Sign preservation analysis of orthogonal greedy algorithms

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

Sign preservation analysis of orthogonal greedy algorithms

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

Abstract

We bring a contribution to the exact recovery theory of a K-sparse vector in the noiseless setting under the standard condition $\mu < 1/(2K-1)$, where μ 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 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; Exact Recovery Condition; Mutual Coherence; Sparse Non-Negative Least Square Problems.

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 2016-2018.

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: {thi-thanh.nguyen,el-hadi.djermoune}@univ-lorraine.fr).

C. Soussen was with CRAN. He is now with CentraleSupélec at the Laboratoire des Signaux et Systèmes (CentraleSupélec-CNRS-Université Paris-Sud, Université Paris-Saclay), 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[1]. 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[2] versions of the candidate atoms [10], [11].

Orthogonal greedy algorithms are able to achieve exact support recovery under certain conditions. The Mutual Incoherence Property (MIP) condition 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]. In the last years, 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) condition [17]–[20]. 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 [21], [22] 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 $K = \mathrm{card}\,(S^*)$ iterations. Our main contribution is to unveil an exact *sign* recovery property regarding the weights of the selected atoms in the best current approximation under the MIP condition and in the noiseless case: at any iteration k between 1 and K, the k coefficients found by Oxx have the same sign as the k corresponding ones in the true decomposition of the data signal.

[1]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].

[2]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 version of OMP and other non-negative greedy schemes. Non-negative OMP (NNOMP) was first introduced by Bruckstein *et al.* [23] under the name OMP, and then renamed NNOMP in [24] (see also [25], [26]). 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 a sign constrained version of OMP, existing exact support recovery analyses are rare, and somewhat discordant. On the one hand, Bruckstein *et al.* conjectured that a MIP type property holds for NNOMP [23], and that the proof is similar to the one given in [12], [27] for OMP. Specifically, [23, Theorem 3] states 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.* elaborate a unified MIP analysis of NNOMP and its generalized version in the multiple measurement vector setting [25, 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, μ is required to be lower than $\frac{1}{K-1} - \frac{K}{2}$, which can occur only when $K = 1$ or 2.

We have been unable to prove [23, Theorem 3] as prescribed by the authors (as a direct extension of the derivations in [12], [27]). 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). Fortunately, the sign preservation property of OMP implies that OMP and NNOMP coincide in the exact support recovery regime, so [23, Theorem 3] becomes a trivial byproduct of our sign preservation analysis. Under the same conditions, OLS and Non-negative OLS (NNOLS) [28] coincide, which allows us to extend [23, Theorem 3] to NNOLS; NNOLS and Suboptimal NNOLS (SNNOLS) [28] also coincide, so the same applies to SNNOLS.

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 the studied algorithms outside the exact support recovery regime.

II. NOTATIONS AND BACKGROUND

A. Notations

Let us denote by $\boldsymbol{y} \in \mathbb{R}^m$ the data signal and by $A \in \mathbb{R}^{m \times n}$ the dictionary of elementary atoms $\boldsymbol{a}_i \in \mathbb{R}^m$, $1 \leq i \leq n$. We are interested in the representation $\boldsymbol{y} = A\boldsymbol{x}$ in which the vector solution $\boldsymbol{x} \in \mathbb{R}^n$ has K non-zero elements. We call this a K-sparse representation. Without loss of generality, we assume that the elementary atoms are normalized, that is $\|\boldsymbol{a}_i\|_2 = 1$. Hereafter, the ℓ_2 norm will also be denoted by $\|\cdot\|$. Notations $^{\mathrm{t}}$ and † stand for the transpose and the Moore-Penrose pseudo-inverse, respectively. For any set of indices $S \subset \{1, \ldots, n\}$, we denote by A_S and $\boldsymbol{x}_S$ the subdictionary and subvector indexed by S, respectively. We denote by $\mathrm{span}(A_S)$ the subspace spanned by the dictionary atoms indexed by S, by $\mathrm{span}(A_S)^{\perp}$ the orthogonal complement of $\mathrm{span}(A_S)$, and by $P_S^{\perp} = I_m - A_S A_S^{\dagger}$ the orthogonal projection onto $\mathrm{span}(A_S)^{\perp}$, 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 $A_S^{\dagger} = (A_S^{\mathrm{t}} A_S)^{-1} A_S^{\mathrm{t}}$. Then we denote by $\tilde{\boldsymbol{a}}_i^S = P_S^{\perp} \boldsymbol{a}_i$ the projected atoms. Clearly, $\tilde{\boldsymbol{a}}_i^S = \boldsymbol{0}$ whenever $i \in S$. Let $\tilde{\boldsymbol{b}}_i^S = \tilde{\boldsymbol{a}}_i^S / \|\tilde{\boldsymbol{a}}_i^S\|$, with the convention that $\tilde{\boldsymbol{b}}_i^S = \boldsymbol{0}$ when $\tilde{\boldsymbol{a}}_i^S = \boldsymbol{0}$. Later, we also use the generic notation

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

to refer to either $\tilde{\boldsymbol{a}}_i^S$ or $\tilde{\boldsymbol{b}}_i^S$ 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 $\tilde{\boldsymbol{a}}_i$ (resp., $\tilde{\boldsymbol{b}}_i$ and $\tilde{\boldsymbol{c}}_i$).

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

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

B. OMP and OLS

OMP and OLS address the following minimization problem:

$$\min_{\|\boldsymbol{x}\|_0 \leq K} \|\boldsymbol{y} - A\boldsymbol{x}\|^2. \tag{2}$$

Both algorithms start with a zero vector solution corresponding to the empty support. Then, at each iteration, a new atom ℓ is selected and added to the support. This process is repeated until the support of

Algorithm 1: Oxx in K steps.

Input: $\boldsymbol{y}$, A

Output: $\hat{\boldsymbol{x}}$ with $\|\hat{\boldsymbol{x}}\|_0 \leq K$

Initialization: $\hat{\boldsymbol{x}} \leftarrow \boldsymbol{0}$; $S \leftarrow \emptyset$;

repeat

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

  Choose an atom ℓ according to either (3) or (4);

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

  $\hat{\boldsymbol{x}}_S \leftarrow A_S^\dagger \boldsymbol{y}$;

until $\mathrm{card}\,(S) = K$;

cardinality K is reached (see Algorithm 1). OMP and OLS share the same coefficient update and only differ by their selection rule:

$$\text{OMP:} \quad \ell \in \underset{i \notin S}{\arg\max} \, |\boldsymbol{r}^{\mathrm{t}} \boldsymbol{a}_i|, \tag{3}$$

$$\text{OLS:} \quad \ell \in \underset{i \notin S}{\arg\max} \, |\boldsymbol{r}^{\mathrm{t}} \tilde{\boldsymbol{b}}_i^S| \tag{4}$$

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

$$\ell \in \underset{i \notin S}{\arg\max} \, |\boldsymbol{r}^{\mathrm{t}} \tilde{\boldsymbol{a}}_i^S| \tag{5}$$

since $\boldsymbol{r}$ is orthogonal to $\mathrm{span}(A_S)$. 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 $\|\boldsymbol{r} - z_i \boldsymbol{a}_i\|^2$ w.r.t. scalar z_i, whereas OLS relies on the minimization of $\|\boldsymbol{y} - A_{S \cup \{i\}} \boldsymbol{z}\|^2$ w.r.t. vector $\boldsymbol{z}$ of length $\mathrm{card}\,(S) + 1$ [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, see, *e.g.,* [2], [4].

Let us state the MIP condition that holds for both OMP and OLS in the noiseless case. Note that under the assumption of Lemma II.1, the K-sparse representation is unique [29, Theorem 7].

Lemma II.1. *[12, Theorem 3.1-3.5], [13, Theorem 2]. Assume that* $\mu(A) < \frac{1}{2K-1}$. *Let* $\boldsymbol{y} = A\boldsymbol{x}^*$ *be a* K*-sparse representation. Then Oxx recover the support of* $\boldsymbol{x}^*$ *in* K *iterations.*

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 $|\boldsymbol{r}^{\mathrm{t}} \tilde{\boldsymbol{c}}_{i_1}^S| = |\boldsymbol{r}^{\mathrm{t}} \tilde{\boldsymbol{c}}_{i_2}^S|$) including an atom that

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

Input: $\boldsymbol{y}$, A

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

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

repeat

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

  Choose an atom ℓ according to prescribed selection rule (SR) in Table I;

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

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

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

TABLE I

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

NNOMP [23]	(SR)	$\underset{i \notin S}{\arg\max}\, \boldsymbol{r}^{\mathrm{t}} \boldsymbol{a}_i$
	(SC)	$\underset{i \notin S}{\max}\, \boldsymbol{r}^{\mathrm{t}} \boldsymbol{a}_i \leq 0$
SNNOLS [28]	(SR)	$\underset{i \notin S}{\arg\max}\, \boldsymbol{r}^{\mathrm{t}} \tilde{\boldsymbol{b}}_i^S$
	(SC)	$\underset{i \notin S}{\max}\, \boldsymbol{r}^{\mathrm{t}} \tilde{\boldsymbol{b}}_i^S \leq 0$
NNOLS [28]	(SR)	$\underset{i \notin S}{\arg\min} \left(\underset{\boldsymbol{z} \geq \mathbf{0}}{\min} \|\boldsymbol{y} - A_{S \cup \{i\}} \boldsymbol{z}\|^2 \right)$
	(SC)	$\underset{i \notin S}{\min} \left(\underset{\boldsymbol{z} \geq \mathbf{0}}{\min} \|\boldsymbol{y} - A_{S \cup \{i\}} \boldsymbol{z}\|^2 \right) = \|\boldsymbol{r}\|^2$

does not belong to the support of $\boldsymbol{x}^*$, 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 [30], microarray data analysis [31], mass spectroscopy [32] and fluid mechanics [33], to name a few, the sought solution is required to be non-negative. Several extensions of OMP and OLS to the non-negative setting have been studied so far, the main being NNOMP [23], [24], NNOLS and SNNOLS [28]. All three solve non-negative least squares (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 absolute value, *i.e.*, ignoring negative inner products. Similarly, the

selection rule of SNNOLS is the same as OLS without absolute values. On the contrary, NNOLS solves as many NNLS subproblems as candidate atoms to find the largest decrease of the residual error. Note that 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 [26], the reader can find a detailed study about the design and the practical implementation of non-negative orthogonal greedy algorithms. In particular, 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.* [23] 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], [27]. Here, we argue that this extension does not actually appear to be trivial.

Tropp's reasoning in [12] consists in lower bounding the absolute value of the inner product $|\boldsymbol{r}^{\mathrm{t}}\boldsymbol{a}_i|$ between the residual and the correct dictionary atoms based on induced matrix norm identities. Unfortunately, similar lower bounds cannot be directly extended when absolute values are dropped. Donoho *et al.*'s recursive proof [27] 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 $\boldsymbol{y} = A\boldsymbol{x}^*$ (*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 $\boldsymbol{r} = A(\boldsymbol{x}^* - \hat{\boldsymbol{x}}^{(k)})$ lays in the subspace spanned by the true atoms, where $\hat{\boldsymbol{x}}^{(k)}$ denotes the NNOMP iterate. Provided that the latter atoms are linearly independent, it is clear that $\boldsymbol{r}$ lays in their positive span if and only if $\hat{\boldsymbol{x}}^{(k)} \leq \boldsymbol{x}^*$. Simple numerical tests show that even for toy problems, $\hat{\boldsymbol{x}}^{(k)} \leq \boldsymbol{x}^*$ may not hold when $\boldsymbol{x}^* \geq \boldsymbol{0}$ 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) ℓ_1 analysis with non-negativity constraints: The theoretical analysis of non-negative versions of Basis Pursuit (BP: $\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1$ s.t. $\boldsymbol{y} = A\boldsymbol{x}$) and Basis Pursuit Denoising (BPDN: $\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|_1$) turns out to be far simpler than the analysis of orthogonal greedy algorithms. Indeed, it is well-known that contrary to greedy algorithms, the exact recovery analysis of Basis Pursuit heavily depends on the sign pattern. Fuchs [34] proved that when $\boldsymbol{y} = A\boldsymbol{x}^*$, BP and BPDN (for sufficiently small λ) both have a unique solution under the MIP assumption, and that this solution identifies with $\boldsymbol{x}^*$ as long as A_{S^*} is full rank and

$$\forall j \notin S^*, \, |\langle \boldsymbol{\sigma}^*, A_{S^*}^{\dagger} \boldsymbol{a}_j \rangle| < 1 \tag{6}$$

where $S^* := \mathrm{supp}(\boldsymbol{x}^*)$ and $\boldsymbol{\sigma}^* := \mathrm{sign}(\boldsymbol{x}^*_{S^*}) \in \{-1, 1\}^K$ denotes the sign pattern. It is noticeable that the latter condition[3] only depends on the sign pattern, but not on the magnitudes of coefficients x_i^*. Denoting by $\mathbf{1}$ the "all-ones" vector of size K, condition (6) with $\boldsymbol{\sigma}^* \leftarrow \mathbf{1}$ thus guarantees exact recovery of $\boldsymbol{x}^*$ for any $\boldsymbol{x}^* \geq \mathbf{0}$ supported by S^*. It follows that $\boldsymbol{x} = \boldsymbol{x}^*$ is not only the unique solution of BP/BPDN but also of their non-negative counterpart

$$\min_{\boldsymbol{x} \geq \mathbf{0}} \|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|_1, \tag{7}$$

often referred to as the non-negative Garrote in statistics [36]. Indeed, the minimum value of $\|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|_1$ over $\boldsymbol{x} \geq \mathbf{0}$ can be reached only for $\boldsymbol{x} = \boldsymbol{x}^*$. Summarizing, the sign preservation property of BP is guaranteed as long as Fuchs condition (6) is met.

3) Extension of K-step exact recovery analysis of ℓ_1 homotopy: Homotopy is a popular greedy algorithm dedicated to ℓ_1 minimization. It was originally proposed in [37] 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 $\min_{\boldsymbol{x}} \|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \lambda\|\boldsymbol{x}\|_1$ for a continuum of λ's. The principle of homotopy is to reconstruct the regularization path (defined as the set of solutions for all λ) by solving the ℓ_1 problem for gradually decreasing λ's starting from $\lambda = +\infty$, with the corresponding solution $\boldsymbol{x} = \mathbf{0}$. 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 $\boldsymbol{x}^*$ from noise-free observations $\boldsymbol{y} = A\boldsymbol{x}^*$ is guaranteed. Not only the support but also the magnitudes of $\boldsymbol{x}^*$ are found after K iterations (no deselection occurs).

[3]referred to as Fuchs corollary condition in [35] as opposed to the sharp (but more involved) Fuchs condition in [35, Theorem 2]. The latter condition solely depends on the sign pattern as well.

Although homotopy has a greedy structure, a major difference with Oxx algorithms is that homotopy solves an ℓ_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 $\mathrm{supp}(\boldsymbol{x}^*)$ 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 $\hat{\boldsymbol{x}}$ satisfies* $\mathrm{sign}(\hat{\boldsymbol{x}}_S) = \mathrm{sign}(A_S^{\mathrm{t}}\boldsymbol{r})$*, where* $S = \mathrm{supp}(\hat{\boldsymbol{x}})$. Donoho and Tsaig showed that when $\mu(A) < \frac{1}{2K-1}$, the magnitudes of the active atoms always increase while λ is decreasing. Since the correct magnitudes are exactly found at iteration K ($\hat{\boldsymbol{x}} = \boldsymbol{x}^*$), it follows that their sign is unchanged throughout the iterations. So, a sign preservation result in the ℓ_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 far less straightforward reasoning. Moreover, the sign agreement property "$\mathrm{sign}(\hat{\boldsymbol{x}}_S) = \mathrm{sign}(A_S^{\mathrm{t}}\boldsymbol{r})$" does not make sense anymore since $A_S^{\mathrm{t}}\boldsymbol{r} = \boldsymbol{0}$, 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, we have decomposed the proof of Theorem III.1 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 $\boldsymbol{y} = A\boldsymbol{x}^*$ be a K-sparse representation. Then Oxx recovers the support of $\boldsymbol{x}^*$ in K iterations, and at each iteration, the weights of selected atoms are of the same sign as the corresponding magnitudes of $\boldsymbol{x}^*$.*

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 $\boldsymbol{y} = A\boldsymbol{x}^*$ be a K-sparse representation with $\boldsymbol{x}^* \geq \boldsymbol{0}$. Then, Oxx recovers the support of $\boldsymbol{x}^*$ 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 $\boldsymbol{y} \in \mathbb{R}^m$ and assume that any combination of K dictionary columns is linearly independent.*

- *Assume that at every iteration $k = 1, \ldots, K$ of OMP, the selection rule yields a unique optimal index ℓ 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 $\boldsymbol{y} = A\boldsymbol{x}^*$ be a K-sparse representation with non-negative weights. Then NNOMP identifies with OMP whereas both NNOLS and SNNOLS identify with OLS. Thus, NNOMP, NNOLS and SNNOLS all recover the support of $\boldsymbol{x}^*$ 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], [38] by Oxx. Cai *et al.* [38] indeed exhibited an equiangular dictionary whose mutual coherence equals $\mu(A) = \frac{1}{2K-1}$ and a vector $\boldsymbol{y}$ having two K-sparse representations $\boldsymbol{y} = A\boldsymbol{x} = A\boldsymbol{z}$ with distinct supports. K-step exact support recovery does not make sense anymore in this situation, since either the support of $\boldsymbol{x}$ or $\boldsymbol{z}$ 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 $\boldsymbol{y}$ 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 $\boldsymbol{y} \in \mathbb{R}^m$ proposed in [38, Section III], the latter vector having two K-sparse representations $\boldsymbol{y} = A\boldsymbol{x} = A\boldsymbol{z}$ with distinct

supports. Since both supports are distinct, one can define the subrogate dictionary $A' \in \mathbb{R}^{m\times n}$ as:

$$\boldsymbol{a}'_i = \begin{cases} \operatorname{sign}(x_i)\,\boldsymbol{a}_i & \text{if } \ i \in \operatorname{supp}(\boldsymbol{x}), \\ \operatorname{sign}(z_i)\,\boldsymbol{a}_i & \text{if } \ i \in \operatorname{supp}(\boldsymbol{z}), \\ \boldsymbol{a}_i & \text{otherwise.} \end{cases} \tag{8}$$

Moreover, let $|\boldsymbol{x}|$ and $|\boldsymbol{z}| \in \mathbb{R}^n$ denote the non-negative vectors whose entries are equal to $|x_i|$ and $|z_i|$, respectively. Obviously, $\boldsymbol{y} = A'|\boldsymbol{x}| = A'|\boldsymbol{z}|$, 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.

Before going further, let us denote by

$$\mathcal{C}_K^+ := \{A\boldsymbol{x}^*,\ \boldsymbol{x}^* > \boldsymbol{0} \text{ and } \|\boldsymbol{x}^*\|_0 = K\} \tag{9}$$

the set of K-sparse representations with positive weights, where the ℓ_0-"norm" $\|\,.\,\|_0$ counts the number of nonzero elements. Now, consider $\boldsymbol{y} = A\boldsymbol{x}^* \in \mathcal{C}_K^+$. From Lemma II.1, Oxx recovers the support S^* of $\boldsymbol{x}^*$ in K iterations. At any iteration $k \leq K$, the support S_k of the current solution $\hat{\boldsymbol{x}}^{(k)}$ is therefore a subset of S^*. Recall that $\hat{\boldsymbol{x}}^{(k)}_{S_k}$ is the unconstrained least squares solution related to support S_k, see Algorithm 1. Let $\boldsymbol{r}_k = P^{\perp}_{S_k}\boldsymbol{y}$ denote the related residual vector, with $\boldsymbol{r}_0 = \boldsymbol{y}$ corresponding to $S_0 = \emptyset$.

We proceed in two distinct steps to prove that $\hat{\boldsymbol{x}}^{(k)}_{S_k} > \boldsymbol{0}$ for all $k \in \{1, \ldots, K\}$. First, we prove that the weight of each newly selected atom $\hat{x}^{(k)}_{S_k\setminus S_{k-1}}$ is positive for any $k \leq K$. Then, we show that the updated coefficients $\hat{\boldsymbol{x}}^{(k)}_{S_{k-1}}$ remain positive. Let us remark that Theorem III.1 states a trivial fact at iteration K, since $\hat{\boldsymbol{x}}^{(K)} = \boldsymbol{x}^*$ according to Lemma II.1.

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

Lemma III.2. *Let $\boldsymbol{y} \in \mathbb{R}^m$, 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*

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

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)**.** *Let* $\boldsymbol{y} \in \mathcal{C}_K^+$ *and assume that* $\mu(A) < \frac{1}{2K-1}$*. For all* $k \in \{1, \ldots, K\}$*,* $\hat{x}^{(k)}_{S_k \setminus S_{k-1}} > 0$*.*

Proof: See Appendix VI-D. ■

Lemma III.4 (non-negativity of updated coefficients)**.** *Let* $\boldsymbol{y} \in \mathcal{C}_K^+$ *and assume that* $\mu(A) < \frac{1}{2K-1}$*. For all* $k \in \{2, \ldots, K\}$*,* $\hat{\boldsymbol{x}}^{(k)}_{S_{k-1}} > \boldsymbol{0}$*.*

Proof: See Appendix VI-E. ■

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 $\boldsymbol{y} = A\boldsymbol{x}^* + \boldsymbol{n}$ with some additive Gaussian noise $\boldsymbol{n}$. 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 $10 \log_{10}(\|A\boldsymbol{x}^*\|^2/\|\boldsymbol{n}\|^2)$), 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

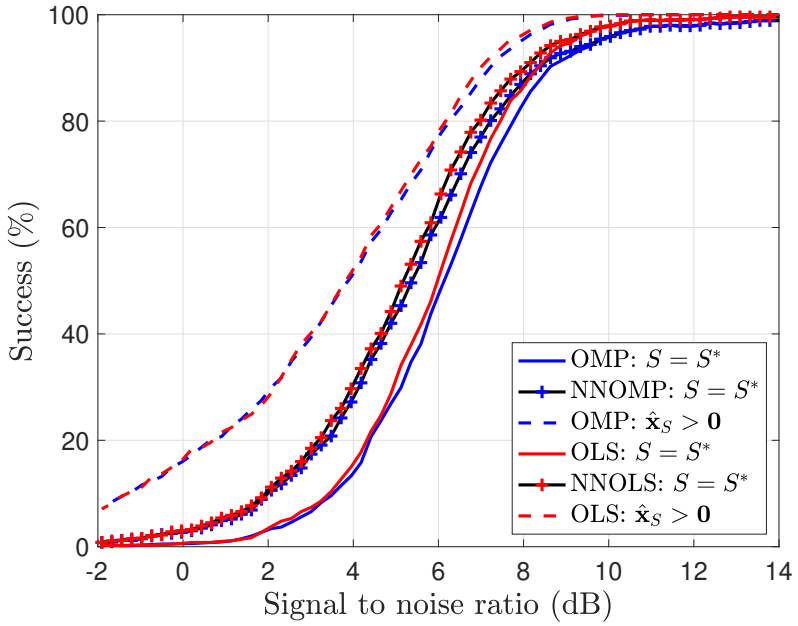
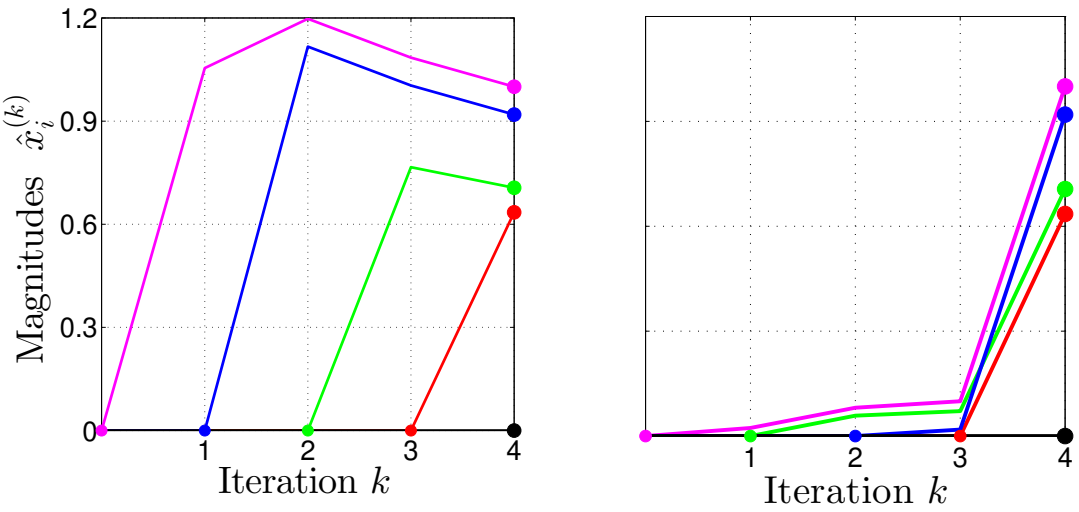


Fig. 1. Rate of exact support recovery ($S_K = S^*$) and non-negativity at all iterations ($\forall k$, $\hat{\boldsymbol{x}}^{(k)} \geq \mathbf{0}$) w.r.t. the signal-to-noise ratio for a simulated data experiment involving Gaussian-shaped atoms. SNNOLS (not shown here) yields the exact same curve as 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, ℓ_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 the iterations. This claim can be proven analytically at iteration 2, by using the fact that $\hat{x}_{S_1}^{(1)} > 0$ and $\hat{\boldsymbol{x}}_{S_2}^{(2)} > \mathbf{0}$ according to Corollary III.1. Since $\hat{\boldsymbol{x}}^{(1)}$ and $\hat{\boldsymbol{x}}^{(2)}$ 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 the inner product. We further compare ℓ_1-homotopy with the 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



(a) OMP/OLS (b) ℓ_1-homotopy

Fig. 2. Behavior of Oxx and ℓ_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 ℓ_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 ($\hat{\boldsymbol{x}}^{(k)} = \boldsymbol{x}^*$).

of A satisfy $\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle = \pm\mu(A)$, where the sign of the inner product is randomly chosen. The ground truth vector $\boldsymbol{x}^*$ 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, which corresponds to the wrong atom $i \notin S^*$, is equal to 0 throughout the iterations. ℓ_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. Conclusion

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. There is still room for improvement though in cases where the dictionary atoms are known to be non-negative valued, since then the MIP condition is not guaranteed to be necessary anymore. This setting covers many application fields ranging from sparse deconvolution [39] to tomographic image reconstruction [40].

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 for noisy observations, Oxx algorithms do yield iterates with negative entries, thus sign-aware versions are worth being considered [26]. Furthermore, the latter versions yield improvement of empirical exact recovery performance for random dictionaries when $\mu(A) > \frac{1}{2K-1}$.

A first perspective of this work is to elaborate on exact recovery conditions when $\mu(A) \geq \frac{1}{2K-1}$. We argued that the MIP condition $\mu(A) < \frac{1}{2K-1}$ is both necessary and sufficient for *uniform* recovery of K-sparse signals, that is, irrespective of the magnitudes of the nonzero coefficients in the K-sparse representation. It is well-known that such worst case analysis is pessimistic since practically, algorithms may succeed far beyond the exact recovery regime $\mu(A) < \frac{1}{2K-1}$. It was shown in [16] that exact recovery guarantees can be obtained for Oxx when $\mu(A) \in [\frac{1}{2K-1}, \frac{1}{K})$ provided that the decay of the magnitudes of the nonzero coefficients is fast enough (the minimum rate of decay required to ensure exact support recovery depends on $\mu(A)$). It would then be interesting to generalize our sign preservation analysis to the case $\mu(A) \in [\frac{1}{2K-1}, \frac{1}{K})$. However, this extension does not appear to be obvious and is left for future work. With the same idea of elaborating on sign preservation under weaker K-step support recovery conditions, Tropp's exact recovery condition (ERC) is worth being considered. Indeed, it is both necessary and sufficient for uniform K-step exact recovery of all representations having a given support S^* with both OMP and OLS [11], [12]. In all our empirical tests, we found that sign preservation holds whenever the ERC is met. However, proving this conjecture would necessitate more involved theoretical analysis than those we derived in this paper.

Another perspective is to address the sign-preservation analysis of Oxx and the exact recovery analysis of non-negative greedy algorithms for noisy observations. Here, the noisy case was addressed from the empirical viewpoint. Theoretical analyses of greedy algorithms for noisy data, see *e.g.,* [14], [15], could be generalized to sign-aware versions, leading to conditions depending on both the mutual coherence and the noise level for various kinds of noise distributions.

VI. APPENDIX

A. Useful lemmas

Let us recall some useful lemmas. Lemma VI.1 provides an upper bound on the ℓ_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. *[41, 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*

$$\|(B^{\mathrm{t}}B)^{-1}\|_{1,1} \leq \frac{1}{1-(k-1)\mu(B)} \tag{11}$$

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 $k = \operatorname{card}(S)$ then*

$$\forall i \notin S,\ \|\tilde{\boldsymbol{a}}_i^S\|^2 \geq \beta_k, \tag{12}$$

$$\forall p \neq q,\ |(\tilde{\boldsymbol{c}}_p^S)^{\mathrm{t}}\tilde{\boldsymbol{a}}_q^S| \leq \mu_k\eta_k, \tag{13}$$

where

$$\beta_k = \frac{(\mu(A)+1)(1-k\mu(A))}{1-(k-1)\mu(A)} \tag{14}$$

$$\mu_k = \frac{\mu(A)}{1-k\mu(A)} \tag{15}$$

$$\eta_k = \begin{cases} \beta_k & \textit{for OMP,} \\ \sqrt{\beta_k} & \textit{for OLS.} \end{cases} \tag{16}$$

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 $A = \left[\boldsymbol{a}_1, A_2\right] \in \mathbb{R}^{p\times q}$ where $\boldsymbol{a}_1 \in \mathbb{R}^p$ and $A_2 \in \mathbb{R}^{p\times(q-1)}$ denotes the submatrix formed of the last $q-1$ column vectors. Then,*

$$\forall \boldsymbol{r} \in \mathbb{R}^p,\ \left(A^{\dagger}\boldsymbol{r}\right)_1 = \frac{\langle \boldsymbol{r}, P_{S_2}^{\perp}\boldsymbol{a}_1\rangle}{\|P_{S_2}^{\perp}\boldsymbol{a}_1\|^2}. \tag{17}$$

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

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

$$\boldsymbol{r} = \boldsymbol{p}_A + \boldsymbol{p}_{A^{\perp}} = (A^{\dagger}\boldsymbol{r})_1\boldsymbol{a}_1 + \boldsymbol{r}_{A_2} + \boldsymbol{p}_{A^{\perp}} \tag{18}$$

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

$$\boldsymbol{r} = \left(A^\dagger \boldsymbol{r}\right)_1 P_{S_2}^\perp \boldsymbol{a}_1 + (\boldsymbol{r}_{A_2} + P_{S_2}\boldsymbol{a}_1) + \boldsymbol{p}_{A^\perp}. \tag{19}$$

(17) is obtained directly from (19) by writing the inner product $\langle \boldsymbol{r}, P_{S_2}^\perp \boldsymbol{a}_1 \rangle$ because the latter decomposition is orthogonal. The denominator in (17) 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 $\hat{\boldsymbol{x}}_{S_k}^{(k)}$ is the unconstrained least squares solution related to subset S_k, we have for $j < k$,

$$\begin{aligned} \hat{\boldsymbol{x}}_{S_k}^{(k)} &= \underset{\boldsymbol{z}\in\mathbb{R}^k}{\arg\min} \|\boldsymbol{y} - A_{S_k}\boldsymbol{z}\|^2 \\ &= \underset{\boldsymbol{v},\boldsymbol{w}}{\arg\min} \|\boldsymbol{y} - A_{S_j}\boldsymbol{v} - A_{S_k\setminus S_j}\boldsymbol{w}\|^2. \end{aligned} \tag{20}$$

In addition,

$$\begin{aligned} \min_{\boldsymbol{v},\boldsymbol{w}} \|\boldsymbol{y} - A_{S_j}\boldsymbol{v} - A_{S_k\setminus S_j}\boldsymbol{w}\|^2 &= \min_{\boldsymbol{w}}\big(\min_{\boldsymbol{v}} \|(\boldsymbol{y} - A_{S_k\setminus S_j}\boldsymbol{w}) - A_{S_j}\boldsymbol{v}\|^2\big) \\ &= \min_{\boldsymbol{w}} \|P_{S_j}^\perp(\boldsymbol{y} - A_{S_k\setminus S_j}\boldsymbol{w})\|^2 \\ &= \min_{\boldsymbol{w}} \|\boldsymbol{r}_j - \tilde{A}_{S_k\setminus S_j}^{S_j}\boldsymbol{w}\|^2. \end{aligned} \tag{21}$$

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 (21) is reached for

$$\boldsymbol{w} = \left[\tilde{A}_{S_k\setminus S_j}^{S_j}\right]^\dagger \boldsymbol{r}_j \tag{22}$$

which identifies with $\hat{\boldsymbol{x}}_{S_k\setminus S_j}^{(k)}$ according to (20).

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 ℓ the index of the atom selected by OMP. The first OMP iterate is the one-sparse vector supported by $\{\ell\}$ such that $\hat{x}_\ell^{(1)} = \boldsymbol{a}_\ell^{\mathrm{t}}\boldsymbol{y}$. By assumption, $\hat{x}_\ell^{(1)} > 0$. Since ℓ is the unique solution to (3) with $\boldsymbol{r}_0 = \boldsymbol{y}$, we have:

$$\{\ell\} = \arg\max_i \{\boldsymbol{a}_i^{\mathrm{t}}\boldsymbol{y}\}, \tag{23}$$

which implies that ℓ is also selected by NNOMP. Since $\boldsymbol{a}_\ell^{\mathrm{t}}\boldsymbol{y} > 0$, $\hat{\boldsymbol{x}}^{(1)} \geq \boldsymbol{0}$ 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 $\boldsymbol{r}_{k-1}$ 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 ℓ the index of the atom selected at iteration k, and let $\hat{\boldsymbol{x}}^{(k)}$ denote the OMP iterate. Lemma III.2 with $j \leftarrow k-1$ implies that:

$$\begin{aligned}
\hat{x}_\ell^{(k)} &= \left[\tilde{\boldsymbol{a}}_\ell^{S_{k-1}}\right]^\dagger \boldsymbol{r}_{k-1} \\
&= \frac{\boldsymbol{r}_{k-1}^{\mathrm{t}}\tilde{\boldsymbol{a}}_\ell^{S_{k-1}}}{\|\tilde{\boldsymbol{a}}_\ell^{S_{k-1}}\|^2} && (24) \\
&= \frac{\boldsymbol{r}_{k-1}^{\mathrm{t}}\boldsymbol{a}_\ell}{\|\tilde{\boldsymbol{a}}_\ell^{S_{k-1}}\|^2}. && (25)
\end{aligned}$$

$\tilde{\boldsymbol{a}}_\ell^{S_{k-1}} \neq \boldsymbol{0}$ in (24) follows from Lemma VI.3 and the full rankness assumption of A_{S_k}. (25) holds because $\tilde{\boldsymbol{a}}_\ell^{S_{k-1}} - \boldsymbol{a}_\ell$ lays in $\mathrm{span}(A_{S_{k-1}})$, and the OMP residual $\boldsymbol{r}_{k-1}$ is orthogonal to $\mathrm{span}(A_{S_{k-1}})$. By assumption, $\hat{\boldsymbol{x}}^{(k)} \geq \boldsymbol{0}$, so (25) implies that $\boldsymbol{r}_{k-1}^{\mathrm{t}}\boldsymbol{a}_\ell \geq 0$. Since ℓ is the unique solution to (3), we have:

$$\{\ell\} = \arg\max_{i \notin S_{k-1}} \{\boldsymbol{r}_{k-1}^{\mathrm{t}}\boldsymbol{a}_i\}. \tag{26}$$

So, ℓ 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 $\boldsymbol{a}_i$ by their normalized projections $\tilde{\boldsymbol{b}}_i^S$.

At the first iteration, $\tilde{\boldsymbol{b}}_i^\emptyset = \tilde{\boldsymbol{a}}_i^\emptyset = \boldsymbol{a}_i$, 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 $\hat{\boldsymbol{x}}^{(k)}$ now denotes the OLS iterate. The right-hand side of (24) rereads $\left(\boldsymbol{r}_{k-1}^{\mathrm{t}}\tilde{\boldsymbol{b}}_\ell^{S_{k-1}}\right)/\|\tilde{\boldsymbol{a}}_\ell^{S_{k-1}}\| \geq 0$, so $\boldsymbol{r}_{k-1}^{\mathrm{t}}\tilde{\boldsymbol{b}}_\ell^{S_{k-1}} \geq 0$. Finally, since ℓ is the only maximizer of the OLS selection rule (4), we have

$$\{\ell\} = \arg\max_{i \notin S_{k-1}} \left\{\boldsymbol{r}_{k-1}^{\mathrm{t}}\tilde{\boldsymbol{b}}_i^{S_{k-1}}\right\}. \tag{27}$$

So, ℓ 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 $\hat{\boldsymbol{x}}^{(k)}_{S_{k-1}} = A^{\dagger}_{S_{k-1}}\boldsymbol{y} > \boldsymbol{0}$. Let ℓ 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 ℓ is unique, we have

$$\forall i \notin S_k,\, \min_{\boldsymbol{z}} \|\boldsymbol{y} - A_{S_k}\boldsymbol{z}\|^2 < \min_{\boldsymbol{z}} \|\boldsymbol{y} - A_{S_{k-1}\cup\{i\}}\boldsymbol{z}\|^2. \tag{28}$$

Clearly,

$$\min_{\boldsymbol{z}} \|\boldsymbol{y} - A_{S_{k-1}\cup\{i\}}\boldsymbol{z}\|^2 \leq \min_{\boldsymbol{z}\geq\boldsymbol{0}} \|\boldsymbol{y} - A_{S_{k-1}\cup\{i\}}\boldsymbol{z}\|^2. \tag{29}$$

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

$$\min_{\boldsymbol{z}} \|\boldsymbol{y} - A_{S_k}\boldsymbol{z}\|^2 = \min_{\boldsymbol{z}\geq\boldsymbol{0}} \|\boldsymbol{y} - A_{S_k}\boldsymbol{z}\|^2. \tag{30}$$

From (28)-(30), we get

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

which implies that ℓ 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. (30) implies that they coincide. Hence, OLS and NNOLS yield the same iterate at iteration k.

D. Proof of Lemma III.3

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

First, the claim is obvious for $K = 1$. We therefore address the case where $K \geq 2$ hereafter. Also, 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

$$\hat{x}^{(k)}_{S_k\setminus S_{k-1}} = \tilde{\boldsymbol{a}}_{\ell}^{\mathrm{t}}\boldsymbol{r}_{k-1}. \tag{31}$$

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

Since $\boldsymbol{a}_\ell$ is selected at the k-th iteration, we have

$$\forall j \notin S_k,\ |\tilde{\boldsymbol{c}}_\ell^{\mathrm{t}} \boldsymbol{r}_{k-1}| \geq |\tilde{\boldsymbol{c}}_j^{\mathrm{t}} \boldsymbol{r}_{k-1}|. \tag{32}$$

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

$$\tilde{\boldsymbol{c}}_\ell^{\mathrm{t}} \boldsymbol{r}_{k-1} > -|\tilde{\boldsymbol{c}}_p^{\mathrm{t}} \boldsymbol{r}_{k-1}|. \tag{33}$$

(33) and (32) with $j \leftarrow p$ indeed imply that $\tilde{\boldsymbol{c}}_\ell^{\mathrm{t}} \boldsymbol{r}_{k-1} > 0$.

Let us introduce the decomposition

$$\boldsymbol{r}_{k-1} = P_{S_{k-1}}^{\perp} \boldsymbol{y} = \sum_{i \notin S_{k-1}} x_i^* \tilde{\boldsymbol{a}}_i. \tag{34}$$

For any $j \in S^* \backslash S_{k-1}$, we deduce

$$\tilde{\boldsymbol{c}}_j^{\mathrm{t}} \boldsymbol{r}_{k-1} = x_j^* \tilde{\boldsymbol{a}}_j^{\mathrm{t}} \tilde{\boldsymbol{c}}_j + \sum_{i \notin S_{k-1} \cup \{j\}} x_i^* \tilde{\boldsymbol{a}}_i^{\mathrm{t}} \tilde{\boldsymbol{c}}_j. \tag{35}$$

Clearly,

$$\tilde{\boldsymbol{a}}_j^{\mathrm{t}} \tilde{\boldsymbol{c}}_j = \begin{cases} \|\tilde{\boldsymbol{a}}_j\|^2 & \text{for OMP,} \\ \|\tilde{\boldsymbol{a}}_j\| & \text{for OLS.} \end{cases} \tag{36}$$

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

$$\tilde{\boldsymbol{a}}_j^{\mathrm{t}} \tilde{\boldsymbol{c}}_j \geq \eta_{k-1}, \tag{37}$$

$$|\tilde{\boldsymbol{a}}_i^{\mathrm{t}} \tilde{\boldsymbol{c}}_j| \leq \mu_{k-1} \eta_{k-1} \tag{38}$$

with μ_{k-1} and η_{k-1} defined in (15)-(16). From (35), we get for $j \in S^* \backslash S_{k-1}$,

$$\tilde{\boldsymbol{c}}_j^{\mathrm{t}} \boldsymbol{r}_{k-1} \geq \eta_{k-1} \Big(x_j^* - \mu_{k-1} \sum_{i \notin S_{k-1} \cup \{j\}} x_i^* \Big). \tag{39}$$

Notice that there are $K - k$ nonzero terms in the sum in (39) since $\boldsymbol{x}^*$ is K-sparse. The latter sum can then be upper bounded as:

$$\sum_{i \notin S_{k-1} \cup \{j\}} x_i^* \leq \begin{cases} x_\ell^* + (K-k-1) x_p^* & \text{if } j \neq \ell, \\ (K-k) x_p^* & \text{if } j = \ell \end{cases} \tag{40}$$

with p defined as:

$$p \in \underset{i \in S^* \backslash S_k}{\arg\max}\, x_i^*. \tag{41}$$

Apply (39) twice with $j \leftarrow p$ and $j \leftarrow \ell$. We get

$$\tilde{\boldsymbol{c}}_p^{\mathrm{t}} \boldsymbol{r}_{k-1} \geq \eta_{k-1} \big((1 - (K-k-1)\mu_{k-1}) x_p^* - \mu_{k-1} x_\ell^* \big),$$

$$\tilde{\boldsymbol{c}}_\ell^{\mathrm{t}} \boldsymbol{r}_{k-1} \geq \eta_{k-1} \big(x_\ell^* - (K-k)\mu_{k-1} x_p^* \big),$$

from which we deduce

$$\tilde{\boldsymbol{c}}_p^{\mathrm{t}}\boldsymbol{r}_{k-1} + \tilde{\boldsymbol{c}}_\ell^{\mathrm{t}}\boldsymbol{r}_{k-1} \geq \eta_{k-1}(\tau_1 x_\ell^* + \tau_2 x_p^*), \tag{42}$$

with

$$\tau_1 := 1 - \mu_{k-1}, \tag{43}$$

$$\tau_2 := 1 - (2K - 2k - 1)\mu_{k-1}. \tag{44}$$

According to (15), $\tau_1 > 0$ if $\mu(A) < \frac{1}{k}$ and $\tau_2 > 0$ if $(2K - k - 2)\mu(A) < 1$. Since $\mu(A) < \frac{1}{2K-1}$, it holds true that $\tau_1 > 0$ and $\tau_2 > 0$. Hence, from (42) we get $\tilde{\boldsymbol{c}}_p^{\mathrm{t}}\boldsymbol{r}_{k-1} + \tilde{\boldsymbol{c}}_\ell^{\mathrm{t}}\boldsymbol{r}_{k-1} > 0$, which implies (33). We thus conclude that $\tilde{\boldsymbol{c}}_\ell^{\mathrm{t}}\boldsymbol{r}_{k-1}$ and $\hat{x}^{(k)}_{S_k\backslash S_{k-1}}$ are positive quantities.

E. Proof of Lemma III.4

In this proof, the abridged notations $\tilde{\boldsymbol{a}}_i$, $\tilde{\boldsymbol{b}}_i$, $\tilde{\boldsymbol{c}}_i$ correspond to projected vectors onto $\mathrm{span}(A_{j-1})^\perp$ (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 $G_S^{\mathrm{OMP}} := \tilde{A}_S^{\mathrm{t}}\tilde{A}_S$ and $G_S^{\mathrm{OLS}} := \tilde{B}_S^{\mathrm{t}}\tilde{B}_S$. For statements that are common to both OMP and OLS, we will use the simpler generic notation $G_S := \tilde{C}_S^{\mathrm{t}}\tilde{C}_S$.

The proof of Lemma III.4 is not recursive. However, in order to prove that $\hat{\boldsymbol{x}}^{(k)}_{S_{k-1}} > \mathbf{0}$, we use the fact that $\hat{x}^{(j)}_{S_j\backslash S_{j-1}} > 0$ 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 ℓ has been selected by Oxx, so that $S_j\backslash S_{j-1} = \{\ell\}$. According to Lemma III.2 and since ℓ is the first entry in the ordered set $S_k\backslash S_{j-1}$, we have

$$\hat{x}_\ell^{(k)} = \big(\tilde{A}^\dagger_{S_k\backslash S_{j-1}}\boldsymbol{r}_{j-1}\big)_1. \tag{45}$$

In order to exploit the Oxx selection rule which is based on the atoms $\tilde{c}_i$ defined in (1), let us rewrite (45) with respect to matrix $\tilde{C}$. Obviously, one can rewrite $\tilde{A}_{S_k\backslash S_{j-1}} = \tilde{C}_{S_k\backslash S_{j-1}}\Delta$, 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 $\|\tilde{\boldsymbol{a}}_i\| > 0$, $i \in S_k\backslash S_{j-1}$ (OLS, $\tilde{C} \leftarrow \tilde{B}$). The positivity property is related to the full rankness of $\tilde{C}_{S_k\backslash S_{j-1}}$, which deduces from that of A_{S_k} according to Lemma VI.3. From standard properties of pseudo-inverses, we have $\tilde{A}^\dagger_{S_k\backslash S_{j-1}} = \Delta^{-1}\tilde{C}^\dagger_{S_k\backslash S_{j-1}}$. (45) yields

$$\begin{aligned}
\hat{x}_\ell^{(k)} \;&\propto^+\; \big(\tilde{C}^\dagger_{S_k\backslash S_{j-1}}\boldsymbol{r}_{j-1}\big)_1 \\
&=\; \Big(\big[\tilde{\boldsymbol{c}}_\ell, \tilde{C}_{S_k\backslash S_j}\big]^\dagger \boldsymbol{r}_{j-1}\Big)_1 \qquad (46)\\
&=\; \frac{\big\langle \boldsymbol{r}_{j-1}, \tilde{P}^\perp_{S_k\backslash S_j}\tilde{\boldsymbol{c}}_\ell\big\rangle}{\big\|\tilde{P}^\perp_{S_k\backslash S_j}\tilde{\boldsymbol{c}}_\ell\big\|^2} \qquad (47)
\end{aligned}$$

where $\propto^+$ indicates proportionality up to a positive factor, $\tilde{P}^{\perp}_{S_k\backslash S_j}$ denotes the orthogonal projection onto the orthogonal complement of $\mathrm{span}(\tilde{C}_{S_k\backslash S_j})$, and (47) deduces from Lemma VI.4.

(47) implies that:

$$\hat{x}_\ell^{(k)} > 0 \iff \langle \tilde{P}^{\perp}_{S_k\backslash S_j} \boldsymbol{r}_{j-1}, \tilde{\boldsymbol{c}}_\ell \rangle > 0 \tag{48}$$

$$\iff \langle \tilde{C}_{S_k\backslash S_j} \tilde{C}^{\dagger}_{S_k\backslash S_j} \boldsymbol{r}_{j-1}, \tilde{\boldsymbol{c}}_\ell \rangle < \langle \boldsymbol{r}_{j-1}, \tilde{\boldsymbol{c}}_\ell \rangle \tag{49}$$

$$\iff \langle \tilde{C}^{\dagger}_{S_k\backslash S_j} \boldsymbol{r}_{j-1}, \tilde{C}^{\mathrm{t}}_{S_k\backslash S_j} \tilde{\boldsymbol{c}}_\ell \rangle < \langle \boldsymbol{r}_{j-1}, \tilde{\boldsymbol{c}}_\ell \rangle. \tag{50}$$

By Hölder's inequality and since $\tilde{C}^{\dagger}_{S_k\backslash S_j} = G^{-1}_{S_k\backslash S_j} \tilde{C}^{\mathrm{t}}_{S_k\backslash S_j}$, the left-hand side (LHS) of (50) is upper bounded by

$$\begin{aligned} \|\tilde{C}^{\mathrm{t}}_{S_k\backslash S_j} \tilde{\boldsymbol{c}}_\ell\|_\infty \|\tilde{C}^{\dagger}_{S_k\backslash S_j} \boldsymbol{r}_{j-1}\|_1 &\leq \|\tilde{C}^{\mathrm{t}}_{S_k\backslash S_j} \tilde{\boldsymbol{c}}_\ell\|_\infty \|G^{-1}_{S_k\backslash S_j}\|_{1,1} \|\tilde{C}^{\mathrm{t}}_{S_k\backslash S_j} \boldsymbol{r}_{j-1}\|_1 \\ &\leq \mu(\tilde{C}) \, \|G^{-1}_{S_k\backslash S_j}\|_{1,1} \|\tilde{C}^{\mathrm{t}}_{S_k\backslash S_j} \boldsymbol{r}_{j-1}\|_1 \\ &\leq \mu(\tilde{C}) \, \|G^{-1}_{S_k\backslash S_j}\|_{1,1} \, (k-j) \, |\langle \boldsymbol{r}_{j-1}, \tilde{\boldsymbol{c}}_\ell \rangle|. \end{aligned} \tag{51}$$

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

$$\forall i \in S_k \backslash S_j, \ |\langle \boldsymbol{r}_{j-1}, \tilde{\boldsymbol{c}}_i \rangle| \leq |\langle \boldsymbol{r}_{j-1}, \tilde{\boldsymbol{c}}_\ell \rangle|. \tag{52}$$

From Lemma III.3, we have $\hat{x}_\ell^{(j)} > 0$, and as already remarked in the proof of Lemma III.3, $\tilde{\boldsymbol{c}}_\ell^{\mathrm{t}} \boldsymbol{r}_{j-1}$ is of same sign as $\hat{x}_\ell^{(j)}$, see (31). The upper bound in (51) thus rewrites

$$(k-j) \, \mu(\tilde{C}) \, \|G^{-1}_{S_k\backslash S_j}\|_{1,1} \, \langle \boldsymbol{r}_{j-1}, \tilde{\boldsymbol{c}}_\ell \rangle \tag{53}$$

and we deduce from (50) that

$$\big((k-j) \mu(\tilde{C}) \|G^{-1}_{S_k\backslash S_j}\|_{1,1} < 1 \big) \Longrightarrow \big(\hat{x}_\ell^{(k)} > 0 \big). \tag{54}$$

Let us now provide some upper bounds on $\mu(\tilde{C})$ and $\|G^{-1}_{S_k\backslash S_j}\|_{1,1}$ in order to show that the LHS of (54) holds true.

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

$$\forall i \notin S_{j-1}, \ \|\tilde{\boldsymbol{a}}_i\|^2 \geq \beta_{j-1}, \tag{55}$$

$$\forall p \neq q, \ |\tilde{\boldsymbol{a}}_p^{\mathrm{t}} \tilde{\boldsymbol{a}}_q| \leq \mu_{j-1} \beta_{j-1}, \tag{56}$$

with μ_{j-1} and β_{j-1} defined in (14)-(15), from which we can easily deduce that

$$\mu(\tilde{C}) = \begin{cases} \mu(\tilde{A}) \ \leq \ \mu_{j-1}\beta_{j-1} & \text{(OMP case)}, \\ \mu(\tilde{B}) \ \leq \mu_{j-1} & \text{(OLS case)}. \end{cases} \tag{57}$$

2) Upper bound on $\|G^{-1}_{S_k\backslash S_j}\|_{1,1}$ in the OLS case: We have shown that for $i \in S_k\backslash S_j$, $\tilde{\boldsymbol{a}}_i \neq \mathbf{0}$, so matrix $\tilde{B}_{S_k\backslash S_j}$ is column-normalized, hence $G^{\mathrm{OLS}}_{S_k\backslash S_j}$ has a unit diagonal. Since $\mu(A) < \frac{1}{2K-1}$, (57) and (15) imply that:

$$\mu(\tilde{B}) \leq \mu_{j-1} \leq \frac{1}{2K-j} \leq \frac{1}{k-j}. \tag{58}$$

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

$$\big\|\big[G^{\mathrm{OLS}}_{S_k\backslash S_j}\big]^{-1}\big\|_{1,1} \leq \frac{1}{1-(k-j-1)\mu(\tilde{B}_{S_k\backslash S_j})} \leq \frac{1}{1-(k-j-1)\mu(\tilde{B})}. \tag{59}$$

It follows from (57) that

$$\begin{aligned}
\mu(\tilde{B})\big\|\big[G^{\mathrm{OLS}}_{S_k\backslash S_j}\big]^{-1}\big\|_{1,1} &\leq \mu_{j-1}\big\|\big[G^{\mathrm{OLS}}_{S_k\backslash S_j}\big]^{-1}\big\|_{1,1} \\
&\leq \frac{\mu_{j-1}}{1-(k-j-1)\mu(\tilde{B})}, \\
&\leq \frac{1}{2K-k+1} &(60)\\
&< \frac{1}{k-j} &(61)
\end{aligned}$$

where (60) follows from the second upper bound in (58). (61) thus implies that the LHS of (54) is true in the OLS case.

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

$$\big[G^{\mathrm{OMP}}_{S_k\backslash S_j}\big]^{-1} = \Delta^{-1}\big[G^{\mathrm{OLS}}_{S_k\backslash S_j}\big]^{-1}\Delta^{-1}, \tag{62}$$

and thus

$$\big\|\big[G^{\mathrm{OMP}}_{S_k\backslash S_j}\big]^{-1}\big\|_{1,1} \leq \big\|\big[G^{\mathrm{OLS}}_{S_k\backslash S_j}\big]^{-1}\big\|_{1,1}\,\|\Delta^{-1}\|^2_{1,1}. \tag{63}$$

Moreover,

$$\|\Delta^{-1}\|^2_{1,1} = \frac{1}{\min_{i\in S_k\backslash S_j}\|\tilde{\boldsymbol{a}}_i\|^2} \leq \frac{1}{\beta_{j-1}} \tag{64}$$

by (55). We have thus

$$\mu(\tilde{A})\big\|\big[G^{\mathrm{OMP}}_{S_k\backslash S_j}\big]^{-1}\big\|_{1,1} \leq \frac{\mu(\tilde{A})}{\beta_{j-1}}\big\|\big[G^{\mathrm{OLS}}_{S_k\backslash S_j}\big]^{-1}\big\|_{1,1} \leq \mu_{j-1}\big\|\big[G^{\mathrm{OLS}}_{S_k\backslash S_j}\big]^{-1}\big\|_{1,1}$$

according to (57). From (61), we conclude that the LHS of (54) is also true in the OMP case.

REFERENCES

[1] Y. C. Pati, R. Rezaiifar, 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 ℓ_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] 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.

[22] 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.

[23] 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.

[24] 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.

[25] D. Kim and J. P. Haldar, "Greedy algorithms for nonnegativity-constrained simultaneous sparse recovery", *Signal Process.*, vol. 125, pp. 274–289, 2016.

[26] T. T. Nguyen, J. Idier, C. Soussen, and E.-H. Djermoune, "Non-negative orthogonal greedy algorithms", Research Report, http://webpages.lss.supelec.fr/charles.soussen/Nguyen19a.pdf, Université de Lorraine, LS2N, CentraleSupélec, Nancy, France, Jan. 2019.

[27] 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.

[28] 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.

[29] D. L. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization", *Proc. Natl. Acad. Sci.* USA, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[30] 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.

[31] 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.

[32] A. Cherni, E. Chouzenoux, and M.-A. Delsuc, "Fast dictionnary-based approach for mass spectrometry data analysis", in *Proc. IEEE ICASSP*, Calgary, Canada, Apr. 2018, pp. 816–820.

[33] 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.

[34] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases", *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.

[35] 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.

[36] A. J. Miller, *Subset selection in regression*, Chapman and Hall, London, UK, 2nd edition, Apr. 2002.

[37] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression", *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, Apr. 2004.

[38] 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.

[39] T. Bendory, "Robust recovery of positive stream of pulses", *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2114–2122, Apr. 2017.

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

[41] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*, Applied and Numerical Harmonic Analysis. Birkhäuser, Basel, June 2013.