma}^*$  equal to the all-ones vector of size  $K$ . Condition (7) thus guarantees exact recovery of  $\mathbf{x}^*$  for any  $\mathbf{x}^* \geq \mathbf{0}$  supported by  $S^*$ . It follows that  $\mathbf{x} = \mathbf{x}^*$  is not only the unique solution of BP/BPDN but also a solution of their non-negative counterpart

$$\min_{\mathbf{x} \geq \mathbf{0}} \|\mathbf{y} - A\mathbf{x}\|^2 + \lambda\|\mathbf{x}\|_1, \quad (8)$$

<sup>3</sup>referred to as Fuchs corollary condition in [37] as opposed to the sharp (but more involved) Fuchs condition in [37, Theorem 2]. The latter condition solely depends on the sign pattern as well.

often referred to as the non-negative Garrote in statistics [38]. Indeed, the minimum value of  $\|\mathbf{y} - A\mathbf{x}\|^2 + \lambda\|\mathbf{x}\|_1$  over  $\mathbf{x} \geq \mathbf{0}$  can be reached only for  $\mathbf{x} = \mathbf{x}^*$ . Summarizing, the sign preservation property of BP is guaranteed as long as Fuchs condition (7) is met.

3) *Extension of  $K$ -step exact recovery analysis of  $\ell_1$  homotopy:* Homotopy is a popular greedy algorithm dedicated to  $\ell_1$  minimization. It was originally proposed in [39] in the undercomplete setting, and named modified Least Angle Regression (LARS). It was later renamed homotopy by Donoho and Tsaig, and further analyzed in the overcomplete setting under the MIP assumption in the same paper [9]. Homotopy aims to solve the BPDN problem  $\min_{\mathbf{x}} \|\mathbf{y} - A\mathbf{x}\|^2 + \lambda\|\mathbf{x}\|_1$  for a continuum of  $\lambda$ 's. The principle of homotopy is to reconstruct the regularization path (defined as the set of solutions for all  $\lambda$ ) by solving the  $\ell_1$  problem for gradually decreasing  $\lambda$ 's starting from  $\lambda = +\infty$ , with the corresponding solution  $\mathbf{x} = \mathbf{0}$ . Homotopy has a stepwise mechanism with an atom selection or deselection at each iteration. It turns out [9, Theorem 1] that when  $\mu(A) < \frac{1}{2K-1}$ ,  $K$ -step recovery of any  $K$ -sparse vector  $\mathbf{x}^*$  from noise-free observations  $\mathbf{y} = A\mathbf{x}^*$  is guaranteed. Not only the support but also the magnitudes of  $\mathbf{x}^*$  are found after  $K$  iterations (no deselection occurs).

Although homotopy has a greedy structure, a major difference with Oxx algorithms is that homotopy solves an  $\ell_1$ -penalized least square problem at each iteration. Therefore, the related exact recovery analysis significantly differs from that of Oxx. The exact recovery analysis in [9, Theorem 1] is based on two ingredients: the correct selection of atoms indexed by  $\text{supp}(\mathbf{x}^*)$  and the so-called *sign agreement* property. The latter ensures that elements from the solution support are never removed, so the deselection event never occurs. Sign agreement is defined as follows: *at every iteration  $k$ , the homotopy iterate  $\hat{\mathbf{x}}$  satisfies  $\text{sign}(\hat{\mathbf{x}}_S) = \text{sign}(A_S^t \mathbf{r})$ , where  $S = \text{supp}(\hat{\mathbf{x}})$ .* Donoho and Tsaig showed that when  $\mu(A) < \frac{1}{2K-1}$ , the magnitudes of the active atoms always increase while  $\lambda$  is decreasing. Since the correct magnitudes are exactly found at iteration  $K$  ( $\hat{\mathbf{x}} = \mathbf{x}^*$ ), it follows that their sign is unchanged throughout the iterations. So, a sign preservation result in the  $\ell_1$  case is truly obtained as a byproduct of [9, Theorem 1]. In the Oxx setting, the sign preservation result will be obtained here using a brandnew reasoning. Moreover, the sign agreement property “ $\text{sign}(\hat{\mathbf{x}}_S) = \text{sign}(A_S^t \mathbf{r})$ ” does not make sense anymore since  $A_S^t \mathbf{r} = \mathbf{0}$ , due to the orthogonality between the residual vector and the selected atoms.

### III. EXACT RECOVERY AND SIGN PRESERVATION

This section contains our main results concerning sign preservation of Oxx when the exact recovery condition in terms of mutual coherence is met. The cornerstone of our study is Theorem III.1, while the other results of Subsection III-A are rather direct consequences. In Subsection III-B, the proof of

Theorem III.1 is decomposed into distinct steps, most technical elements being postponed in Appendix.

### A. Main results

The sign preservation property for arbitrary  $K$ -sparse representations is stated now.

**Theorem III.1.** *Assume that  $\mu(A) < \frac{1}{2K-1}$ . Let  $\mathbf{y} = A\mathbf{x}^* + \mathbf{n}$  be a noisy  $K$ -sparse representation, where  $\mathbf{n}$  satisfies (6). Then Oxx recovers the support of  $\mathbf{x}^*$  in  $K$  iterations, and at each iteration, the weights of selected atoms are of the same sign as the corresponding magnitudes of  $\mathbf{x}^*$ .*

An obvious corollary can be stated in the special case of  $K$ -sparse representations with non-negative weights.

**Corollary III.1.** *Assume that  $\mu(A) < \frac{1}{2K-1}$ . Let  $\mathbf{y} = A\mathbf{x}^* + \mathbf{n}$  be a noisy  $K$ -sparse representation, where  $\mathbf{x}^* \geq \mathbf{0}$  and  $\mathbf{n}$  satisfies (6). Then, Oxx recovers the support of  $\mathbf{x}^*$  in  $K$  iterations, and at each iteration, the weights of selected atoms are positive.*

In the non-negative setting, Corollary III.1 has interesting implications concerning non-negative versions of Oxx. Let us start with the following lemma.

**Lemma III.1.** *Let  $\mathbf{y} \in \mathbb{R}^m$  and assume that any combination of  $K$  dictionary columns is linearly independent.*

- *Assume that at every iteration  $k = 1, \dots, K$  of OMP, the selection rule yields a unique optimal index  $\ell$  and that the weights of selected atoms are all positive. Then NNOMP provides the same iterates as OMP (i.e., with the same support and the same weights at every iteration).*
- *The same applies if OMP and NNOMP are replaced by OLS and SNNOLS, respectively.*
- *It also applies if OMP and NNOMP are replaced by OLS and NNOLS, respectively.*

This result is intuitive since at each iteration of Oxx, the weights of selected atoms form a vector defined as a least squares solution. Clearly, if an unconstrained least squares solution is positive, then it is also the solution of the corresponding NNLS problem.

*Proof:* See Appendix VI-C. ■

From Corollary III.1 and Lemma III.1, we can deduce the following result.

**Corollary III.2.** *Assume that  $\mu(A) < \frac{1}{2K-1}$ . Let  $\mathbf{y} = A\mathbf{x}^* + \mathbf{n}$  be a noisy  $K$ -sparse representation, where  $\mathbf{x}^* \geq \mathbf{0}$  and  $\mathbf{n}$  satisfies (6). Then NNOMP identifies with OMP whereas both NNOLS and SNNOLS identify with OLS. Thus, NNOMP, NNOLS and SNNOLS all recover the support of  $\mathbf{x}^*$  in  $K$  iterations.*

It is known that the MIP condition  $\mu(A) < \frac{1}{2K-1}$  is not only sufficient but also necessary for *uniform* (i.e., irrespective of the magnitudes of nonzero coefficients in the sparse representation and of the choice of the dictionary)  $K$ -step exact support recovery [13], [40] by Oxx. Cai *et al.* [40] indeed exhibited an equiangular dictionary whose mutual coherence equals  $\mu(A) = \frac{1}{2K-1}$  and a vector  $\mathbf{y}$  having two  $K$ -sparse representations  $\mathbf{y} = A\mathbf{x} = A\mathbf{z}$  with distinct supports.  $K$ -step exact support recovery does not make sense anymore in this situation, since either the support of  $\mathbf{x}$  or  $\mathbf{z}$  cannot be reconstructed in  $K$  steps. The same analysis can be made concerning non-negative extensions of Oxx.

**Corollary III.3.** *The MIP condition  $\mu(A) < \frac{1}{2K-1}$  is necessary for  $K$ -step exact recovery of any non-negative  $K$ -sparse vector by non-negative orthogonal greedy algorithms, since there exists a dictionary  $A$  with  $\mu(A) = \frac{1}{2K-1}$  and a vector  $\mathbf{y}$  having two non-negative  $K$ -sparse representations with distinct supports.*

*Proof:* Consider the dictionary  $A \in \mathbb{R}^{m \times n}$  with  $\mu(A) = \frac{1}{2K-1}$  and the vector  $\mathbf{y} \in \mathbb{R}^m$  proposed in [40, Section III], the latter vector having two  $K$ -sparse representations  $\mathbf{y} = A\mathbf{x} = A\mathbf{z}$  with distinct supports. Since both supports are distinct, one can define the subrogate dictionary  $A' \in \mathbb{R}^{m \times n}$  as:

$$\mathbf{a}'_i = \begin{cases} \text{sign}(x_i) \mathbf{a}_i & \text{if } i \in \text{supp}(\mathbf{x}), \\ \text{sign}(z_i) \mathbf{a}_i & \text{if } i \in \text{supp}(\mathbf{z}), \\ \mathbf{a}_i & \text{otherwise.} \end{cases} \quad (9)$$

Moreover, let  $|\mathbf{x}|$  and  $|\mathbf{z}| \in \mathbb{R}^n$  denote the non-negative vectors whose entries are equal to  $|x_i|$  and  $|z_i|$ , respectively. Obviously,  $\mathbf{y} = A'|\mathbf{x}| = A'|\mathbf{z}|$ , and  $\mu(A') = \mu(A) = \frac{1}{2K-1}$ . ■

### B. Proof of Theorem III.1

Let us remark that it is sufficient to consider the case of non-negative weights. Indeed, an arbitrary  $K$ -sparse representation can be reduced to the case with non-negative weights, by an obvious transformation where each negative weight is replaced by its opposite value, the corresponding atom being also replaced by the opposite one. Moreover, it is straightforward to check that the list of atoms selected by Oxx is invariant through such a transformation. We therefore simply need to prove the result for non-negative weights, which corresponds to the setting of Corollary III.1.

Let us consider a noisy  $K$ -sparse representation  $\mathbf{y} = A\mathbf{x}^* + \mathbf{n}$  with positive weights and a noise term satisfying (6). From Lemma II.1, Oxx recovers the support  $S^*$  of  $\mathbf{x}^*$  in  $K$  iterations. At any iteration  $k \leq K$ , the support  $S_k$  of the current solution  $\hat{\mathbf{x}}_{S_k}^{(k)}$  is therefore a subset of  $S^*$ . Recall that  $\hat{\mathbf{x}}_{S_k}^{(k)}$  is the

unconstrained least squares solution related to support  $S_k$ , see Algorithm 1. Let  $\mathbf{r}_k = P_{S_k}^\perp \mathbf{y}$  denote the related residual vector, with  $\mathbf{r}_0 = \mathbf{y}$  corresponding to  $S_0 = \emptyset$ .

We proceed in two distinct steps to prove that  $\hat{\mathbf{x}}_{S_k}^{(k)} > \mathbf{0}$  for all  $k \in \{1, \dots, K\}$ . First, we prove that the weight of each newly selected atom  $\hat{x}_{S_k \setminus S_{k-1}}^{(k)}$  is positive for any  $k \leq K$ . Then, we show that the updated coefficients  $\hat{\mathbf{x}}_{S_{k-1}}^{(k)}$  remain positive.

Let us first characterize the last  $k - j$  coefficients of  $\hat{\mathbf{x}}_{S_k}^{(k)}$ ,  $j < k$  being an arbitrary index.

**Lemma III.2.** *Let  $\mathbf{y} \in \mathbb{R}^m$ , and let  $j$  and  $k$  be two iteration indices such that  $0 \leq j < k$ . Assume that  $A_{S_k}$  is full column rank. Then, the  $k$ -th iterate of Oxx satisfies*

$$\hat{\mathbf{x}}_{S_k \setminus S_j}^{(k)} = \left[ \tilde{A}_{S_k \setminus S_j}^{S_j} \right]^\dagger \mathbf{r}_j. \quad (10)$$

*Proof:* See Appendix VI-B. ■

Then, the two main steps of the proof of Theorem III.1 correspond to the following lemmas.

**Lemma III.3** (non-negativity of new coefficient). *Assume that  $\mu(A) < \frac{1}{2K-1}$ . Let  $\mathbf{y} = A\mathbf{x}^* + \mathbf{n}$  be a noisy  $K$ -sparse representation, where  $\mathbf{x}^* \geq \mathbf{0}$  and  $\mathbf{n}$  satisfies (6). For all  $k \in \{1, \dots, K\}$ ,  $\hat{x}_{S_k \setminus S_{k-1}}^{(k)} > 0$ .*

*Proof:* See Appendix VI-D. ■

**Lemma III.4** (non-negativity of updated coefficients). *Assume that  $\mu(A) < \frac{1}{2K-1}$ . Let  $\mathbf{y} = A\mathbf{x}^* + \mathbf{n}$  be a noisy  $K$ -sparse representation, where  $\mathbf{x}^* \geq \mathbf{0}$  and  $\mathbf{n}$  satisfies (6). For all  $k \in \{2, \dots, K\}$ ,  $\hat{\mathbf{x}}_{S_{k-1}}^{(k)} > \mathbf{0}$ .*

*Proof:* See Appendix VI-E. ■

Let us remark that the result of Lemmas III.3 and III.4 is straightforward at the last iteration  $k = K$ . Indeed, the correct support  $S^*$  is found by Oxx according to Lemma II.1. Moreover, Oxx algorithms yield the same output  $\hat{\mathbf{x}}^{(K)}$ , with  $\hat{\mathbf{x}}_{S^*}^{(K)} = A_{S^*}^\dagger \mathbf{y}$ . According to the local stability analysis performed in [29, Theorem 5.1], we have

$$\|\mathbf{x}^* - \hat{\mathbf{x}}^{(K)}\|^2 \leq \frac{\varepsilon^2}{1 - (K-1)\mu(A)} \quad (11)$$

under the assumptions  $\mu(A) < \frac{1}{2K-1}$  and  $\|\mathbf{n}\| \leq \varepsilon$ . Using the upper bound in (6):

$$\varepsilon = \frac{1 - (2K-1)\mu(A)}{2} \min\{|x_i^*|, x_i^* \neq 0\} \quad (12)$$

on can easily check<sup>4</sup> that

$$\|\mathbf{x}^* - \hat{\mathbf{x}}^{(K)}\|^2 < \left(\min_{i \in S^*} x_i^*\right)^2, \quad (13)$$

hence for all  $j \in S^*$ ,

$$|x_j^* - \hat{x}_j^{(K)}|^2 \leq \left(\min_{i \in S^*} x_i^*\right)^2 \leq \left(x_j^*\right)^2 \quad (14)$$

which proves that the entries of  $\mathbf{x}^{(K)}$  and  $\mathbf{x}^*$  have same signs.

#### IV. NUMERICAL STUDY

##### A. Comparison of Oxx and their sign-aware versions

The previous section shows that in some specifically favorable situations, greedy algorithms such as OMP have not only the capacity to recover the support of the true solution, but also to recover the correct weight signs. In such conditions, according to Lemma III.1, implementing non-negative (or more generally, sign-aware) versions of such greedy algorithms is useless. However, one cannot generalize such a conclusion to more realistic scenarios. On the contrary, one can empirically observe that sign-aware greedy algorithms tend to reach superior performance, which is in agreement with the fact that they exploit more information than usual greedy algorithms. To illustrate this fact, let us consider a dictionary  $A$  with 22 regularly spaced, discretized Gaussian-shaped atoms, with a constant standard deviation  $\sigma = 0.5$ , and a mutual coherence  $\mu(A) = 0.37$ . We randomly choose  $K = 10$  atoms in  $A$ , whose location in the dictionary are distributed according to the uniform distribution. The atoms are all equally weighted with a unit weight, and we generate data  $\mathbf{y} = A\mathbf{x}^* + \mathbf{n}$  with some additive Gaussian noise  $\mathbf{n}$ . Note that  $\mu(A) > \frac{1}{2K-1}$ , which implies that exact support recovery is not mathematically guaranteed even at low noise. Within a certain range of signal-to-noise ratio (SNR, defined by  $10 \log_{10}(\|A\mathbf{x}^*\|^2/\|\mathbf{n}\|^2)$ ), we have generated average performance for OMP, NNOMP, OLS, SNNOLS and NNOLS in terms of rate of exact support recovery, the stopping condition being that the cardinality of the estimated support should be 10. This experiment has been repeated 1000 times to obtain the average results shown in Fig. 1. Several empirical conclusions can be drawn. Some of them are already acknowledged facts. For instance, greedy algorithms keep some exact recovery capacities far beyond the “zero defect” area. In contrast, in the low SNR regime, the exact recovery capacity almost surely vanishes. The intermediate zone is the most interesting. Specifically, one can notice that there is a significant difference of performance between the usual greedy algorithms and their non-negative extensions. We have also performed a sign-preservation

<sup>4</sup>The proof is omitted for brevity.

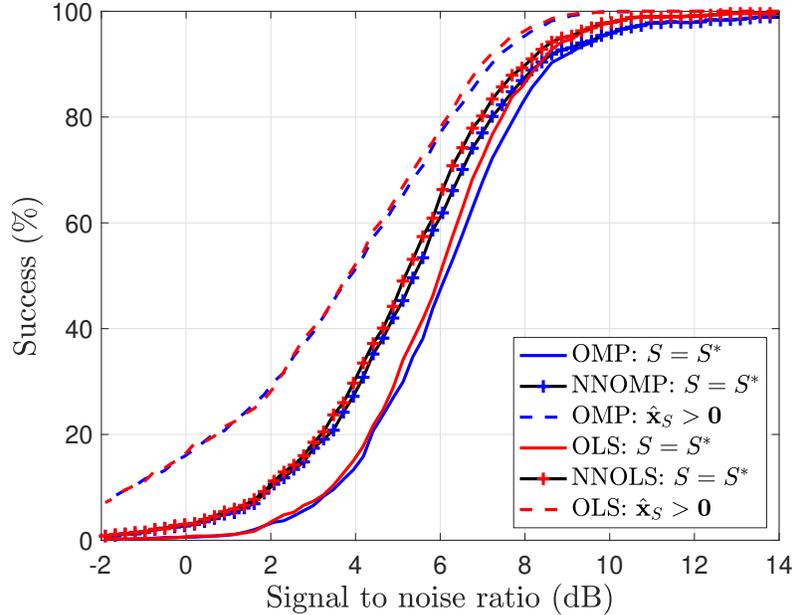


Fig. 1. Rate of exact support recovery ( $S_K = S^*$ ) and non-negativity at all iterations ( $\forall k, \hat{\mathbf{x}}^{(k)} \geq \mathbf{0}$ ) w.r.t. the signal-to-noise ratio for a simulated data experiment involving Gaussian-shaped atoms. SNNOLS (not shown here) yields the exact same curve as NNOLS.

test for OMP and OLS that simply consists in checking whether at all iterations, all estimated weights are positive. Fig. 1 shows that such a sign-preservation property is rather robust, and that empirically, sign-preservation is always reached with Oxx whenever exact support recovery is found.

### B. Non-monotony of the magnitude variations

As argued in § II-D3,  $\ell_1$  homotopy is a stepwise greedy algorithm for which the sign preservation guarantee holds whenever  $\mu(A) < \frac{1}{2K-1}$ . In [9], Donoho and Tsaig proved a stronger result under the mutual coherence condition: the magnitudes  $\hat{x}_i^{(k)}$  of the selected atoms keep increasing while  $k$  is increasing. In contrast, such monotony property does not hold for Oxx algorithms, since the magnitudes may either increase or decrease during iterations. This claim can be proven analytically at iteration 2, by using the fact that  $\hat{x}_{S_1}^{(1)} > 0$  and  $\hat{x}_{S_2}^{(2)} > \mathbf{0}$  according to Corollary III.1. Since  $\hat{\mathbf{x}}^{(1)}$  and  $\hat{\mathbf{x}}^{(2)}$  have closed-form expressions, a simple calculation (skipped for brevity reasons) shows that the first magnitude is decreasing, *i.e.*,  $\hat{x}_{S_1}^{(1)} - \hat{x}_{S_1}^{(2)} > 0$  if and only if the inner product between the atoms selected in the first two iterations is non-negative. So, the magnitude of  $\hat{x}_{S_1}$  may either increase or decrease depending on the sign of this inner product. We further compare  $\ell_1$ -homotopy with Oxx algorithms for a toy problem of dimension  $(m, n) = (5, 5)$ , with an equiangular dictionary  $A$  such that  $\mu(A) = \frac{0.9}{2K-1}$ . The columns

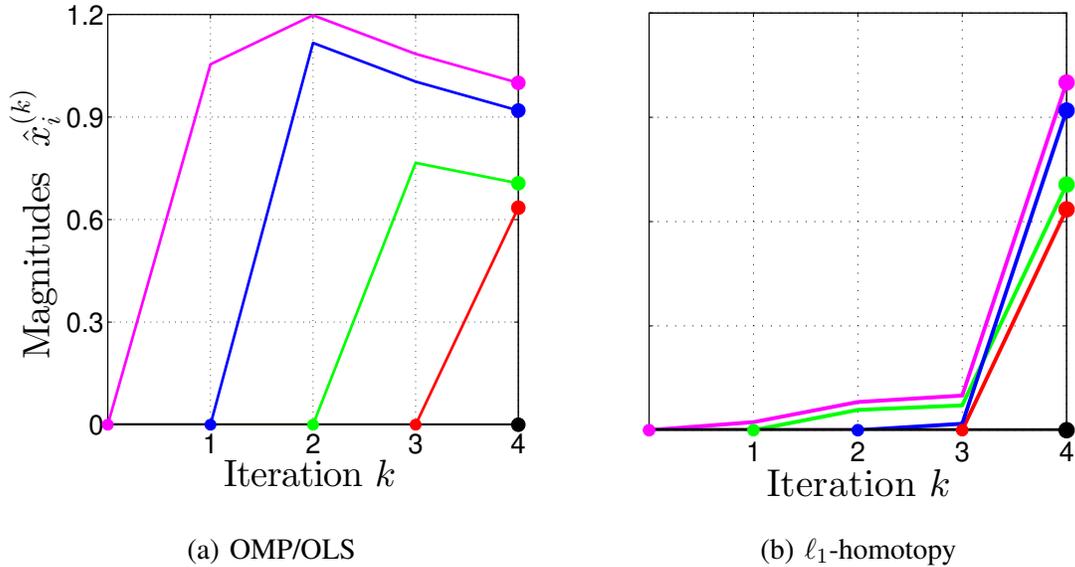


Fig. 2. Behavior of Oxx and  $\ell_1$ -homotopy in the case of a toy problem ( $m = n = 5$ ) corresponding to a 4-sparse noiseless representation. (a) The OMP and OLS iterates are identical and yield magnitudes  $\hat{x}_i^{(k)}$  with non-monotonous variations throughout the iterations  $k$ . (b) On the contrary, the magnitudes of the  $\ell_1$ -homotopy iterates are always increasing. The ground truth magnitudes  $x_i^*$  are represented with bullets for  $k = 4$ , and are exactly recovered after  $K = 4$  iterations ( $\hat{\mathbf{x}}^{(k)} = \mathbf{x}^*$ ).

of  $A$  satisfy  $\langle \mathbf{a}_i, \mathbf{a}_j \rangle = \pm\mu(A)$ , where the sign of the inner product is randomly chosen. The ground truth vector  $\mathbf{x}^*$  is  $K$ -sparse with  $K = 4$ , with nonzero magnitudes drawn from the uniform distribution  $\mathcal{U}([0.6, 1])$ . Since  $\mu(A) < \frac{1}{2K-1}$ ,  $K$ -step exact recovery holds for all considered algorithms. In Fig. 2, the variation of each entry  $\hat{x}_i^{(k)}$  with respect to  $k$  is represented with a specific color. As expected (since exact support recovery holds), the black magnitude corresponding to the wrong atom  $i \notin S^*$ , is equal to 0 throughout the iterations.  $\ell_1$ -homotopy yields magnitudes that are indeed increasing with  $k$ , which is consistent with the theoretical result in [9]. On the contrary, the OMP and OLS iterates (which are identical here; the same indices are selected at each iteration) are non-monotonous.

## V. DISCUSSION AND CONCLUSION

### A. Contributions and links with alternative algorithms

We have established that OMP and OLS satisfy the sign preservation property when dealing with sufficiently incoherent dictionaries: in the exact recovery regime  $\mu(A) < \frac{1}{2K-1}$ , the Oxx estimates are guaranteed to keep the same sign as the ground-truth sparse vector, at all iterations. This interesting property enables us to establish the first  $K$ -step recovery analysis of three non-negative greedy algorithms proposed previously, namely NNOMP, NNOLS and SNNOLS, under the MIP condition  $\mu(A) < \frac{1}{2K-1}$ .

This exact recovery condition turns out to be identical for both Oxx algorithms and their non-negative extensions. Moreover, it is not only a sufficient but also a (worst-case) necessary condition of exact recovery. Therefore, one cannot distinguish the performance of Oxx and their nonnegative counterparts from this theoretical viewpoint.

A strong feature of our analysis is that the three algorithms NNOMP, NNOLS and SNNOLS are analyzed in a unified way. In contrast, many exact recovery analyses of the literature related to OMP and OLS are done separately, because the OLS selection rule is more sophisticated than that of OMP (see the next paragraph for details regarding the related bibliography). One may wonder whether the proposed sign preservation analysis is valid for other approaches and algorithms. In Section IV-B, we pointed out that the  $\ell_1$  homotopy algorithm [9] not only enjoys a sign preservation property akin to Oxx algorithms, but moreover, the magnitudes of the selected atoms keep increasing during iterations. Consequently, the non-negativity constrained LARS (NLARS) algorithm proposed in [41] yields the same iterates as the LARS algorithm, thus the true support can be recovered in  $K$  iterations.

The analysis of the non-negative extensions of CoSaMP [42] and Subspace Pursuit (SP) [43] appears to be less straightforward. The so-called NN-CoSaMP and NN-SP algorithms proposed in [27], aim to update an estimated support of constant size  $K$  by repeatedly performing support merging, NNLS estimation of coefficients and support pruning. In [27, Theorem 1], an exact recovery analysis was carried out under mutual coherence assumptions for the multiple measurement vector problem, where one aims to recover non-negative representations of a set of vectors having a simultaneous support. The latter condition turns out to be very restrictive in the single measurement case since it can be met only when  $K \leq 2$ . Therefore, a thorough analysis using mutual coherence assumptions is still an open issue. Since the classical analyses of CoSaMP and SP are substantially different from those of greedy algorithms (for the former, exact recovery analyses usually aim to guarantee that the coefficient error  $\|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^*\|$  is decreasing at each iteration  $k$ ), we conjecture that the sign of coefficients  $\hat{\mathbf{x}}^{(k)}$  does not play such a critical role as compared to Oxx algorithms, thus the line of reasoning may be substantially different from the one elaborated here.

### *B. Towards analyses for dictionaries of higher coherence*

Empirically, we observed that the sign preservation property holds for noiseless scenarios when  $\mu(A)$  is substantially greater than  $\frac{1}{2K-1}$ . However, for truly coherent dictionaries and/or noisy observations, Oxx algorithms do yield iterates with negative entries, thus sign-aware versions are worth being considered [28]. Furthermore, the latter versions yield improved empirical exact recovery performance for random dictionaries when  $\mu(A) > \frac{1}{2K-1}$ , as shown in the numerical comparisons in [28].

The current analysis has to be understood as the first thorough analysis of non-negative greedy algorithms in terms of mutual coherence, with optimal bounds. Since real world problems usually involve dictionaries with higher coherence, let us now discuss the main challenges of deriving improved analyses for correlated dictionaries, in order to fill the gap between the theoretical understanding for low coherence dictionaries and the practitioner knowledge [28].

1) *ERC based analyses*: In the proposed analysis (Theorem III.1), the MIP property plays two distinct roles:

- the capacity to select true atoms derives from the low mutual coherence *between true and wrong atoms*;
- the sign preservation property holds because the mutual coherence is low *between true atoms*.

With a view to further distinguish the influence of true and wrong atoms and to relax the conditions of Theorem III.1, Tropp's exact recovery condition (ERC) [12] is worth being considered. The ERC

$$\max_{j \notin S^*} \|A_{S^*}^\dagger \mathbf{a}_j\|_1 < 1 \quad (\text{ERC}(A, S^*))$$

turns out to be necessary and sufficient for uniform (*i.e.*, for any  $\mathbf{x}^*$  supported by  $S^*$ )  $K$ -step recovery using both OMP [12] and OLS [11]. It is therefore weaker than the MIP and RIP conditions which hold for all  $K$ -sparse representations irrespective of their supports. Specifically, true atoms are considered aside from wrong atoms in  $\text{ERC}(A, S^*)$ , which allows one to carry out finer exact recovery analyses. To the best of our knowledge, the fact that the ERC ensures exact support recovery with signed versions of OMP and OLS is an open question. Hereafter, we elaborate on the fact that the sign preservation property is not guaranteed under the ERC.

**Proposition V.1.** *For any  $K \geq 4$ , there exists a dictionary  $A$  and a non-negative  $K$  sparse representation for which  $\text{ERC}(A, S^*)$  holds and the non-negative signs are not preserved with OMP and OLS.*

*Proof:* In the following counterexample, we exploit that the ERC condition remains true when wrong atoms are orthogonal to the span of the true atoms, regardless of the correlation between the true atoms. We further exhibit a non-negative  $K$ -sparse representation  $\mathbf{y}$  for which an atom with a negative weight is selected at the first iteration of Oxx.

Let us consider a dictionary  $A \in \mathbb{R}^{(K+1) \times (K+1)}$  made of  $K$  atoms forming an equiangular subdic-

tionary, and a last one being orthogonal to the others. Specifically,  $A$  is such that

$$(A^t A)_{ij} = \begin{cases} 1 & \text{if } i = j \\ -\frac{1}{K} & \text{if } i \neq j \text{ and } i, j \leq K \\ 0 & \text{if } i \neq j \text{ and } \max(i, j) = K + 1. \end{cases}$$

The true support is set to  $S^* = \{1, \dots, K\}$ . Clearly, the MIP does not hold since  $\mu(A) = \frac{1}{K} > \frac{1}{2K-1}$ . On the contrary,  $\text{ERC}(A, S^*)$  is trivially met since the single wrong atom  $\mathbf{a}_{K+1}$  is orthogonal to the true ones.

Let us define  $\mathbf{y} = A\mathbf{x}^*$  with a true weight vector

$$\mathbf{x}^* = [a, \dots, a, 1, 0]^t, \text{ with } a > \frac{K-1}{K-3}.$$

Obviously,  $K$ -step exact recovery is achieved by Oxx since  $\text{ERC}(A, S^*)$  is met [11], [12]. However, the Oxx iterates and their nonnegative versions differ from the very first iteration. Indeed, we have  $A^t \mathbf{y} = [b, \dots, b, c, 0]^t$  with  $b = \frac{2a-1}{K} > 0$  and  $c = \frac{K-(K-1)a}{K} < 0$ . Since  $c < -b$ ,  $\mathbf{a}_K$  is the first atom to be selected by Oxx algorithms (in violation with the sign preservation property, since  $\mathbf{y}$  is negatively correlated with  $\mathbf{a}_K$ ), whereas NNOMP, NNOLS, and SNNOLS pick any of the first  $(K-1)$  atoms. ■

This example tends to indicate that the sign preservation property of Oxx is only met in restrictive situations, while the  $K$ -step exact recovery property of non-negative versions could hold under broader conditions. The exact recovery analyses that can be foreseen would be substantially different from the one we proposed, since sign-preservation with Oxx is not guaranteed anymore.

2) *RIP based analyses*: In recent years, various analyses of OMP and OLS algorithms have been carried out to derive  $K$ -step recovery guarantees using the restricted isometry property. The OMP algorithm was analyzed first. Improved conditions have been gradually proposed on the RIP factor or order  $K+1$ , denoted by  $\delta_{K+1}$  [17], [20], [21], [24], [44]–[46]. The OLS algorithm was analyzed more recently with increasingly sharp bounds [22]–[24]. For both OMP and OLS, it was shown in [24] that when  $N \geq 4$ , the condition  $\delta_{K+1} < \frac{1}{\sqrt{K}}$  is a sharp bound; on the one hand, it is sufficient for  $K$ -step exact recovery of any  $K$ -sparse representation. On the other hand, there exist dictionaries for which  $\delta_{K+1} = \frac{1}{\sqrt{K}}$  and Oxx fail to recover a good atom in the first iteration.

A natural perspective of our work is to investigate whether exact recovery guarantees could be derived under restricted isometry assumptions. The fact that the sign preservation property is guaranteed when  $\delta_{K+1} < \frac{1}{\sqrt{K}}$  is an open question. It turns out that the mutual coherence analysis of Lemma III.3 cannot be easily adapted since the related proof strongly relies on the derivation of lower bounds of inner products  $\tilde{\mathbf{c}}_j^t \mathbf{r}$ , where  $\mathbf{r}$  and  $\tilde{\mathbf{c}}_j$  respectively stand for the current residual and the projected (possibly normalized)

atoms, in order to show that (58) holds true. In contrast, the classical RIP analyses [24] consist of deriving (i) an accurate lower bound of the *modulus* of the inner products  $|\tilde{\mathbf{c}}_j^t \mathbf{r}|$  of the residual with true atoms  $\mathbf{a}_j$ , and (ii) an upper bound of the modulus of inner products related to wrong atoms, in order to show that  $|\tilde{\mathbf{c}}_j^t \mathbf{r}|$  is maximal when  $\mathbf{a}_j$  is a true atom. Regarding step (i), the most accurate techniques make use of the inequality  $\|\mathbf{t}\|_\infty \geq \|\mathbf{t}\|_2 / \sqrt{\|\mathbf{t}\|_0}$  to lower bound  $\max_{j \in S^* \setminus S_{k-1}} |\tilde{\mathbf{c}}_j^t \mathbf{r}| = \|\tilde{\mathbf{C}}_{S^* \setminus S_{k-1}}^t \mathbf{r}\|_\infty$ , using  $\mathbf{t} = \tilde{\mathbf{C}}_{S^* \setminus S_{k-1}}^t \mathbf{r}$ , see, e.g., [19]. Unfortunately, the latter techniques cannot be used to lower bound the *signed* inner product<sup>5</sup>  $\max_{j \in S^* \setminus S_{k-1}} \tilde{\mathbf{c}}_j^t \mathbf{r}$ . Since the standard RIP analysis techniques are not sign aware, it seems that brandnew analysis techniques need to be elaborated in the non-negative setting.

## VI. APPENDIX

### A. Useful lemmas

Let us recall some useful lemmas. Lemma VI.1 provides an upper bound on the  $\ell_1$ -norm of the Gram matrix columns by means of mutual coherence. Lemma VI.2 provides lower and upper bounds on the inner product of projected atoms. Lemma VI.3 is related to the full rankness of the matrix of projected atoms. Lemma VI.4 is a simple algebraic manipulation related to the pseudo-inverse.

**Lemma VI.1.** [47, Theorem 5.3], [12, Proposition 2.1, Theorem 3.5]. *If  $B$  is a column-normalized matrix with  $k$  columns and  $\mu(B) < \frac{1}{k-1}$  then  $B$  is full column rank and*

$$\|(B^t B)^{-1}\|_{1,1} \leq \frac{1}{1 - (k-1)\mu(B)} \quad (15)$$

wherein  $\|\cdot\|_{1,1}$  equals the maximum absolute column sum of its argument.

**Lemma VI.2.** [16, Lemmas 2 & 3]. *If  $\mu(A) \leq \frac{1}{k+1}$  with  $k = \text{card}(S)$  then*

$$\forall i \notin S, \|\tilde{\mathbf{a}}_i^S\|^2 \geq \beta_k, \quad (16)$$

$$\forall p \neq q, |(\tilde{\mathbf{c}}_p^S)^t \tilde{\mathbf{a}}_q^S| \leq \mu_k \eta_k, \quad (17)$$

<sup>5</sup>The same limitation can be pointed out regarding the generalization of Tropp's ERC analysis to the non-negative setting [12, Theorem 3.1] as argued in Section II-D.

where

$$\beta_k = \frac{(1 + \mu(A))(1 - k\mu(A))}{1 - (k-1)\mu(A)} \quad (18)$$

$$\mu_k = \frac{\mu(A)}{1 - k\mu(A)} \quad (19)$$

$$\eta_k = \begin{cases} \beta_k & \text{for OMP,} \\ \sqrt{\beta_k} & \text{for OLS.} \end{cases} \quad (20)$$

**Lemma VI.3.** [11, Lemma 8] If  $S \cap S' = \emptyset$  and  $A_{S \cup S'}$  is full column rank, then matrices  $\tilde{A}_{S'}^S$  and  $\tilde{B}_{S'}^S$  are full column rank.

**Lemma VI.4.** Consider a full column rank matrix  $A = [\mathbf{a}_1, A_2] \in \mathbb{R}^{p \times q}$  where  $\mathbf{a}_1 \in \mathbb{R}^p$  and  $A_2 \in \mathbb{R}^{p \times (q-1)}$  denotes the submatrix formed of the last  $q-1$  column vectors. Then,

$$\forall \mathbf{r} \in \mathbb{R}^p, (A^\dagger \mathbf{r})_1 = \frac{\langle \mathbf{r}, P_{S_2}^\perp \mathbf{a}_1 \rangle}{\|P_{S_2}^\perp \mathbf{a}_1\|^2}. \quad (21)$$

where the index set  $S_2 = \{2, \dots, q\}$  corresponds to the columns of  $A_2$ .

*Proof:* Let  $\mathbf{r} \in \mathbb{R}^p$ . Since  $A$  is full column rank,  $\mathbf{r}$  can be uniquely decomposed as

$$\mathbf{r} = \mathbf{p}_A + \mathbf{p}_{A^\perp} = (A^\dagger \mathbf{r})_1 \mathbf{a}_1 + \mathbf{r}_{A_2} + \mathbf{p}_{A^\perp} \quad (22)$$

where  $\mathbf{p}_A$  and  $\mathbf{p}_{A^\perp}$  are the orthogonal projections of  $\mathbf{r}$  onto  $\text{span}(A)$  and  $\text{span}(A)^\perp$ , respectively, and  $\mathbf{r}_{A_2} \in \text{span}(A_2)$ . Note though that the decomposition  $\mathbf{p}_A = (A^\dagger \mathbf{r})_1 \mathbf{a}_1 + \mathbf{r}_{A_2}$  is not orthogonal. Rewriting  $\mathbf{a}_1$  as  $P_{S_2} \mathbf{a}_1 + P_{S_2}^\perp \mathbf{a}_1$  yields the orthogonal decomposition:

$$\mathbf{r} = (A^\dagger \mathbf{r})_1 P_{S_2}^\perp \mathbf{a}_1 + (\mathbf{r}_{A_2} + P_{S_2} \mathbf{a}_1) + \mathbf{p}_{A^\perp}. \quad (23)$$

(21) is obtained directly from (23) by writing the inner product  $\langle \mathbf{r}, P_{S_2}^\perp \mathbf{a}_1 \rangle$  because the latter decomposition is orthogonal. The denominator in (21) is nonzero because of the full rankness of  $A$ . ■

### B. Proof of Lemma III.2

Let us now start with the proof of Lemma III.2, since this Lemma will be used later in the proofs of Lemmas III.1, III.3, and III.4.

Since  $\hat{\mathbf{x}}_{S_k}^{(k)}$  is the unconstrained least squares solution related to subset  $S_k$ , we have for  $j < k$ ,

$$\begin{aligned} \hat{\mathbf{x}}_{S_k}^{(k)} &= \arg \min_{\mathbf{z} \in \mathbb{R}^k} \|\mathbf{y} - A_{S_k} \mathbf{z}\|^2 \\ &= \arg \min_{\mathbf{v}, \mathbf{w}} \|\mathbf{y} - A_{S_j} \mathbf{v} - A_{S_k \setminus S_j} \mathbf{w}\|^2. \end{aligned} \quad (24)$$

In addition,

$$\begin{aligned}
 \min_{\mathbf{v}, \mathbf{w}} \|\mathbf{y} - A_{S_j} \mathbf{v} - A_{S_k \setminus S_j} \mathbf{w}\|^2 &= \min_{\mathbf{w}} \left( \min_{\mathbf{v}} \|\mathbf{y} - A_{S_k \setminus S_j} \mathbf{w} - A_{S_j} \mathbf{v}\|^2 \right) \\
 &= \min_{\mathbf{w}} \|P_{S_j}^\perp (\mathbf{y} - A_{S_k \setminus S_j} \mathbf{w})\|^2 \\
 &= \min_{\mathbf{w}} \|\mathbf{r}_j - \tilde{A}_{S_k \setminus S_j}^{S_j} \mathbf{w}\|^2.
 \end{aligned} \tag{25}$$

Since  $A_{S_k}$  is full column rank,  $\tilde{A}_{S_k \setminus S_j}^{S_j}$  is full column rank as well according to Lemma VI.3. The minimum corresponding to (25) is reached for

$$\mathbf{w} = \left[ \tilde{A}_{S_k \setminus S_j}^{S_j} \right]^\dagger \mathbf{r}_j \tag{26}$$

which identifies with  $\hat{\mathbf{x}}_{S_k \setminus S_j}^{(k)}$  according to (24).

### C. Proof of Lemma III.1

We prove by induction that the supports found by OMP and NNOMP (respectively, by OLS and NNOLS/SNNOLS) coincide at each iteration.

1) *NNOMP vs OMP*: Let us first prove the claim at the first iteration. Let us denote by  $\ell$  the index of the atom selected by OMP. The first OMP iterate is the one-sparse vector supported by  $\{\ell\}$  such that  $\hat{\mathbf{x}}_\ell^{(1)} = \mathbf{a}_\ell^t \mathbf{y}$ . By assumption,  $\hat{\mathbf{x}}_\ell^{(1)} > 0$ . Since  $\ell$  is the unique solution to (3) with  $\mathbf{r}_0 = \mathbf{y}$ , we have:

$$\{\ell\} = \arg \max_i \{\mathbf{a}_i^t \mathbf{y}\}, \tag{27}$$

which implies that  $\ell$  is also selected by NNOMP. Since  $\mathbf{a}_\ell^t \mathbf{y} > 0$ ,  $\hat{\mathbf{x}}^{(1)} \geq \mathbf{0}$  corresponds to both the unconstrained and non-negative solutions related to subset  $\{\ell\}$ , that is, to both NNOMP and OMP very first iterates.

Let us assume that OMP and NNOMP deliver the same support  $S_{k-1}$  and solution after  $k-1$  iterations. The residual vectors  $\mathbf{r}_{k-1}$  corresponding to OMP and NNOMP therefore coincide. Let  $S_k = S_{k-1} \cup \{\ell\}$  denote the support found by OMP at iteration  $k$  with  $\ell$  the index of the atom selected at iteration  $k$ , and let  $\hat{\mathbf{x}}^{(k)}$  denote the OMP iterate. Lemma III.2 with  $j \leftarrow k-1$  implies that:

$$\hat{\mathbf{x}}_\ell^{(k)} = \left[ \tilde{\mathbf{a}}_\ell^{S_{k-1}} \right]^\dagger \mathbf{r}_{k-1} = \frac{\mathbf{r}_{k-1}^t \tilde{\mathbf{a}}_\ell^{S_{k-1}}}{\|\tilde{\mathbf{a}}_\ell^{S_{k-1}}\|^2} \tag{28}$$

$$= \frac{\mathbf{r}_{k-1}^t \mathbf{a}_\ell}{\|\tilde{\mathbf{a}}_\ell^{S_{k-1}}\|^2}. \tag{29}$$

$\tilde{\mathbf{a}}_\ell^{S_{k-1}} \neq \mathbf{0}$  in (28) follows from Lemma VI.3 and the full rankness assumption of  $A_{S_k}$ . (29) holds because  $\tilde{\mathbf{a}}_\ell^{S_{k-1}} - \mathbf{a}_\ell$  lays in  $\text{span}(A_{S_{k-1}})$ , and the OMP residual  $\mathbf{r}_{k-1}$  is orthogonal to  $\text{span}(A_{S_{k-1}})$ . By

assumption,  $\hat{\mathbf{x}}^{(k)} \geq \mathbf{0}$ , so (29) implies that  $\mathbf{r}_{k-1}^t \mathbf{a}_\ell \geq 0$ . Since  $\ell$  is the unique solution to (3), we have:

$$\{\ell\} = \arg \max_{i \notin S_{k-1}} \{\mathbf{r}_{k-1}^t \mathbf{a}_i\}. \quad (30)$$

So,  $\ell$  is also selected by NNOMP at iteration  $k$ , leading to the same subset  $S_k$  as for OMP. Because the OMP iterate is non-negative, it is also the NNLS solution corresponding to  $S_k$ . Hence, NNOMP yields the same iterate as OMP.

2) *SNNOLS vs OLS*: We can make a similar argument as in the previous case by replacing the dictionary atoms  $\mathbf{a}_i$  by their normalized projections  $\tilde{\mathbf{b}}_i^S$ .

At the first iteration,  $\tilde{\mathbf{b}}_i^\emptyset = \tilde{\mathbf{a}}_i^\emptyset = \mathbf{a}_i$ , so the very first iterates of SNNOLS and OLS respectively identify with those of NNOMP and OMP. They coincide according to § VI-C1. At iteration  $k$ , the proof of § VI-C1 can be repeated, where  $\hat{\mathbf{x}}^{(k)}$  now denotes the OLS iterate. The right-hand side of (28) rereads  $(\mathbf{r}_{k-1}^t \tilde{\mathbf{b}}_\ell^{S_{k-1}}) / \|\tilde{\mathbf{a}}_\ell^{S_{k-1}}\| \geq 0$ , so  $\mathbf{r}_{k-1}^t \tilde{\mathbf{b}}_\ell^{S_{k-1}} \geq 0$ . Finally, since  $\ell$  is the only maximizer of the OLS selection rule (4), we have

$$\{\ell\} = \arg \max_{i \notin S_{k-1}} \{\mathbf{r}_{k-1}^t \tilde{\mathbf{b}}_i^{S_{k-1}}\}. \quad (31)$$

So,  $\ell$  is also selected by SNNOLS. Similar to the NNOMP vs OMP case, we conclude that NNOLS and SNNOLS yield the same iterate at iteration  $k$ .

3) *NNOLS vs OLS*: The very first iterates of OLS and NNOLS identify to those of OMP and NNOMP, respectively. We have proved above that they coincide.

Assume that OLS and NNOLS deliver the same support  $S_{k-1}$  after  $k-1$  iterations, and the same iterate  $\hat{\mathbf{x}}_{S_{k-1}}^{(k)} = A_{S_{k-1}}^\dagger \mathbf{y} > \mathbf{0}$ . Let  $\ell$  denote the atom selected by OLS at iteration  $k$ , with  $S_k = S_{k-1} \cup \{\ell\}$ . Since OLS selects the atom yielding the least squared error and the optimal index  $\ell$  is unique, we have

$$\forall i \notin S_k, \min_{\mathbf{z}} \|\mathbf{y} - A_{S_k} \mathbf{z}\|^2 < \min_{\mathbf{z}} \|\mathbf{y} - A_{S_{k-1} \cup \{i\}} \mathbf{z}\|^2. \quad (32)$$

Clearly,

$$\min_{\mathbf{z}} \|\mathbf{y} - A_{S_{k-1} \cup \{i\}} \mathbf{z}\|^2 \leq \min_{\mathbf{z} \geq \mathbf{0}} \|\mathbf{y} - A_{S_{k-1} \cup \{i\}} \mathbf{z}\|^2. \quad (33)$$

Moreover, since the OLS iterate at iteration  $k$  is non-negative, it is also an NNLS solution related to  $S_k$ . Therefore,

$$\min_{\mathbf{z}} \|\mathbf{y} - A_{S_k} \mathbf{z}\|^2 = \min_{\mathbf{z} \geq \mathbf{0}} \|\mathbf{y} - A_{S_k} \mathbf{z}\|^2. \quad (34)$$

From (32)-(34), we get

$$\forall i \notin S_k, \min_{\mathbf{z} \geq \mathbf{0}} \|\mathbf{y} - A_{S_k} \mathbf{z}\|^2 < \min_{\mathbf{z} \geq \mathbf{0}} \|\mathbf{y} - A_{S_{k-1} \cup \{i\}} \mathbf{z}\|^2,$$

which implies that  $\ell$  is also selected by NNOLS. By assumption,  $A_{S_k}$  is full column rank, so the unconstrained and non-negative least-squares solutions related to  $S_k$  are unique. (34) implies that they coincide. Hence, OLS and NNOLS yield the same iterate at iteration  $k$ .

#### D. Proof of Lemma III.3

Here, the abridged notations  $\tilde{\mathbf{y}}, \tilde{\mathbf{a}}_i, \tilde{\mathbf{b}}_i, \tilde{\mathbf{c}}_i, \tilde{\mathbf{n}}$  correspond to projected vectors onto  $\text{span}(A_{S_{k-1}})^\perp$ . We denote by  $\ell$  the atom selected at iteration  $k$ , so that  $S_k \setminus S_{k-1} = \{\ell\}$ . In a somewhat counter-intuitive manner, our proof of Lemma III.3 is not recursive: to prove that  $\hat{x}_\ell^{(k)} > 0$ , we exploit that good atoms have been selected at previous iterations, but we do not consider the current weight signs.

First, it should be noticed that  $\mu(A) < \frac{1}{2K-1}$  implies that  $A_{S^*}$  is full column rank according to Lemma VI.1. Applying Lemma III.2 for  $j \leftarrow k-1$ , we have

$$\hat{x}_\ell^{(k)} = \tilde{\mathbf{a}}_\ell^t \mathbf{r}_{k-1}. \quad (35)$$

Therefore,  $\hat{x}_\ell^{(k)}$  has the same sign as  $\tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1}$  since  $\tilde{\mathbf{c}}_\ell = \tilde{\mathbf{a}}_\ell$  or  $\frac{\tilde{\mathbf{a}}_\ell}{\|\tilde{\mathbf{a}}_\ell\|}$  ( $\tilde{\mathbf{a}}_\ell \neq \mathbf{0}$  according to Lemma VI.3). The remaining part of the proof consists in showing that  $\tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1} > 0$ .

Since atom  $\ell$  is selected at the  $k$ -th iteration, we have

$$\forall j \notin S_k, |\tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1}| \geq |\tilde{\mathbf{c}}_j^t \mathbf{r}_{k-1}|. \quad (36)$$

To prove that  $\tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1} > 0$ , we are going to exhibit an index  $j \notin S_k$  such that

$$\tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1} > -|\tilde{\mathbf{c}}_j^t \mathbf{r}_{k-1}|. \quad (37)$$

Indeed, (37) can be consistent with (36), *i.e.*,  $-|\tilde{\mathbf{c}}_j^t \mathbf{r}_{k-1}| \geq -|\tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1}|$  only if  $\tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1} > 0$ . It is therefore sufficient to show that (37) holds for some  $j$  to complete the proof. Specifically, we show (37) for any

$$j \in \arg \max_{i \in \{1, \dots, n\} \setminus S_k} x_i^*. \quad (38)$$

When  $k = K$ ,  $S_k$  identifies with  $S^*$ , thus  $j$  in (38) refers to any wrong atom, with  $x_j^* = 0$ . On the contrary, when  $k < K$ , we have  $j \in S^* \setminus S_k$ , *i.e.*,  $x_j^* > 0$ .

Let us introduce the decomposition

$$\mathbf{r}_{k-1} = P_{S_{k-1}}^\perp \mathbf{y} = \sum_{i \notin S_{k-1}} x_i^* \tilde{\mathbf{a}}_i + \tilde{\mathbf{n}}. \quad (39)$$

For  $p \notin S_{k-1}$ , we have

$$\tilde{\mathbf{c}}_p^t \mathbf{r}_{k-1} = x_p^* \tilde{\mathbf{a}}_p^t \tilde{\mathbf{c}}_p + \sum_{i \notin S_{k-1} \cup \{p\}} x_i^* \tilde{\mathbf{a}}_i^t \tilde{\mathbf{c}}_p + \tilde{\mathbf{n}}^t \tilde{\mathbf{c}}_p \quad (40)$$

where

$$\tilde{\mathbf{a}}_p^t \tilde{\mathbf{c}}_p = \begin{cases} \|\tilde{\mathbf{a}}_p\|^2 & \text{for OMP,} \\ \|\tilde{\mathbf{a}}_p\| & \text{for OLS.} \end{cases} \quad (41)$$

Since  $K \geq k \geq 1$ ,  $\mu(A) < \frac{1}{2K-1} \leq \frac{1}{k}$ . Then, Lemma VI.2 yields:

$$\tilde{\mathbf{a}}_p^t \tilde{\mathbf{c}}_p \geq \eta_{k-1}, \quad (42)$$

$$|\tilde{\mathbf{a}}_i^t \tilde{\mathbf{c}}_p| \leq \mu_{k-1} \eta_{k-1} \quad (43)$$

with  $\mu_{k-1}$  and  $\eta_{k-1}$  defined in (19)-(20). By the Cauchy-Schwarz inequality, we have

$$|\tilde{\mathbf{n}}^t \tilde{\mathbf{c}}_p| \leq \|\tilde{\mathbf{n}}\| \|\tilde{\mathbf{c}}_p\| \leq \|\tilde{\mathbf{n}}\| \leq \|\mathbf{n}\| \quad (44)$$

since  $\|\tilde{\mathbf{c}}_p\| \leq 1$ . (40) yields

$$\tilde{\mathbf{c}}_p^t \mathbf{r}_{k-1} \geq \eta_{k-1} \left( x_p^* - \mu_{k-1} \sum_{i \notin S_{k-1} \cup \{p\}} x_i^* \right) - \|\mathbf{n}\|. \quad (45)$$

Notice that there are at most  $K - k + 1$  nonzero terms in the sum in (45) since  $\mathbf{x}^*$  is  $K$ -sparse, and exactly  $K - k$  nonzeros in the special case  $p = \ell$ . The latter sum can then be upper bounded as:

$$\sum_{i \notin S_{k-1} \cup \{p\}} x_i^* \leq \begin{cases} x_\ell^* + (K - k)x_j^* & \text{if } p \neq \ell, \\ (K - k)x_j^* & \text{if } p = \ell \end{cases} \quad (46)$$

according to the definition of  $j$  in (38).

Apply (45) twice with  $p \leftarrow j$  and  $p \leftarrow \ell$ . Since  $j \neq \ell$ , we get

$$\tilde{\mathbf{c}}_j^t \mathbf{r}_{k-1} \geq \eta_{k-1} \left( (1 - (K - k)\mu_{k-1})x_j^* - \mu_{k-1}x_\ell^* \right) - \|\mathbf{n}\|,$$

$$\tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1} \geq \eta_{k-1} \left( x_\ell^* - (K - k)\mu_{k-1}x_j^* \right) - \|\mathbf{n}\|,$$

from which we deduce

$$\tilde{\mathbf{c}}_j^t \mathbf{r}_{k-1} + \tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1} \geq \eta_{k-1} (\tau_1 x_\ell^* + \tau_2 x_j^*) - 2\|\mathbf{n}\|, \quad (47)$$

with

$$\tau_1 := 1 - \mu_{k-1}, \quad (48)$$

$$\tau_2 := 1 - (2K - 2k)\mu_{k-1}. \quad (49)$$

One can easily check from (19) that  $\tau_1 > 0$  whenever  $\mu(A) < \frac{1}{k}$ , and  $\tau_2 > 0$  when  $(2K - k - 1)\mu(A) < 1$ . Since  $\mu(A) < \frac{1}{2K-1}$ , it holds true that  $\tau_1 > 0$  and  $\tau_2 > 0$ . Therefore, (47) yields

$$\begin{aligned} & \tilde{\mathbf{c}}_j^t \mathbf{r}_{k-1} + \tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1} \\ & \geq \eta_{k-1} \tau_1 x_\ell^* - 2\|\mathbf{n}\| \\ & > \left[ \eta_{k-1} \tau_1 - (1 - (2K - 1)\mu(A)) \right] \min_{i \in S^*} x_i^* \end{aligned} \quad (50)$$

where we exploit that

$$2\|\mathbf{n}\| < (1 - (2K - 1)\mu(A)) \min_{i \in S^*} x_i^* \quad (51)$$

since  $\mathbf{n}$  satisfies (6). According to (16),  $\beta_{k-1} \leq 1$  which implies  $\beta_{k-1} \leq \sqrt{\beta_{k-1}}$  and then

$$\eta_{k-1} \geq \beta_{k-1} \quad (52)$$

using (20). It follows from (50) that

$$\begin{aligned} & \tilde{\mathbf{c}}_j^t \mathbf{r}_{k-1} + \tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1} \\ & > \left[ \beta_{k-1} \tau_1 - (1 - (2K - 1)\mu(A)) \right] \min_{i \in S^*} x_i^*. \end{aligned} \quad (53)$$

According to (48) and (18)-(19),

$$\begin{aligned} \beta_{k-1} \tau_1 &= \beta_{k-1} (1 - \mu_{k-1}) \\ &= \frac{(1 + \mu(A))}{1 - (k - 2)\mu(A)} (1 - k\mu(A)) \end{aligned} \quad (54)$$

$$\geq \begin{cases} 1 - \mu(A) & \text{if } k = 1, \\ (1 + \mu(A)) (1 - k\mu(A)) & \text{if } k \geq 2. \end{cases} \quad (55)$$

When  $k \geq 2$ , (55) rereads

$$\beta_{k-1} \tau_1 \geq 1 - (k - 1)\mu(A) - k\mu(A)^2. \quad (56)$$

Since  $\mu(A) \leq 1$ , one has  $\mu(A)^2 \leq \mu(A)$ , thus

$$\beta_{k-1} \tau_1 \geq 1 - (2k - 1)\mu(A) \geq 1 - (2K - 1)\mu(A). \quad (57)$$

When  $k = 1$ , (57) also holds from (55). It follows from (53) that for any  $k \in \{1, \dots, K\}$ ,

$$\tilde{\mathbf{c}}_j^t \mathbf{r}_{k-1} + \tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1} > 0, \quad (58)$$

which implies (37). We conclude from (36) that  $\tilde{\mathbf{c}}_\ell^t \mathbf{r}_{k-1}$  and then  $\hat{x}_\ell^{(k)}$  are positive quantities.

### E. Proof of Lemma III.4

In this proof, the abridged notations  $\tilde{\mathbf{a}}_i, \tilde{\mathbf{b}}_i, \tilde{\mathbf{c}}_i$  correspond to projected vectors onto  $\text{span}(A_{S_{j-1}})^\perp$  (the latter space being different from that of Subsection VI-D). Similarly, we will use the abridged notations  $\tilde{A}, \tilde{B}$  to refer to  $\tilde{A}^{S_{j-1}}, \tilde{B}^{S_{j-1}}$  and  $\tilde{C}^{S_{j-1}}$ , respectively. For any subset  $S$ , the Gram matrices of  $\tilde{A}_S$  and  $\tilde{B}_S$  will be respectively denoted by  $G_S^{\text{OMP}} := \tilde{A}_S^t \tilde{A}_S$  and  $G_S^{\text{OLS}} := \tilde{B}_S^t \tilde{B}_S$ . For statements that are common to both OMP and OLS, we will use the simpler generic notation  $G_S := \tilde{C}_S^t \tilde{C}_S$ .

The proof of Lemma III.4 is not recursive. However, in order to prove that  $\hat{\mathbf{x}}_{S_{k-1}}^{(k)} > \mathbf{0}$ , we use the fact that  $\hat{\mathbf{x}}_{S_j \setminus S_{j-1}}^{(j)} > 0$  for  $j < k$ , which holds according to Lemma III.3.

Let  $\ell \in S_{k-1}$  and denote by  $j < k$  the iteration at which the atom indexed by  $\ell$  has been selected by Oxx, so that  $S_j \setminus S_{j-1} = \{\ell\}$ . According to Lemma III.2 and since  $\ell$  is the first entry in the ordered set  $S_k \setminus S_{j-1}$ , we have

$$\hat{\mathbf{x}}_\ell^{(k)} = (\tilde{A}_{S_k \setminus S_{j-1}}^\dagger \mathbf{r}_{j-1})_1. \quad (59)$$

In order to exploit the Oxx selection rule which is based on the atoms  $\tilde{\mathbf{c}}_i$  defined in (1), let us rewrite (59) with respect to matrix  $\tilde{C}$ . Obviously, one can rewrite  $\tilde{A}_{S_k \setminus S_{j-1}} = \tilde{C}_{S_k \setminus S_{j-1}} \Delta$ , where  $\Delta > \mathbf{0}$  is a square diagonal matrix whose diagonal elements are either equal to 1 (OMP,  $\tilde{C} \leftarrow \tilde{A}$ ) or to  $\|\tilde{\mathbf{a}}_i\| > 0$ ,  $i \in S_k \setminus S_{j-1}$  (OLS,  $\tilde{C} \leftarrow \tilde{B}$ ). The positivity property is related to the full rankness of  $\tilde{C}_{S_k \setminus S_{j-1}}$ , which deduces from that of  $A_{S_k}$  according to Lemma VI.3. From standard properties of pseudo-inverses, we have  $\tilde{A}_{S_k \setminus S_{j-1}}^\dagger = \Delta^{-1} \tilde{C}_{S_k \setminus S_{j-1}}^\dagger$ . (59) yields

$$\begin{aligned} \hat{\mathbf{x}}_\ell^{(k)} &\overset{\dagger}{\propto} (\tilde{C}_{S_k \setminus S_{j-1}}^\dagger \mathbf{r}_{j-1})_1 = \left( [\tilde{\mathbf{c}}_\ell, \tilde{C}_{S_k \setminus S_j}]^\dagger \mathbf{r}_{j-1} \right)_1 \\ &= \frac{\langle \mathbf{r}_{j-1}, \tilde{P}_{S_k \setminus S_j}^\perp \tilde{\mathbf{c}}_\ell \rangle}{\|\tilde{P}_{S_k \setminus S_j}^\perp \tilde{\mathbf{c}}_\ell\|^2} \end{aligned} \quad (60)$$

where  $\overset{\dagger}{\propto}$  indicates proportionality up to a positive factor,  $\tilde{P}_{S_k \setminus S_j}^\perp$  denotes the orthogonal projection onto the orthogonal complement of  $\text{span}(\tilde{C}_{S_k \setminus S_j})$ , and (60) deduces from Lemma VI.4.

(60) implies that:

$$\hat{\mathbf{x}}_\ell^{(k)} > 0 \iff \langle \tilde{P}_{S_k \setminus S_j}^\perp \mathbf{r}_{j-1}, \tilde{\mathbf{c}}_\ell \rangle > 0 \quad (61)$$

$$\iff \langle \tilde{C}_{S_k \setminus S_j} \tilde{C}_{S_k \setminus S_j}^\dagger \mathbf{r}_{j-1}, \tilde{\mathbf{c}}_\ell \rangle < \langle \mathbf{r}_{j-1}, \tilde{\mathbf{c}}_\ell \rangle \quad (62)$$

$$\iff \langle \tilde{C}_{S_k \setminus S_j}^\dagger \mathbf{r}_{j-1}, \tilde{C}_{S_k \setminus S_j}^t \tilde{\mathbf{c}}_\ell \rangle < \langle \mathbf{r}_{j-1}, \tilde{\mathbf{c}}_\ell \rangle. \quad (63)$$

By Hölder's inequality and since  $\tilde{C}_{S_k \setminus S_j}^\dagger = G_{S_k \setminus S_j}^{-1} \tilde{C}_{S_k \setminus S_j}^t$ , the left-hand side (LHS) of (63) is upper

bounded by

$$\begin{aligned}
 \|\tilde{C}_{S_k \setminus S_j}^t \tilde{\mathbf{c}}_\ell\|_\infty \|\tilde{C}_{S_k \setminus S_j}^\dagger \mathbf{r}_{j-1}\|_1 &\leq \|\tilde{C}_{S_k \setminus S_j}^t \tilde{\mathbf{c}}_\ell\|_\infty \|G_{S_k \setminus S_j}^{-1}\|_{1,1} \|\tilde{C}_{S_k \setminus S_j}^t \mathbf{r}_{j-1}\|_1 \\
 &\leq \mu(\tilde{C}) \|G_{S_k \setminus S_j}^{-1}\|_{1,1} \|\tilde{C}_{S_k \setminus S_j}^t \mathbf{r}_{j-1}\|_1 \\
 &\leq \mu(\tilde{C}) \|G_{S_k \setminus S_j}^{-1}\|_{1,1} (k-j) |\langle \mathbf{r}_{j-1}, \tilde{\mathbf{c}}_\ell \rangle|.
 \end{aligned} \tag{64}$$

To obtain the last inequality, we exploit that  $\mathbf{a}_\ell$  has been selected at the  $j$ -th iteration of Oxx, therefore:

$$\forall i \in S_k \setminus S_j, |\langle \mathbf{r}_{j-1}, \tilde{\mathbf{c}}_i \rangle| \leq |\langle \mathbf{r}_{j-1}, \tilde{\mathbf{c}}_\ell \rangle|. \tag{65}$$

From Lemma III.3, we have  $\hat{x}_\ell^{(j)} > 0$ , and as already remarked in the proof of Lemma III.3,  $\tilde{\mathbf{c}}_\ell^t \mathbf{r}_{j-1}$  is of same sign as  $\hat{x}_\ell^{(j)}$ , see (35). The upper bound in (64) thus rewrites

$$(k-j) \mu(\tilde{C}) \|G_{S_k \setminus S_j}^{-1}\|_{1,1} \langle \mathbf{r}_{j-1}, \tilde{\mathbf{c}}_\ell \rangle \tag{66}$$

and we deduce from (63) that

$$((k-j) \mu(\tilde{C}) \|G_{S_k \setminus S_j}^{-1}\|_{1,1} < 1) \implies (\hat{x}_\ell^{(k)} > 0). \tag{67}$$

Let us now provide some upper bounds on  $\mu(\tilde{C})$  and  $\|G_{S_k \setminus S_j}^{-1}\|_{1,1}$  in order to show that the LHS of (67) holds true.

1) *Upper bound on  $\mu(\tilde{C})$* : Since  $\mu(A) < \frac{1}{2K-1} < \frac{1}{j}$ , Lemma VI.2 yields:

$$\forall i \notin S_{j-1}, \|\tilde{\mathbf{a}}_i\|^2 \geq \beta_{j-1}, \tag{68}$$

$$\forall p \neq q, |\tilde{\mathbf{a}}_p^t \tilde{\mathbf{a}}_q| \leq \mu_{j-1} \beta_{j-1}, \tag{69}$$

with  $\mu_{j-1}$  and  $\beta_{j-1}$  defined in (18)-(19), from which we can easily deduce that

$$\mu(\tilde{C}) = \begin{cases} \mu(\tilde{A}) \leq \mu_{j-1} \beta_{j-1} & \text{(OMP case),} \\ \mu(\tilde{B}) \leq \mu_{j-1} & \text{(OLS case).} \end{cases} \tag{70}$$

2) *Upper bound on  $\|G_{S_k \setminus S_j}^{-1}\|_{1,1}$  in the OLS case*: We have shown that for  $i \in S_k \setminus S_j$ ,  $\tilde{\mathbf{a}}_i \neq \mathbf{0}$ , so matrix  $\tilde{B}_{S_k \setminus S_j}$  is column-normalized, hence  $G_{S_k \setminus S_j}^{\text{OLS}}$  has a unit diagonal. Since  $\mu(A) < \frac{1}{2K-1}$ , (70) and (19) imply that:

$$\mu(\tilde{B}) \leq \mu_{j-1} \leq \frac{1}{2K-j} \leq \frac{1}{k-j}. \tag{71}$$

Lemma VI.1 then applies to matrix  $\tilde{B}_{S_k \setminus S_j} \in \mathbb{R}^{m \times (k-j)}$ , so the latter is full column rank, and

$$\|[G_{S_k \setminus S_j}^{\text{OLS}}]^{-1}\|_{1,1} \leq \frac{1}{1 - (k-j-1)\mu(\tilde{B}_{S_k \setminus S_j})} \leq \frac{1}{1 - (k-j-1)\mu(\tilde{B})}. \tag{72}$$

It follows from (70) that

$$\begin{aligned} \mu(\tilde{B}) \| [G_{S_k \setminus S_j}^{\text{OLS}}]^{-1} \|_{1,1} &\leq \mu_{j-1} \| [G_{S_k \setminus S_j}^{\text{OLS}}]^{-1} \|_{1,1} \\ &\leq \frac{\mu_{j-1}}{1 - (k - j - 1)\mu(\tilde{B})}, \\ &\leq \frac{1}{2K - k + 1} \end{aligned} \quad (73)$$

$$< \frac{1}{k - j} \quad (74)$$

where (73) follows from the second upper bound in (71). (74) thus implies that the LHS of (67) is true in the OLS case.

3) *Upper bound on  $\|G_{S_k \setminus S_j}^{-1}\|_{1,1}$  in the OMP case:* Contrary to the OLS case, the diagonal elements of  $G_{S_k \setminus S_j}^{\text{OMP}}$  ( $\|\tilde{\mathbf{a}}_i\|^2$ ,  $i \in S_k \setminus S_j$ ) are not equal to 1, so Lemma VI.1 does not apply. Let  $\Delta$  be the square diagonal matrix with the elements  $\|\tilde{\mathbf{a}}_i\|$ ,  $i \in S_k \setminus S_j$  on its diagonal. Clearly,  $\tilde{A}_{S_k \setminus S_j} = \tilde{B}_{S_k \setminus S_j} \Delta$ , hence

$$[G_{S_k \setminus S_j}^{\text{OMP}}]^{-1} = \Delta^{-1} [G_{S_k \setminus S_j}^{\text{OLS}}]^{-1} \Delta^{-1}, \quad (75)$$

and thus

$$\| [G_{S_k \setminus S_j}^{\text{OMP}}]^{-1} \|_{1,1} \leq \| [G_{S_k \setminus S_j}^{\text{OLS}}]^{-1} \|_{1,1} \| \Delta^{-1} \|_{1,1}^2. \quad (76)$$

Moreover,

$$\| \Delta^{-1} \|_{1,1}^2 = \frac{1}{\min_{i \in S_k \setminus S_j} \|\tilde{\mathbf{a}}_i\|^2} \leq \frac{1}{\beta_{j-1}} \quad (77)$$

by (68). We have thus

$$\mu(\tilde{A}) \| [G_{S_k \setminus S_j}^{\text{OMP}}]^{-1} \|_{1,1} \leq \frac{\mu(\tilde{A})}{\beta_{j-1}} \| [G_{S_k \setminus S_j}^{\text{OLS}}]^{-1} \|_{1,1} \leq \mu_{j-1} \| [G_{S_k \setminus S_j}^{\text{OLS}}]^{-1} \|_{1,1}$$

according to (70). From (74), we conclude that the LHS of (67) is also true in the OMP case.

## REFERENCES

- [1] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition”, in *Proc. 27th Asilomar Conf. on Signals, Systems and Computers*, Nov. 1993, vol. 1, pp. 40–44.
- [2] S. Chen, S. A. Billings, and W. Luo, “Orthogonal least squares methods and their application to non-linear system identification”, *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.
- [3] S. F. Cotter, J. Adler, B. D. Rao, and K. Kreutz-Delgado, “Forward sequential algorithms for best basis selection”, *IEE Proc. Vision, Image and Signal Processing*, vol. 146, no. 5, pp. 235–244, Oct. 1999.
- [4] L. Rebollo-Neira and D. Lowe, “Optimized orthogonal matching pursuit approach”, *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 137–140, Apr. 2002.

- [5] S. Foucart, “Stability and robustness of weak orthogonal matching pursuits”, in *Recent advances in harmonic analysis and applications*, D. Bilyk, L. De Carli, A. Petukhov, A. M. Stokolos, and B. D. Wick, Eds. 2013, vol. 25, pp. 395–405, Springer Proceedings in Mathematics & Statistics.
- [6] B. K. Natarajan, “Sparse approximate solutions to linear systems”, *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.
- [7] S. Bourguignon, J. Ninin, H. Carfantan, and M. Mongeau, “Exact sparse approximation problems via mixed-integer programming: Formulations and computational performance”, *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1405–1419, Mar. 2016.
- [8] R. Tibshirani, “Regression shrinkage and selection via the Lasso”, *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.
- [9] D. L. Donoho and Y. Tsaig, “Fast solution of  $\ell_1$ -norm minimization problems when the solution may be sparse”, *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.
- [10] T. Blumensath and M. E. Davies, “On the difference between Orthogonal Matching Pursuit and Orthogonal Least Squares”, Tech. Rep., University of Edinburgh, Mar. 2007.
- [11] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, “Joint  $k$ -step analysis of orthogonal matching pursuit and orthogonal least squares”, *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3158–3174, May 2013.
- [12] J. A. Tropp, “Greed is good: Algorithmic results for sparse approximation”, *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [13] C. Herzet, C. Soussen, J. Idier, and R. Gribonval, “Exact recovery conditions for sparse representations with partial support information”, *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7509–7524, Nov. 2013.
- [14] Z. Ben-Haim, Y. C. Eldar, and M. Elad, “Coherence-based performance guarantees for estimating a sparse vector under random noise”, *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.
- [15] T. T. Cai and L. Wang, “Orthogonal matching pursuit for sparse signal recovery with noise”, *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4680–4688, July 2011.
- [16] C. Herzet, A. Drémeau, and C. Soussen, “Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay”, *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 459–470, Jan. 2016.
- [17] M. A. Davenport and M. B. Wakin, “Analysis of orthogonal matching pursuit using the restricted isometry property”, *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sept. 2010.
- [18] S. Satpathi, R. Lochan Das, and M. Chakraborty, “Improving the bound on the RIP constant in generalized orthogonal matching pursuit”, *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1074–1077, Nov. 2013.
- [19] B. Li, Y. Shen, Z. Wu, and J. Li, “Sufficient conditions for generalized orthogonal matching pursuit in noisy case”, *Signal Process.*, vol. 108, pp. 111–123, Mar. 2015.
- [20] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, “A sharp condition for exact support recovery with orthogonal matching pursuit”, *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1370–1382, Mar. 2017.
- [21] C. Liu, Y. Fang, and J. Liu, “Some new results about sufficient conditions for exact support recovery of sparse signals via orthogonal matching pursuit”, *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4511–4524, Sept. 2017.
- [22] J. Wen, J. Wang, and Q. Zhang, “Nearly optimal bounds for orthogonal least squares”, *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5347–5356, Oct. 2017.
- [23] J. Wang and P. Li, “Recovery of sparse signals using multiple orthogonal least squares”, *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2049–2062, Apr. 2017.

- [24] J. Kim, J. Wang, and B. Shim, “An optimal restricted isometry condition for exact sparse recovery with orthogonal least squares”, Tech. Rep., Seoul National University, Korea, Oct. 2019.
- [25] 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.
- [26] 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.
- [27] D. Kim and J. P. Haldar, “Greedy algorithms for nonnegativity-constrained simultaneous sparse recovery”, *Signal Process.*, vol. 125, pp. 274–289, 2016.
- [28] T. T. Nguyen, J. Idier, C. Soussen, and E.-H. Djermoune, “Non-negative orthogonal greedy algorithms”, *IEEE Trans. Signal Process.*, vol. 67, no. 21, pp. 5643–5658, Nov. 2019.
- [29] D. L. Donoho, M. Elad, and V. N. Temlyakov, “Stable recovery of sparse overcomplete representations in the presence of noise”, *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
- [30] M. Yaghoobi and M. E. Davies, “Fast non-negative orthogonal least squares”, in *Proc. Eur. Sig. Proc. Conf.*, Nice, France, Aug. 2015, pp. 479–483.
- [31] D. L. Donoho and M. Elad, “Optimally sparse representation in general (non-orthogonal) dictionaries via  $\ell^1$  minimization”, *Proc. Natl. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.
- [32] 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.
- [33] H. Kim and H. Park, “Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis”, *Bioinformatics*, vol. 23, no. 12, pp. 1495–1502, June 2007.
- [34] A. Cherni, E. Chouzenoux, and M.-A. Delsuc, “Fast dictionary-based approach for mass spectrometry data analysis”, in *Proc. IEEE ICASSP*, Calgary, Canada, Apr. 2018, pp. 816–820.
- [35] I. Barbu and C. Herzet, “A new approach for volume reconstruction in tomoPIV with the alternating direction method of multipliers”, *Measurement Science and Technology*, vol. 27, no. 10, pp. 104002, Oct. 2016.
- [36] J.-J. Fuchs, “On sparse representations in arbitrary redundant bases”, *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.
- [37] M. D. Plumbley, “On polar polytopes and the recovery of sparse representations”, *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3188–3195, Sept. 2007.
- [38] A. J. Miller, *Subset selection in regression*, Chapman and Hall, London, UK, 2nd edition, Apr. 2002.
- [39] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression”, *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, Apr. 2004.
- [40] T. T. Cai, L. Wang, and G. Xu, “Stable recovery of sparse signals using an oracle inequality”, *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3516–3522, July 2010.
- [41] M. Mørup, K. H. Madsen, and L. K. Hansen, “Approximate  $L_0$  constrained non-negative matrix and tensor factorization”, in *IEEE Int. Symp. Circuits Syst.*, Seattle, WA, May 2008, pp. 1328–1331.
- [42] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples”, *Appl. Comp. Harmonic Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.
- [43] W. Dai and O. Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction”, *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

- [44] Q. Mo and Y. Shen, “A remark on the restricted isometry property in orthogonal matching pursuit”, *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3654–3656, June 2012.
- [45] J. Wang and B. Shim, “On the recovery limit of sparse signals using orthogonal matching pursuit”, *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4973–4976, Sept. 2012.
- [46] L.-H. Chang and J.-Y. Wu, “An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit”, *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5702–5715, Sept. 2014.
- [47] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Applied and Numerical Harmonic Analysis. Birkhäuser, Basel, June 2013.