



HAL
open science

K-step analysis of orthogonal greedy algorithms for non-negative sparse representations

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

► **To cite this version:**

Thi Thanh Nguyen, Charles Soussen, Jérôme Idier, El-Hadi Djermoune. K-step analysis of orthogonal greedy algorithms for non-negative sparse representations. *Signal Processing*, 2021, 188, pp.108185. 10.1016/j.sigpro.2021.108185 . hal-01971697v5

HAL Id: hal-01971697

<https://hal.science/hal-01971697v5>

Submitted on 20 Jul 2021

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.

K -step analysis of orthogonal greedy algorithms for non-negative sparse representations

Thanh T. Nguyen^a, Charles Soussen^b, Jérôme Idier^c, El-Hadi Djermoune^a

^a *Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France*

^b *Université Paris-Saclay, CNRS, CentraleSupélec, L2S, F-91192 Gif-sur-Yvette, France*

^c *Laboratoire des Sciences du Numérique de Nantes, CNRS, F-44321 Nantes, France*

Abstract

This paper proposes an exact recovery analysis of greedy algorithms for non-negative sparse representations. Orthogonal greedy algorithms such as Orthogonal Matching Pursuit (OMP) and Orthogonal Least Squares (OLS) consist of gradually increasing the solution support and updating the nonzero coefficients in the least squares sense. From a theoretical viewpoint, greedy algorithms have been extensively studied in terms of exact support recovery. In contrast, the exact recovery analysis of their non-negative extensions (NNOMP, NNOLS) remains an open problem. We show that when the mutual coherence μ is lower than $\frac{1}{2K-1}$, the iterates of NNOMP / NNOLS coincide with those of OMP / OLS, respectively, the latter being known to reach K -step exact recovery. Our analysis heavily relies on a sign preservation property satisfied by OMP and OLS. This property is of stand-alone interest and constitutes our second important contribution. Finally, we provide an extended discussion of the main challenges of deriving improved analyses for correlated dictionaries.

Keywords: Greedy; Orthogonal Matching Pursuit; Orthogonal Least Squares; Non-Negativity; Mutual Coherence; Non-Negative Least Squares.

*Corresponding author: Charles Soussen.

Email addresses: thanh59hnue@gmail.com (Thanh T. Nguyen), charles.soussen@centralesupelec.fr (Charles Soussen), jerome.idier@ls2n.fr (Jérôme Idier), el-hadi.djermoune@univ-lorraine.fr (El-Hadi Djermoune)

1. Introduction

Greedy schemes are standard techniques for sparse approximation, with a relatively low computing cost compared to exact methods [4, 28], while convex relaxation yields another important branch of approximate methods [33, 14]. The principle of greedy schemes is to sequentially select atoms from a given dictionary in order to decrease the residual error. In *orthogonal* greedy algorithms, the decrease is maximal in the least square sense, and the approximated signal is computed as the orthogonal projection of the data vector onto the subspace spanned by the selected atoms. Orthogonal Matching Pursuit (OMP) [30] and Orthogonal Least Squares (OLS) [9] are two well-known instances of orthogonal greedy algorithms which only differ by their atom selection rule. For OMP, it simply consists in maximizing the magnitude of the inner product between the residual vector and the candidate atoms, assumed normalized. The OLS rule can be interpreted similarly, but the involved atoms are renormalized, projected versions of the candidate atoms [3, 32]. OLS is known under many other names, *e.g.*, ORMP [11] and OOMP [31]. Throughout this paper, we will use the generic acronym Oxx in statements that refer to both OMP and OLS.

In many applications such as multispectral unmixing [20], machine learning [23], mass spectroscopy [10] and fluid mechanics [1], the sought solution is required to be sparse and non-negative. Non-negative sparse reconstruction can be addressed using iterative thresholding algorithms (including NNSP, NNCoSaMP, and NNHTP) [21], but also using orthogonal greedy algorithms. The latter were naturally extended to the non-negativity setting [5, 38, 21], the main impact of the sign constraint is that Non-Negative Least-Square (NNLS) problems need to be solved to update the sparse approximation coefficients. This yields an increase of computation time since NNLS subproblems do not have closed-form solutions, and an iterative subroutine is needed. In [29], following the early work of [38], we proposed fully recursive implementations. We showed that non-negative greedy algorithms yield accurate empirical results and that their computation cost is of the same order of magnitude as those of Oxx

for moderate size problems.

The primary motivation of this paper is to elaborate mathematical conditions guaranteeing that the support of non-negative K -sparse representations is exactly recovered in K steps. While there is a rich literature on K -step recovery analysis with Oxx (based on *e.g.*, mutual incoherence [2, 6, 19] and restricted isometry assumptions [24, 36, 35, 22]) and other greedy algorithms, much less attention was paid to their non-negative versions. The existing analyses are scarce, and sharp worst-case exact recovery conditions are not even available. Our objective is to fill this gap in the literature and to derive sharp conditions for exact support recovery with non-negative versions of OMP and OLS.

Non-negative OMP was first introduced by Bruckstein *et al.* [5] under the name OMP, and then renamed NNOMP in [38] (see also [21, 29]). It relies on the repeated maximization of the *positive* inner product between the residual vector and the dictionary atoms, followed by the resolution of an NNLS problem. Existing analyses of NNOMP are rare and somewhat discordant. On the one hand, Bruckstein *et al.* claimed that the Mutual Incoherence Property (MIP) $\mu < \frac{1}{2K-1}$ holds for NNOMP [5] and that the proof should be similar to the one given in [34, 13] for OMP. Specifically, [5, Th. 3] states without proof that any K -sparse representation can be exactly recovered in K steps using NNOMP as long as $\mu < \frac{1}{2K-1}$. On the other hand, Kim *et al.* elaborated a unified MIP analysis of NNOMP and its generalized version in the multiple measurement vector setting [21, Th. 1]. In the specific case of NNOMP, *i.e.*, for single measurement vectors, the related MIP turns out to be very restrictive: $\mu < \frac{1}{K-1} - \frac{1}{2}$ is required, which can occur only when $K \in \{1, 2\}$.

Indeed, we think that it is impossible to prove [5, Th. 3] as a direct extension of [34, 13], 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 2.4 for further details). This is precisely why Kim *et al.*'s analysis, closely following Donoho *et al.*'s approach [13], yields over-pessimistic guarantees, instead of the expected result.

Our first contribution is to show that any non-negative K -sparse representation can be exactly recovered in K steps with NNOMP when the mutual coherence of the dictionary is less than $\frac{1}{2K-1}$. We further show that under the same condition, the non-negative extensions of OLS proposed in [37], named NNOLS and Suboptimal NNOLS (SNNOLS), are also guaranteed to recover the true support in K steps. To the best of our knowledge, the latter algorithms have never been analyzed. The analysis of NNOMP, NNOLS and SNNOLS is carried out in a unified way, and applies to noisy cases with bounded noise.

Our second contribution is to unveil a sign preservation property satisfied by Oxx for non-negative sparse representations. It is well-known that when $\mu < \frac{1}{2K-1}$, Oxx algorithms achieve K -step exact support recovery [34, 19]. We further show that at any iteration, the nonzero coefficients found by Oxx are positive. This property is of stand-alone interest, and turns out to be the cornerstone of our recovery analysis of non-negative extensions of Oxx. It enables us to prove that OMP and NNOMP coincide when $\mu < \frac{1}{2K-1}$, so [5, Th. 3] becomes a byproduct of our sign preservation analysis. Under the same conditions, we prove that OLS coincides with both NNOLS and SNNOLS [37].

The paper is organized as follows. Section 2 recalls known results about greedy algorithms and their non-negative versions. Section 3 contains our K -step analysis of non-negative greedy algorithms. The central sign preservation property of Oxx is stated as Theorem 3.1 and proved in the same section, most technical steps being postponed in Appendix. The numerical simulations of Section 4 illustrate the average behavior of algorithms outside the exact support recovery regime. In Section 5, an extensive discussion is provided on possible analyses for coherent dictionaries, and using other analysis techniques.

2. Notations and background

2.1. 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 so-called

K -sparse representation $\mathbf{y} = A\mathbf{x}$, in which the vector $\mathbf{x} \in \mathbb{R}^n$ has K non-zero elements. Without loss of generality, the atoms are assumed to be normalized, that is $\|\mathbf{a}_i\| = 1$, where $\|\cdot\|$ denotes the ℓ_2 norm. Notations † and ‡ 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 $\text{span}(A_S)^\perp$ its orthogonal complement, and by $P_S^\perp = I_m - A_S A_S^\ddagger$ the orthogonal projection onto $\text{span}(A_S)^\perp$, where I_m stands for the identity matrix of size m . We place the convention that $P_\emptyset^\perp = I_m$. When A_S is full column rank, one has a further explicit formulation $A_S^\ddagger = (A_S^\dagger A_S)^{-1} A_S^\dagger$. 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 $\tilde{\mathbf{b}}_i^S = \frac{\tilde{\mathbf{a}}_i^S}{\|\tilde{\mathbf{a}}_i^S\|}$ with the convention that $\tilde{\mathbf{b}}_i^S = \mathbf{0}$ when $\tilde{\mathbf{a}}_i^S = \mathbf{0}$. We will denote by \tilde{A}^S (resp., \tilde{B}^S) the matrix of size $m \times n$ gathering all projected atoms $\tilde{\mathbf{a}}_i^S$ (resp., $\tilde{\mathbf{b}}_i^S$).

The mutual coherence is defined by $\mu(A) = \max_{i \neq j} |\mathbf{a}_i^\dagger \mathbf{a}_j|$. This quantity tells us how much the dictionary atoms look alike. 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, $\mathbf{r}_0 = \mathbf{y}$ and $S_0 = \emptyset$. Whenever unambiguous, the simpler notations \mathbf{r} and S will be used.

2.2. OMP and OLS

OMP and OLS are heuristics aiming to minimize the approximation error $\|\mathbf{y} - A\mathbf{x}\|^2$ subject to $\|\mathbf{x}\|_0 \leq K$, where $\|\mathbf{x}\|_0$ counts the number of nonzero elements. Both algorithms start with the zero vector corresponding to the empty support. At each iteration, a new atom ℓ is added to the support. This process is repeated until K atoms are selected (see Algorithm 1). OMP and OLS share the same coefficient update and only differ by their selection rule:

$$\text{OMP: } \ell \in \arg \max_{i \notin S} |\mathbf{r}^\dagger \mathbf{a}_i|, \quad (1)$$

$$\text{OLS: } \ell \in \arg \max_{i \notin S} |\mathbf{r}^\dagger \tilde{\mathbf{b}}_i^S| \quad (2)$$

Algorithm 1: Oxx in K steps.

Input: \mathbf{y} , A , K

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 $S \leftarrow S \cup \{\ell\}$, with atom ℓ chosen according to (1) or (2);

5 $\hat{\mathbf{x}}_S \leftarrow A_S^\dagger \mathbf{y}$;

6 **until** $\text{card}(S) = K$;

where $\mathbf{r} = P_S^\perp \mathbf{y}$ denotes the current residual. (1) can also be written as:

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

since \mathbf{r} is orthogonal to $\text{span}(A_S)$. These inner product expressions come from the geometrical interpretation of OMP and OLS [3]. 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} [3]. So, it is clear that the OLS selection rule is the most costly, although recursive update schemes are available [9].

In [34] and [19], it was shown that when $\mu(A) < \frac{1}{2K-1}$, any K -sparse representation $\mathbf{y} = A\mathbf{x}^*$ can be exactly recovered by OMP and OLS, respectively, in K iterations. These analyses were then generalized to noisy settings. Let us recall the results in [6, 18] in the case of a bounded noise satisfying

$$\|\mathbf{n}\| < \frac{1 - (2K-1)\mu(A)}{2} \min\{|x_i^*|, x_i^* \neq 0\}. \quad (4)$$

Lemma 2.1. [6, Th. 1], [18, Th. 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 (4). Then Oxx recover the support of \mathbf{x}^* in K iterations.

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

Input: \mathbf{y} , A , K

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 Choose atom ℓ according to prescribed selection rule (SR) in Table 1;

4 $S \leftarrow S \cup \{\ell\}$;

5 $\hat{\mathbf{x}}_S \leftarrow \arg \min_{\mathbf{z} \geq \mathbf{0}} \|\mathbf{y} - A_S \mathbf{z}\|^2$;

6 **until** $\text{card}(S) = K$ or prescribed stopping condition (SC) in Table 1;

	(SR)	(SC)
NNOMP [5]	$\max_{i \notin S} \mathbf{r}^t \mathbf{a}_i$	$\max_{i \notin S} \mathbf{r}^t \mathbf{a}_i \leq 0$
SNNOLS [37]	$\max_{i \notin S} \mathbf{r}^t \tilde{\mathbf{b}}_i^S$	$\max_{i \notin S} \mathbf{r}^t \tilde{\mathbf{b}}_i^S \leq 0$
NNOLS [37]	$\min_{i \notin S} \min_{\mathbf{z} \geq \mathbf{0}} \ \mathbf{y} - A_{S \cup \{i\}} \mathbf{z}\ ^2$	$\ \mathbf{r}\ ^2 = \min_{i \notin S} \min_{\mathbf{z} \geq \mathbf{0}} \ \mathbf{y} - A_{S \cup \{i\}} \mathbf{z}\ ^2$

Table 1: Possible selection rules (SR) and stopping conditions (SC) of non-negative Oxx. The current residual is computed as $\mathbf{r} = \mathbf{y} - A_S \hat{\mathbf{x}}_S$.

2.3. Non-negative setting

The main non-negative extensions of OMP and OLS are NNOMP [38, 5], NNOLS and SNNOLS [37]. All three solve NNLS subproblems to update the selected coefficients according to Algorithm 2, while they differ at the atom selection stage, see Table 1. 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. NNOLS is built upon the dual formulation of OLS in the optimization viewpoint: NNOLS solves as many NNLS subproblems as candidate atoms to find the lowest residual error. Efficient implementation schemes can be found in [29] using a fully recursive implementation, including huge possibilities of accelerations for NNOLS.

2.4. Exact recovery analysis in the non-negative setting

2.4.1. Extension of K -step exact recovery analysis of OMP

Bruckstein *et al.* [5] claimed that the K -step exact recovery analysis of NNOMP can be carried out as a straightforward extension of the classical analyses of OMP [34, 13]. We argue that this extension is a more complex matter.

Tropp's reasoning [34] consists of minorizing the modulus $|\mathbf{r}^\dagger \mathbf{a}_i|$ for correct dictionary atoms \mathbf{a}_i . Unfortunately, similar bounds cannot be obtained when the modulus is dropped. Donoho *et al.*'s recursive proof [13] exploits that for sparse inputs, the residual lays in the subspace spanned by the true atoms when true atoms have been selected at previous iterations. Therefore, the reasoning at the first iteration can be repeated in the next iterations. To generalize this proof, one would need that for non-negative representations (*i.e.*, for inputs laying in the positive span of the true atoms), the residual lays in the same positive span. This conjecture turns out to be false. Indeed, for noiseless inputs $\mathbf{y} = A\mathbf{x}^*$, if true atoms are selected until iteration k , the residual reads $\mathbf{r} = A(\mathbf{x}^* - \hat{\mathbf{x}}^{(k)})$, where \mathbf{x}^* and $\hat{\mathbf{x}}^{(k)}$ respectively denote the true solution and the NNOMP iterate. So, \mathbf{r} lays in the positive span of true atoms if and only if $\hat{\mathbf{x}}^{(k)} \leq \mathbf{x}^*$. Simple numerical tests show that the latter condition may not hold when $\mathbf{x}^* \geq \mathbf{0}$ and $\mu(A) < \frac{1}{2K-1}$: see § 4.2 and Fig. 3(a). Consequently, Donoho *et al.*'s reasoning at the first iteration cannot be fruitfully repeated at the subsequent ones.

2.4.2. ℓ_1 analysis with non-negativity constraints

The 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 [17] proved that when $\mathbf{y} = A\mathbf{x}^*$ with $\mathbf{x}^* \geq \mathbf{0}$, BP and BPDN (for sufficiently small λ) both have a unique solution under the MIP assumption, which identifies with \mathbf{x}^* as long as A_{S^*} is full rank and $\forall j \notin S^*, |\mathbf{1}^\dagger (A_{S^*}^\dagger \mathbf{a}_j)| < 1$, where $S^* := \text{supp}(\mathbf{x}^*)$ and $\mathbf{1}$ denotes the all-ones vector of size $\text{card}(S^*)$. The latter condition does not depend on the magnitudes of

coefficients x_i^* , therefore it is a uniform exact recovery condition for BP/BPDN for any $\mathbf{x}^* \geq \mathbf{0}$ supported by S^* . Since the cost function $\|\mathbf{y} - A\mathbf{x}\|^2 + \lambda\|\mathbf{x}\|_1$ is minimum for $\mathbf{x} = \mathbf{x}^* \geq \mathbf{0}$, it follows that \mathbf{x}^* is also the unique solution of the so-called non-negative Garrote [25] $\min_{\mathbf{x} \geq \mathbf{0}} \|\mathbf{y} - A\mathbf{x}\|^2 + \lambda\|\mathbf{x}\|_1$.

2.4.3. Extension of K -step exact recovery analysis of ℓ_1 homotopy

Homotopy is a popular greedy algorithm dedicated to BPDN for a continuum of λ 's [15]. The principle is to reconstruct the regularization path (defined as the set of solutions for all λ) for gradually decreasing λ 's starting from $+\infty$. Homotopy has a stepwise mechanism akin to OMP with an atom selection or deselection at each iteration. When $\mu(A) < \frac{1}{2K-1}$, K -step recovery is guaranteed for any K -sparse representation $\mathbf{y} = A\mathbf{x}^*$ [14]. This analysis is based on two ingredients: the correct selection of atoms indexed by $\text{supp}(\mathbf{x}^*)$ and the so-called *sign agreement* property¹, ensuring that no deselection occurs. Donoho and Tsai further 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 after K iterations, their sign is preserved throughout the iterations. Consequently, the non-negative extension (NLARS) of ℓ_1 homotopy proposed in [27] yields the same iterates as homotopy for all K -sparse non-negative representations, and NLARS achieves uniform K -step recovery when $\mu(A) < \frac{1}{2K-1}$. Although OMP shares structural similarities with ℓ_1 homotopy, the analysis of NNOMP turns out to be far more difficult than that of NLARS. In particular, the sign agreement property of [14] (which essentially depends on the sign of the inner products between the current residual and the active atoms) does not make sense anymore in the context of OMP, since the latter inner products vanish. We will devise a brandnew sign preservation property, that will be proved valid for both OMP and OLS.

¹Sign agreement is defined as follows: at every iteration of the homotopy algorithm, the iterate $\hat{\mathbf{x}}$ satisfies $\text{sign}(\hat{\mathbf{x}}_S) = \text{sign}(A_S^T \mathbf{r})$, where $S = \text{supp}(\hat{\mathbf{x}})$.

3. Exact recovery and sign preservation

This section contains our main results concerning the exact recovery analysis of non-negative greedy algorithms under the MIP assumption. The cornerstone of our study is Theorem 3.1, the subsequent exact recovery results being direct consequences. In Subsection 3.2, the proof of Theorem 3.1 is decomposed into distinct steps, most technical elements being postponed in Appendix.

3.1. Main results

Theorem 3.1 (sign preservation). *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 (4). Then, Oxx recovers the support of \mathbf{x}^* in K iterations, and at each iteration, the estimated coefficients of selected atoms are positive.*

Theorem 3.1 has interesting implications concerning non-negative versions of Oxx. Let us start with the following lemma.

Lemma 3.1. *Let $\mathbf{y} \in \mathbb{R}^m$ and assume that any set of K dictionary columns is linearly independent.*

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

This result is intuitive since the nonzero coefficients are estimated in the least squares sense. Clearly, if an unconstrained least squares solution is positive, then it is also the solution of the corresponding NNLS problem.

Proof. See Appendix 6.3. □

From Theorem 3.1 and Lemma 3.1, we can deduce the following result.

Corollary 3.1. *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 (4). 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 [7, 19] by Oxx. Cai *et al.* [7] 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 disjoint 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 3.2. *Condition $\mu(A) < \frac{1}{2K-1}$ is necessary for uniform non-negative K -step recovery: 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 disjoint 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 [7, Sect. III], the latter having two K -sparse representations $\mathbf{y} = A\mathbf{x} = A\mathbf{z}$ with disjoint supports. Since both supports are disjoint, one can define the subrogate dictionary $A' \in \mathbb{R}^{m \times n}$ as $\mathbf{a}'_i = \text{sign}(x_i) \mathbf{a}_i$ if $i \in \text{supp}(\mathbf{x})$, $\mathbf{a}'_i = \text{sign}(z_i) \mathbf{a}_i$ if $i \in \text{supp}(\mathbf{z})$, and $\mathbf{a}'_i = \mathbf{a}_i$ otherwise.

Moreover, let $|\mathbf{x}|$ denote the vector whose entries are equal to $|x_i|$, and define $|\mathbf{z}|$ similarly. Obviously, $\mathbf{y} = A'|\mathbf{x}| = A'|\mathbf{z}|$, and $\mu(A') = \mu(A) = \frac{1}{2K-1}$. \square

3.2. Proof of Theorem 3.1

Let $\mathbf{y} = A\mathbf{x}^* + \mathbf{n}$ with $\mathbf{x}^* \geq \mathbf{0}$ and a noise term satisfying (4). According to Lemma 2.1, Oxx recovers the support S^* of \mathbf{x}^* in K iterations. For any $k \leq K$, the support S_k of the current solution $\hat{\mathbf{x}}^{(k)}$ is therefore a subset of S^* . Recall that $\hat{\mathbf{x}}_{S_k}^{(k)}$ is the unconstrained least squares solution related to S_k , see Algorithm 1. Let $\mathbf{r}_k = P_{S_k}^\perp \mathbf{y}$ denote the related residual, with $\mathbf{r}_0 = \mathbf{y}$.

We proceed in two steps to show that $\hat{\mathbf{x}}_{S_k}^{(k)} > \mathbf{0}$ for all $k \in \{1, \dots, K\}$. First, we prove that the coefficient of each newly selected atom $\hat{x}_{S_k \setminus S_{k-1}}^{(k)}$ is positive. Then, we show that the updated coefficients $\hat{\mathbf{x}}_{S_{k-1}}^{(k)}$ remain positive. Let us first characterize the coefficients $\hat{\mathbf{x}}_{S_k \setminus S_j}^{(k)}$, $j < k$ being an arbitrary iteration.

Lemma 3.2. *Let $\mathbf{y} \in \mathbb{R}^m$, and let j and k be two iteration indices with $0 \leq j < k$. Assume that A_{S_k} is full column rank. Then, the k -th iterate of Oxx satisfies*

$$\hat{\mathbf{x}}_{S_k \setminus S_j}^{(k)} = (\tilde{A}_{S_k \setminus S_j}^{S_j})^\dagger \mathbf{r}_j. \quad (5)$$

Proof. See Appendix 6.2. □

The two steps of the proof of Theorem 3.1 are stated now.

Lemma 3.3 (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 (4). For all $k \in \{1, \dots, K\}$, $\hat{x}_{S_k \setminus S_{k-1}}^{(k)} > 0$.*

Proof. See Appendix 6.4. □

Lemma 3.4 (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 (4). For all $k \in \{2, \dots, K\}$, $\hat{\mathbf{x}}_{S_{k-1}}^{(k)} > \mathbf{0}$.*

Proof. See Appendix 6.5. □

Notice that Lemmas 3.3 and 3.4 are straightforward results at the last iteration $k = K$. Indeed, the correct support S^* is found by Oxx according to Lemma 2.1. Thus, Oxx algorithms yield the output $\hat{\mathbf{x}}^{(K)}$ supported by S^* , with $\hat{\mathbf{x}}_{S^*}^{(K)} = A_{S^*}^\dagger \mathbf{y}$. According to the local stability analysis in [13, Th. 5.1],

$$\|\mathbf{x}^* - \hat{\mathbf{x}}^{(K)}\|^2 \leq \frac{\|\mathbf{n}\|^2}{1 - (K-1)\mu(A)}. \quad (6)$$

Combining (6) with (4), one can check that $\|\mathbf{x}^* - \hat{\mathbf{x}}^{(K)}\| < \min\{x_i^*, i \in S^*\}$. Hence, for all $i \in S^*$, $|x_i^* - \hat{x}_i^{(K)}| \leq x_i^*$, which proves that the entries of $\hat{\mathbf{x}}^{(K)}$ are non-negative since $\mathbf{x}^* \geq \mathbf{0}$.

4. Numerical study

4.1. Comparison of Oax and their non-negative versions

The previous section showed that in some specifically favorable situations, greedy algorithms such as OMP not only recover the support of the true solution, but also yield sparse representations with non-negative weights. In such conditions, according to Lemma 3.1, implementing non-negative versions of greedy algorithms is useless. On the contrary, one can empirically observe that non-negative greedy algorithms reach superior performance for coherent dictionaries and for noisy scenarios [29], which is in agreement with the fact that they exploit more information than usual greedy algorithms. The next two numerical simulations further illustrate this fact.

4.1.1. Deconvolution problem

Let us consider a dictionary $A \in \mathbb{R}^{50 \times 22}$ corresponding to a convolution operation $y(t) = (h * x)(t)$ with a Gaussian-shaped impulse response h . The dictionary atoms therefore identify with delayed and discretized versions of $h(t)$. Specifically, 22 atoms are considered, corresponding to regularly spaced integer delays. The width of the Gaussian impulse response is $\sigma = 0.5$, yielding a mutual coherence $\mu(A) = 0.37$. We randomly choose $K = 10$ atoms in A , whose locations in the dictionary are drawn with a uniform distribution. The atoms are equally weighted with $x_i^* = 1$ for all $i \in S^*$, and we generate data $\mathbf{y} = A\mathbf{x}^* + \mathbf{n}$ with additive Gaussian noise \mathbf{n} . Note that $\mu(A) \approx \frac{7}{2K-1}$, so exact support recovery is not guaranteed even at low noise. Within a certain range of signal-to-noise ratio (SNR, defined by $10 \log_{10}(\|A\mathbf{x}^*\|^2 / \|\mathbf{n}\|^2)$), generated average performance have been measured for OMP, NNOMP, OLS, SNNOLS and NNOLS in terms of K -step support recovery rates. 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 [5, 21]. For high SNRs, greedy algorithms keep some exact recovery capacities far beyond the zero defect area $\mu \in [0, \frac{1}{2K-1}]$. In the low

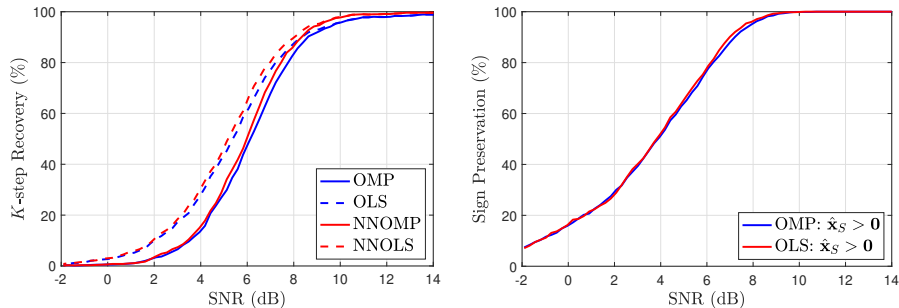


Figure 1: Rates of trials in which K -step exact support recovery (left) and sign preservation (right) is achieved, for a simulated data experiment involving a convolutive dictionary with Gaussian-shaped atoms. The sign preservation rate refers to the trials in which for all iteration $k \leq K$, $\hat{\mathbf{x}}^{(k)} \geq \mathbf{0}$. SNNOLS (not shown here) yields the exact same curve as NNOLS.

SNR regime, the exact recovery capacity almost surely vanishes. For intermediate SNRs, one can notice a significant difference of performance between the usual greedy algorithms and their non-negative extensions. We also performed a sign-preservation test, checking that at all iterations of Oxx, the estimated weights are all positive, regardless of whether K -step recovery is achieved or not. Fig. 1 shows that this sign-preservation property is rather robust. Specifically, sign-preservation occurs whenever the support is exactly recovered by Oxx. In alternative cases, the non-negativity constraint is truly effective within NNOMP/NNOLS, which makes it possible to achieve higher K -step recovery rates as compared to Oxx. Note that in the simulation of Fig. 1, the SNR is lower than 15 dB, so greedy algorithms mainly capture the main broad signal features embedded in noise. Therefore, OMP and OLS show similar performance. Their performance can be more clearly distinguished for problems involving correlated dictionaries and for larger values of SNR and K , see, *e.g.*, [29].

4.1.2. Random dictionaries

The goal of this simulation is to assess the performance of algorithms with respect to the size (m, n) of the dictionary and the subset cardinality K . Here, the dictionary entries are drawn independently using a standard Gaussian distribution. We choose $A \in \mathbb{R}^{m \times n}$ with fixed $n = 500$ and with m ranging from

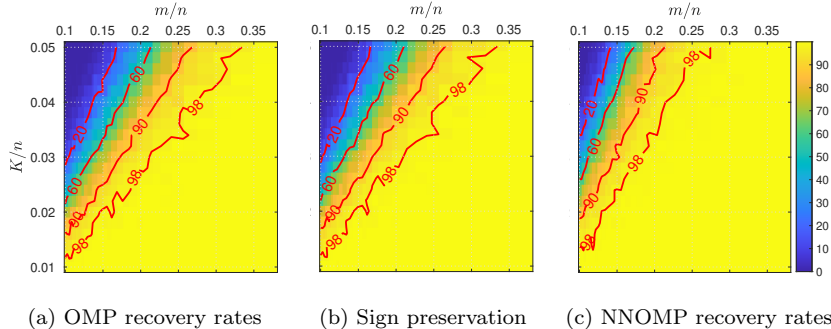


Figure 2: K -step recovery and sign preservation evaluation for Gaussian random dictionaries of variable size (m/n) and for various sparsity levels (K/n). (a,c) For both OMP and NNOMP, empirical phase transition curves (in red) are the contour plots corresponding to exact recovery rates of 20, 60, 90 and 98%, respectively. (b) Similar contour plots are computed for the OMP sign preservation rates.

50 to 190. Moreover, for a given size (m, n), the numerical tests of § 4.1.1 (evaluation of K -step recovery and sign preservation rates) are repeated for various support cardinalities K . For a given setting (m, n, K), the support of \mathbf{x}^* is randomly chosen, and the coefficients x_i^* are distributed according to the uniform distribution in $(0, 1)$, yielding K -sparse representations $\mathbf{y} = A\mathbf{x}^*$. The performance of OMP and NNOMP, averaged over 300 trials, are displayed in Fig. 2 where the horizontal and vertical axes respectively refer to the dictionary size (m/n) and sparsity level (K/n). As one would expect, the exact recovery rates increase with the number of measurements (high values of m/n) and for highly sparse representations (low values of K/n). The OMP and NNOMP performance are further compared by means of phase transition curves, defined as contour plots of Figs. 2(a,c). We observe similar conclusions as in § 4.1.1: the phase transition curves of NNOMP are shifted to the left as compared to those of OMP, which highlights the benefit of using non-negative versions. Moreover, the OMP sign preservation rates (Fig. 2(b)) are very similar to the OMP recovery rates (Fig. 2(a)). This result is consistent with that of Fig. 1 for high SNRs.

The experiment of Fig. 2 was repeated for OLS and SNNOLS, and OLS and

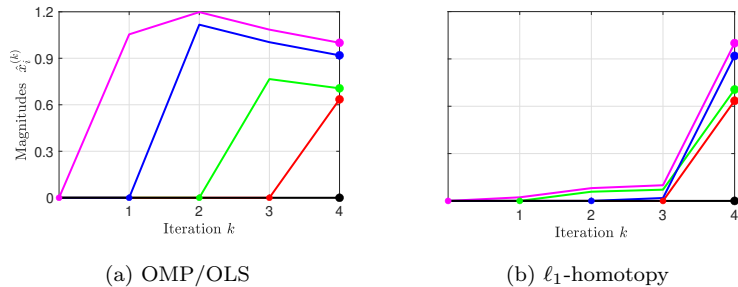


Figure 3: Behavior of Oxx and ℓ_1 -homotopy for a toy problem ($m = n = 5$) corresponding to a 4-sparse noiseless representation. (a) The OMP and OLS iterates are identical and yield magnitudes $\hat{x}_i^{(k)}$ with non-monotonous variations. (b) On the contrary, the magnitudes of the ℓ_1 -homotopy iterates are always increasing. The ground truth magnitudes x_i^* are represented with bullets. They are exactly recovered after $K = 4$ iterations ($\hat{\mathbf{x}}^{(K)} = \mathbf{x}^*$).

NNOLS. The related results are very similar to those obtained with OMP and NNOMP, with slight increases of the recovery rates of Fig. 1(a,c).

4.2. Non-monotony of the magnitude variations

As argued in § 2.4, ℓ_1 homotopy is a stepwise greedy algorithm for which sign preservation holds whenever $\mu(A) < \frac{1}{2K-1}$. In [14], 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, we observe that this monotony property does not hold for Oxx algorithms, since the magnitudes may either increase or decrease during iterations. We indeed compare ℓ_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 of A satisfy $\mathbf{a}_i^\dagger \mathbf{a}_j = \pm\mu(A)$ for $i \neq j$, 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. 3, 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. ℓ_1 -homotopy yields magnitudes that are increasing with k , which

is consistent with the theoretical result in [14]. On the contrary, the OMP and OLS iterates (which are identical here; the same indices are selected at each iteration) are non-monotonous.

5. Discussion and conclusion

5.1. Contributions and links with alternative algorithms

Our analysis of non-negative greedy algorithms essentially relies on the discovery of the sign preservation property for Oxx algorithms. Indeed, we could show that the Oxx algorithms yield the same iterates as their non-negative extensions under the MIP condition. Moreover, the latter condition is not only a sufficient but also a (worst-case) necessary condition of exact recovery. A strong feature of our analysis is that NNOMP, NNOLS and SNNOLS are analyzed in a unified way. In contrast, many OMP and OLS analyses in the literature are done separately, because the OLS selection rule is more sophisticated than that of OMP (see the next paragraph for details). One may wonder whether the proposed analysis applies to other approaches and algorithms. In Section 4.2, we pointed out that ℓ_1 algorithms enjoy a sign preservation property, hence their non-negative extensions achieve K -step recovery when $\mu(A) < \frac{1}{2K-1}$. The analysis of the non-negative versions of CoSaMP and Subspace Pursuit (SP) appears to be less straightforward. The so-called NN-CoSaMP and NN-SP algorithms proposed in [21], aim to update an estimated support of constant size K by repeatedly performing support merging, NNLS estimation of coefficients and support pruning. To our knowledge, no thorough exact recovery analysis of these algorithms has been carried out in the literature. Since the classical analyses of CoSaMP and SP are substantially different from those of greedy algorithms (the former 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, thus the line of reasoning may be substantially different from the one elaborated here.

5.2. *Towards analyses for dictionaries of higher coherence*

The experiments of Section 4 confirm that non-negative greedy algorithms remain accurate when $\mu(A)$ is far greater than $\frac{1}{2K-1}$. Similar observations were made in [5, 21]. In [29], an extensive set of numerical tests were carried out for highly coherent dictionaries and/or noisy observations. It turns out that Oxx algorithms do yield iterates with negative entries (the sign preservation property is violated), and that their non-negative versions are worth being considered. 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 [29].

5.2.1. *ERC based analyses*

With a view to further distinguish the influence of true and wrong atoms and to relax the conditions of Theorem 3.1, Tropp’s exact recovery condition (ERC) [34] 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 [34] and OLS [32]. It is therefore weaker than the MIP and Restricted Isometry Property (RIP) conditions which hold for all K -sparse representations irrespective of their supports. To the best of our knowledge, the fact that K -step support recovery is achieved with non-negative versions of Oxx under the ERC is an open question. Hereafter, we highlight that the sign preservation property is not guaranteed under the ERC.

Proposition 5.1. *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. The following counterexample exploits that the ERC remains true when wrong atoms are orthogonal to the span of the true atoms, regardless of the correlation between the true atoms. Let us consider a normalized dictionary

$A \in \mathbb{R}^{(K+1) \times (K+1)}$ such that $\mathbf{a}_i^t \mathbf{a}_j = -\frac{1}{K}$ for $i \neq j$ and $i, j \leq K$, and $\mathbf{a}_{K+1}^t \mathbf{a}_j = 0$ for all $j \leq K$, *i.e.*, the last atom is orthogonal to the others.

Let us further define $\mathbf{y} = A\mathbf{x}^*$ with $\mathbf{x}^* = [a, \dots, a, 1, 0]^t$ and $a > \frac{K-1}{K-3}$. The related true support reads $S^* = \{1, \dots, K\}$. 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. Therefore, S^* is exactly recovered in K steps by Oxx. However, the iterates of Oxx and their nonnegative versions differ from the very first iteration. Indeed, we have $A^t \mathbf{y} = (A^t A)\mathbf{x}^* = [b, \dots, b, c, 0]^t$ with $b = \frac{2a-1}{K} > 0$ and $c = 1 - \frac{(K-1)a}{K} < 0$. 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. \square

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 proposed here, since sign-preservation with Oxx is not guaranteed anymore.

5.2.2. RIP based analyses

In recent years, various K -step recovery analyses of Oxx algorithms have been proposed under the restricted isometry property. OMP was analyzed first. Improved conditions have been gradually proposed on the RIP factor or order $K+1$, denoted by δ_{K+1} [12, 26, 8, 36, 22]. The OLS algorithm was analyzed more recently with increasingly sharp bounds [35, 22]. For both OMP and OLS, it was shown in [22] 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 3.3 cannot be easily adapted since it strongly relies on the derivation of lower bounds of signed inner products such as $\mathbf{r}^t \tilde{\mathbf{a}}_j$. In contrast, classical RIP analyses [22] consist of deriving (i) an accurate lower bound of the *modulus* $|\mathbf{r}^t \tilde{\mathbf{a}}_j|$ for true atoms \mathbf{a}_j , and (ii) an upper bound of $|\mathbf{r}^t \tilde{\mathbf{a}}_j|$ for wrong atoms, in order to show that the maximum of $|\mathbf{r}^t \tilde{\mathbf{a}}_j|$ is reached for a true atom. Regarding step (i), the most accurate techniques make use of the inequality $\|\mathbf{t}\|_\infty \geq \|\mathbf{t}\|/\sqrt{\|\mathbf{t}\|_0}$ to minorize $\max\{|\mathbf{r}^t \tilde{\mathbf{a}}_j|, j \in S^* \setminus S_{k-1}\} = \|\tilde{A}_{S^* \setminus S_{k-1}}^t \mathbf{r}\|_\infty$, see, e.g., [24]. Unfortunately, the latter techniques cannot be used to minorize the *signed* inner product $\max\{\mathbf{r}^t \tilde{\mathbf{a}}_j, j \in S^* \setminus S_{k-1}\}$. Therefore, it seems that brandnew analysis techniques need to be elaborated in the non-negative setting.

Acknowledgment

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

6. Appendix

6.1. Useful lemmas

Let us recall useful lemmas. Lemma 6.1 provides an upper bound on the ℓ_1 -norm of the columns of the inverse Gram matrix by means of mutual coherence. Lemma 6.2 bounds the inner product between projected atoms. Lemma 6.3 is related to the full rankness of the matrices formed of projected atoms. Lemma 6.4 is a simple algebraic manipulation related to the pseudo-inverse.

Lemma 6.1. [16, Th. 5.3], [34, Th. 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*

$$\|(B^t B)^{-1}\|_{1,1} \leq \frac{1}{1 - (k-1)\mu(B)} \quad (7)$$

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

Lemma 6.2. [18, Lemma 1]. If $\mu(A) \leq \frac{1}{k+1}$ with $k = \text{card}(S)$ then

$$\forall i \notin S, \|\tilde{\mathbf{a}}_i^S\|^2 \geq \beta_k, \quad (8)$$

$$\forall p \neq q, |(\tilde{\mathbf{a}}_p^S)^\dagger \tilde{\mathbf{a}}_q^S| \leq \mu_k \beta_k, \quad (9)$$

where

$$\beta_k = \frac{(1 + \mu(A))(1 - k\mu(A))}{1 - (k-1)\mu(A)} \quad \text{and} \quad \mu_k = \frac{\mu(A)}{1 - k\mu(A)}. \quad (10)$$

Lemma 6.3. [32, 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 6.4. Let $A = [\mathbf{a}_1, A_2] \in \mathbb{R}^{p \times q}$ be a full column rank matrix, where $\mathbf{a}_1 \in \mathbb{R}^p$ and $A_2 \in \mathbb{R}^{p \times (q-1)}$ is formed of the last $q-1$ columns of A . Then,

$$\forall \mathbf{r} \in \mathbb{R}^p, (A^\dagger \mathbf{r})_1 = \frac{\mathbf{r}^\dagger (P_{S_2}^\perp \mathbf{a}_1)}{\|P_{S_2}^\perp \mathbf{a}_1\|^2} \quad (11)$$

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

Proof. Any $\mathbf{r} \in \mathbb{R}^p$ can be uniquely decomposed as $\mathbf{r} = \mathbf{p}_A + \mathbf{p}_{A^\perp}$ where \mathbf{p}_A and \mathbf{p}_{A^\perp} are the orthogonal projections of \mathbf{r} onto $\text{span}(A)$ and $\text{span}(A)^\perp$. Moreover, the vector $A^\dagger \mathbf{r}$ gathers the weights of the decomposition of \mathbf{p}_A in the column span of A . Specifically, we have $\mathbf{p}_A = (A^\dagger \mathbf{r})_1 \mathbf{a}_1 + A_2 \boldsymbol{\alpha}$ with $\boldsymbol{\alpha} \in \mathbb{R}^{q-1}$. Now, we make use of the orthogonal decomposition $\mathbf{a}_1 = P_{S_2}^\perp \mathbf{a}_1 + A_2 \boldsymbol{\beta}$, with $\boldsymbol{\beta} \in \mathbb{R}^{q-1}$ to derive the following orthogonal decomposition of \mathbf{r} :

$$\mathbf{r} = (A^\dagger \mathbf{r})_1 P_{S_2}^\perp \mathbf{a}_1 + A_2 (\boldsymbol{\alpha} + (A^\dagger \mathbf{r})_1 \boldsymbol{\beta}) + \mathbf{p}_{A^\perp}. \quad (12)$$

(11) is obtained directly from (12) by calculating the inner product $\mathbf{r}^\dagger (P_{S_2}^\perp \mathbf{a}_1)$. The denominator in (11) is nonzero because A is the full rank. \square

Let us now start with the proof of Lemma 3.2, since this lemma will be used later in the proofs of Lemmas 3.1, 3.3, and 3.4.

6.2. Proof of Lemma 3.2

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

$$\hat{\mathbf{x}}_{S_k}^{(k)} = \arg \min_{\mathbf{z} \in \mathbb{R}^k} \|\mathbf{y} - A_{S_k} \mathbf{z}\|^2 = \arg \min_{\mathbf{v}, \mathbf{w}} \|\mathbf{y} - A_{S_j} \mathbf{v} - A_{S_k \setminus S_j} \mathbf{w}\|^2. \quad (13)$$

In addition, the minimum squared error reads

$$\begin{aligned}
\min_{\mathbf{v}, \mathbf{w}} \|\mathbf{y} - A_{S_j} \mathbf{v} - A_{S_k \setminus S_j} \mathbf{w}\|^2 &= \min_{\mathbf{w}} \left(\min_{\mathbf{v}} \|\mathbf{y} - A_{S_k \setminus S_j} \mathbf{w} - A_{S_j} \mathbf{v}\|^2 \right) \\
&= \min_{\mathbf{w}} \|P_{S_j}^\perp (\mathbf{y} - A_{S_k \setminus S_j} \mathbf{w})\|^2 \\
&= \min_{\mathbf{w}} \|\mathbf{r}_j - \tilde{A}_{S_k \setminus S_j}^{S_j} \mathbf{w}\|^2. \tag{14}
\end{aligned}$$

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 6.3. The minimum corresponding to (14) is reached for $\mathbf{w} = (\tilde{A}_{S_k \setminus S_j}^{S_j})^\dagger \mathbf{r}_j$ which identifies with $\hat{\mathbf{x}}_{S_k \setminus S_j}^{(k)}$ according to (13).

6.3. Proof of Lemma 3.1

We prove by induction that the supports found by OMP and NNOMP (resp., by OLS and NNOLS/SNNOLS) coincide. This is sufficient to prove the claim, since when the supports coincide, the estimated coefficients coincide as well. Indeed, denoting $\hat{\mathbf{x}}^{(k)}$ the coefficients yielded by Oxx, $\hat{\mathbf{x}}^{(k)}$ reads as the unconstrained least-square solution related to S_k . By assumption, $\hat{\mathbf{x}}^{(k)} \geq 0$, thus

$$\hat{\mathbf{x}}_{S_k}^{(k)} = \arg \min_{\mathbf{z}} \|\mathbf{y} - A_{S_k} \mathbf{z}\|^2 = \arg \min_{\mathbf{z} \geq \mathbf{0}} \|\mathbf{y} - A_{S_k} \mathbf{z}\|^2, \tag{15}$$

so $\hat{\mathbf{x}}^{(k)}$ is also the NNLS solution related to S_k .

6.3.1. NNOMP vs OMP

The first atom selected by OMP is indexed by $\{\ell\} = \arg \max_i |\mathbf{y}^\dagger \mathbf{a}_i|$. The related 1-sparse approximation reads $\mathbf{y} \approx \hat{x}_\ell^{(1)} \mathbf{a}_\ell$ with $\hat{x}_\ell^{(1)} = \mathbf{y}^\dagger \mathbf{a}_\ell$. By assumption, $\hat{x}_\ell^{(1)} > 0$. It follows that $\{\ell\} = \arg \max_i \{\mathbf{y}^\dagger \mathbf{a}_i\}$, so ℓ is also selected at the first iteration of NNOMP with input \mathbf{y} .

The same argument is repeated at iteration k . Assume OMP and NNOMP deliver the same iterate after $k-1$ iterations, and denote by \mathbf{r}_{k-1} the (identical) residual vector. Let $S_k = S_{k-1} \cup \{\ell\}$ and $\hat{\mathbf{x}}^{(k)}$ denote the support and iterate found by OMP at iteration k . Applying Lemma 3.2 with $j \leftarrow k-1$, we get $\hat{x}_\ell^{(k)} = (\tilde{\mathbf{a}}_\ell^{S_{k-1}})^\dagger \mathbf{r}_{k-1}$ where $(\tilde{\mathbf{a}}_\ell^{S_{k-1}})^\dagger = (\tilde{\mathbf{a}}_\ell^{S_{k-1}})^t / \|\tilde{\mathbf{a}}_\ell^{S_{k-1}}\|^2$ and $\tilde{\mathbf{a}}_\ell^{S_{k-1}} \neq \mathbf{0}$

according to Lemma 6.3 and the full rankness of A_{S_k} . It follows that

$$\hat{x}_\ell^{(k)} = \frac{\mathbf{r}_{k-1}^t \tilde{\mathbf{a}}_\ell^{S_{k-1}}}{\|\tilde{\mathbf{a}}_\ell^{S_{k-1}}\|^2} = \frac{\mathbf{r}_{k-1}^t \mathbf{a}_\ell}{\|\tilde{\mathbf{a}}_\ell^{S_{k-1}}\|^2} \quad (16)$$

where the last equality comes from the fact that $\tilde{\mathbf{a}}_\ell^{S_{k-1}} = P_{S_{k-1}}^\perp \mathbf{a}_\ell$ and $\mathbf{r}_{k-1} \in \text{span}(A_{S_{k-1}})^\perp$. By assumption, $\hat{\mathbf{x}}^{(k)} \geq \mathbf{0}$. (16) implies that $\mathbf{r}_{k-1}^t \mathbf{a}_\ell \geq 0$. Since ℓ is the unique solution to (1), we also have $\{\ell\} = \arg \max_{i \notin S_{k-1}} \{\mathbf{r}_{k-1}^t \mathbf{a}_i\}$. So, ℓ is also selected at the k -th iteration of NNOMP with input \mathbf{y} .

6.3.2. SNNOLS vs OLS

The very first iterates of SNNOLS and OLS respectively identify with those of NNOMP and OMP. They coincide according to § 6.3.1. At iteration k , the previous proof can be repeated, where $\hat{\mathbf{x}}^{(k)}$ now denotes the OLS iterate. Introducing the normalized projected atoms in (16), we get $\hat{x}_\ell^{(k)} = (\mathbf{r}_{k-1}^t \tilde{\mathbf{b}}_\ell^{S_{k-1}}) / \|\tilde{\mathbf{a}}_\ell^{S_{k-1}}\|$, and $\mathbf{r}_{k-1}^t \tilde{\mathbf{b}}_\ell^{S_{k-1}} \geq 0$ since $x_\ell^{(k)} \geq 0$. Since ℓ is the maximizer of the OLS selection rule (2), we have $\{\ell\} = \arg \max_{i \notin S_{k-1}} \{\mathbf{r}_{k-1}^t \tilde{\mathbf{b}}_i^{S_{k-1}}\}$. So, ℓ is also selected by SNNOLS.

6.3.3. 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} , and let us denote by $S_k = S_{k-1} \cup \{\ell\}$ the OLS support at iteration k . Since OLS selects the atom yielding the minimum squared error, we have for all $i \notin S_k$,

$$\min_{\mathbf{z}} \|\mathbf{y} - A_{S_k} \mathbf{z}\|^2 < \min_{\mathbf{z}} \|\mathbf{y} - A_{S_{k-1} \cup \{i\}} \mathbf{z}\|^2. \quad (17)$$

The left-hand side (LHS) of (17) rereads $\min_{\mathbf{z} \geq \mathbf{0}} \|\mathbf{y} - A_{S_k} \mathbf{z}\|^2$ according to (15). Also, it is clear that the right-hand side (RHS) of (17) is upper bounded by $\min_{\mathbf{z} \geq \mathbf{0}} \|\mathbf{y} - A_{S_{k-1} \cup \{i\}} \mathbf{z}\|^2$. Therefore, we get for all $i \notin S_k$,

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

which implies that ℓ is also selected by NNOLS.

6.4. Proof of Lemma 3.3

Here, the abridged tilded notations $\tilde{\mathbf{y}}, \tilde{\mathbf{a}}_i, \tilde{\mathbf{b}}_i, \tilde{\mathbf{n}}$ correspond to projected vectors onto $\text{span}(A_{S_{k-1}})^\perp$. We also use the context-dependent notation

$$\tilde{\mathbf{c}}_i = \begin{cases} \tilde{\mathbf{a}}_i & (\text{OMP case}), \\ \tilde{\mathbf{b}}_i & (\text{OLS case}) \end{cases} \quad (18)$$

to refer to $\tilde{\mathbf{a}}_i$ or $\tilde{\mathbf{b}}_i$. ℓ denotes the atom selected at iteration k , so $S_k \setminus S_{k-1} = \{\ell\}$. Let us recall that $S_k \subset S^*$ (Lemma 2.1). Note also that A_{S_k} is full column rank since $\mu(A) < \frac{1}{2K-1}$ (Lemma 6.1) and then $\tilde{\mathbf{a}}_\ell \neq \mathbf{0}$ (Lemma 6.3). Applying Lemma 3.2 for $j \leftarrow k-1$ yields (16), so $\hat{x}_\ell^{(k)}$ has the same sign as $\mathbf{r}_{k-1}^\top \tilde{\mathbf{a}}_\ell$ and as $\mathbf{r}_{k-1}^\top \tilde{\mathbf{b}}_\ell$. The remaining part of the proof consists in showing that $\mathbf{r}_{k-1}^\top \tilde{\mathbf{c}}_\ell > 0$.

Since atom ℓ is selected at the k -th iteration, we have

$$\forall j \notin S_k, |\mathbf{r}_{k-1}^\top \tilde{\mathbf{c}}_\ell| \geq |\mathbf{r}_{k-1}^\top \tilde{\mathbf{c}}_j|. \quad (19)$$

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

$$|\mathbf{r}_{k-1}^\top \tilde{\mathbf{c}}_j| > -\mathbf{r}_{k-1}^\top \tilde{\mathbf{c}}_\ell. \quad (20)$$

(20) can be consistent with (19) only if $\mathbf{r}_{k-1}^\top \tilde{\mathbf{c}}_\ell > 0$. It is therefore sufficient to show that (20) holds to complete the proof. Specifically, we show (20) for any

$$j \in \arg \max_{i \in \{1, \dots, n\} \setminus S_k} x_i^*. \quad (21)$$

Let $\mathbf{y} = A\mathbf{x}^* + \mathbf{n} = \sum_{i \in S^*} x_i^* \mathbf{a}_i + \mathbf{n}$. The residual at iteration $k-1$ reads

$$\mathbf{r}_{k-1} = P_{S_{k-1}}^\perp \mathbf{y} = \sum_{i \in S^* \setminus S_{k-1}} x_i^* \tilde{\mathbf{a}}_i + \tilde{\mathbf{n}}. \quad (22)$$

So for any $p \in S^* \setminus S_{k-1}$,

$$\mathbf{r}_{k-1}^\top \tilde{\mathbf{c}}_p = x_p^* \tilde{\mathbf{a}}_p^\top \tilde{\mathbf{c}}_p + \sum_{i \in S^* \setminus (S_{k-1} \cup \{p\})} x_i^* \tilde{\mathbf{a}}_i^\top \tilde{\mathbf{c}}_p + \tilde{\mathbf{n}}^\top \tilde{\mathbf{c}}_p. \quad (23)$$

Since $x_p^* = 0$ when $p \notin S^*$, (23) is also met for $p \notin S^*$. Thus, (23) holds for all $p \notin S_{k-1}$.

Since $K \geq k \geq 1$, $\mu(A) < \frac{1}{2K-1} \leq \frac{1}{k}$. Lemma 6.2 implies that $\tilde{\mathbf{a}}_p^t \tilde{\mathbf{a}}_p \geq \beta_{k-1}$ and $|\tilde{\mathbf{a}}_i^t \tilde{\mathbf{a}}_p| \leq \mu_{k-1} \beta_{k-1}$ with β_{k-1} and μ_{k-1} defined in (10). From $\tilde{\mathbf{b}}_p = \frac{\tilde{\mathbf{a}}_p}{\|\tilde{\mathbf{a}}_p\|}$, we get $\tilde{\mathbf{a}}_p^t \tilde{\mathbf{b}}_p = \|\tilde{\mathbf{a}}_p\| \geq \beta_{k-1}^{1/2}$ and $|\tilde{\mathbf{a}}_i^t \tilde{\mathbf{b}}_p| \leq \mu_{k-1} \beta_{k-1}^{1/2}$. The latter bounds can be rewritten in a unified way: $\tilde{\mathbf{a}}_p^t \tilde{\mathbf{c}}_p \geq \eta_{k-1}$ and $|\tilde{\mathbf{a}}_i^t \tilde{\mathbf{c}}_p| \leq \mu_{k-1} \eta_{k-1}$ with

$$\eta_{k-1} = \begin{cases} \beta_{k-1} & \text{if } \tilde{\mathbf{c}} \leftarrow \tilde{\mathbf{a}}, \\ \beta_{k-1}^{1/2} & \text{if } \tilde{\mathbf{c}} \leftarrow \tilde{\mathbf{b}}. \end{cases} \quad (24)$$

Furthermore, the Cauchy-Schwarz inequality implies that $|\tilde{\mathbf{n}}^t \tilde{\mathbf{c}}_p| \leq \|\tilde{\mathbf{n}}\| \leq \|\mathbf{n}\|$ since $\|\tilde{\mathbf{c}}_p\| \leq 1$. (23) yields

$$\mathbf{r}_{k-1}^t \tilde{\mathbf{c}}_p \geq \eta_{k-1} \left(x_p^* - \mu_{k-1} \sum_{i \in S^* \setminus (S_{k-1} \cup \{p\})} x_i^* \right) - \|\mathbf{n}\|. \quad (25)$$

Notice that $S_{k-1} \cup \{p\} = S_k$ for $p = \ell$. Using the definition of j in (21) and the fact that $\ell \in S^* \setminus S_{k-1}$, we have

$$\sum_{i \in S^* \setminus (S_{k-1} \cup \{\ell\})} x_i^* = \sum_{i \in S^* \setminus S_k} x_i^* \leq (K-k)x_j^*, \quad (26)$$

$$\forall p \neq \ell, \quad \sum_{i \in S^* \setminus (S_{k-1} \cup \{p\})} x_i^* \leq \sum_{i \in S^* \setminus S_{k-1}} x_i^* = x_\ell^* + (K-k)x_j^* \quad (27)$$

with $j \neq \ell$ because $j \notin S_k$. Now, apply (25) twice with $p \leftarrow \ell$ and $p \leftarrow j$:

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

Summing up both inequalities, we get

$$\mathbf{r}_{k-1}^t \tilde{\mathbf{c}}_j + \mathbf{r}_{k-1}^t \tilde{\mathbf{c}}_\ell \geq \eta_{k-1} \left((1 - \mu_{k-1})x_\ell^* + (1 - 2(K-k)\mu_{k-1})x_j^* \right) - 2\|\mathbf{n}\|.$$

One can easily check from (10) that $\mu_{k-1} < \frac{1}{2(K-k)}$ when $\mu(A) < \frac{1}{2K-1}$. Thus,

$$\begin{aligned} \mathbf{r}_{k-1}^t \tilde{\mathbf{c}}_j + \mathbf{r}_{k-1}^t \tilde{\mathbf{c}}_\ell &\geq \eta_{k-1} (1 - \mu_{k-1})x_\ell^* - 2\|\mathbf{n}\| \\ &> \left(\eta_{k-1} (1 - \mu_{k-1}) - (1 - (2K-1)\mu(A)) \right) \min_{i \in S^*} x_i^* \quad (28) \end{aligned}$$

where the last inequality follows from (4). (8) and (24) imply that $\beta_{k-1} \leq 1$,

thus $\eta_{k-1} \geq \beta_{k-1}$. Since $\mu_{k-1} < 1$ and using (10), we get

$$\begin{aligned} \eta_{k-1}(1 - \mu_{k-1}) &\geq \beta_{k-1}(1 - \mu_{k-1}) = \frac{(1 + \mu(A))}{1 - (k-2)\mu(A)} (1 - k\mu(A)) \\ &\geq \begin{cases} 1 - \mu(A) & \text{if } k = 1, \\ (1 + \mu(A))(1 - k\mu(A)) & \text{if } k \geq 2. \end{cases} \end{aligned} \quad (29)$$

When $k = 2$, the RHS in (29) can be expanded and then lower bounded by $1 - (2k - 1)\mu(A)$ by noticing that $\mu(A)^2 \leq \mu(A) \leq 1$. The latter bound holds as well when $k = 1$.

From (28), we get $\mathbf{r}_{k-1}^t \tilde{\mathbf{c}}_j + \mathbf{r}_{k-1}^t \tilde{\mathbf{c}}_\ell > 0$, which implies (20). We conclude from (19) that $\mathbf{r}_{k-1}^t \tilde{\mathbf{c}}_\ell > 0$ and thus $\hat{x}_\ell^{(k)} > 0$.

6.5. Proof of Lemma 3.4

In this proof, the abridged tilded notations $\tilde{\mathbf{a}}_i$ and $\tilde{\mathbf{b}}_i$ refer to projected vectors onto $\text{span}(A_{S_{j-1}})^\perp$ (the latter space differs from that of Subsection 6.4). We still use the generic notation $\tilde{\mathbf{c}}_i$ in (18) to refer to either $\tilde{\mathbf{a}}_i$ or $\tilde{\mathbf{b}}_i$ depending on the context. Similarly, \tilde{C} refers to the matrices \tilde{A} and \tilde{B} gathering the projected atoms. To prove that $\hat{\mathbf{x}}_{S_{k-1}}^{(k)} > \mathbf{0}$, we will exploit that $\hat{x}_{S_j \setminus S_{j-1}}^{(j)} > 0$ for $j < k$, which holds according to Lemma 3.3.

Let $\ell \in S_{k-1}$ and denote by $j < k$ the iteration at which the atom \mathbf{a}_ℓ has been selected by Oxx, so that $S_j \setminus S_{j-1} = \{\ell\}$. According to Lemma 3.2 and since ℓ is the first entry in the ordered set $S_k \setminus S_{j-1}$, we have

$$\hat{x}_\ell^{(k)} = ((\tilde{A}_{S_k \setminus S_{j-1}})^\dagger \mathbf{r}_{j-1})_1. \quad (30)$$

The full rankness of A_{S_k} implies that $\tilde{A}_{S_k \setminus S_{j-1}}$ is full column rank (see Lemma 6.3). Then, Lemma 6.4 yields:

$$\hat{x}_\ell^{(k)} \overset{+}{\propto} \mathbf{r}_{j-1}^t (\tilde{\mathbf{a}}_\ell - \tilde{P}_{S_k \setminus S_j} \tilde{\mathbf{a}}_\ell) \quad (31)$$

where $\overset{+}{\propto}$ refers to proportionality up to a positive factor, and $\tilde{P}_{S_k \setminus S_j}$ denotes the orthogonal projection onto $\text{span}(\tilde{A}_{S_k \setminus S_j})$. Since for all i , $\tilde{\mathbf{b}}_i$ is proportional to $\tilde{\mathbf{a}}_i$, we have $\text{span}(\tilde{A}_{S_k \setminus S_j}) = \text{span}(\tilde{B}_{S_k \setminus S_j})$, thus $\tilde{P}_{S_k \setminus S_j} = \tilde{C}_{S_k \setminus S_j} G_C^{-1} \tilde{C}_{S_k \setminus S_j}^t$

holds for both $\tilde{C} = \tilde{A}$ and $\tilde{C} = \tilde{B}$, with $G_C := \tilde{C}_{S_k \setminus S_j}^t \tilde{C}_{S_k \setminus S_j}$. Since $\tilde{\mathbf{a}}_\ell$ is proportional to $\tilde{\mathbf{c}}_\ell$ up to a positive factor, (31) implies that

$$\hat{x}_\ell^{(k)} > 0 \iff \mathbf{r}_{j-1}^t \tilde{C}_{S_k \setminus S_j} G_C^{-1} \tilde{C}_{S_k \setminus S_j}^t \tilde{\mathbf{c}}_\ell < \mathbf{r}_{j-1}^t \tilde{\mathbf{c}}_\ell. \quad (32)$$

By Hölder's inequality, the LHS of (32) is upper bounded by

$$\begin{aligned} \|\tilde{C}_{S_k \setminus S_j}^t \tilde{\mathbf{c}}_\ell\|_\infty \|G_C^{-1} \tilde{C}_{S_k \setminus S_j}^t \mathbf{r}_{j-1}\|_1 &\leq \mu(\tilde{C}) \|G_C^{-1}\|_{1,1} \|\tilde{C}_{S_k \setminus S_j}^t \mathbf{r}_{j-1}\|_1 \\ &\leq (k-j) \mu(\tilde{C}) \|G_C^{-1}\|_{1,1} |\mathbf{r}_{j-1}^t \tilde{\mathbf{c}}_\ell| \end{aligned} \quad (33)$$

where the last inequality exploits that index ℓ is selected at the j -th iteration of Oxx, *i.e.*, $\forall i, |\mathbf{r}_{j-1}^t \tilde{\mathbf{c}}_i| \leq |\mathbf{r}_{j-1}^t \tilde{\mathbf{c}}_\ell|$. From Lemma 3.3, we have $\hat{x}_\ell^{(j)} > 0$, hence $\mathbf{r}_{j-1}^t \tilde{\mathbf{c}}_\ell > 0$, see (16). We deduce from (32)-(33) that

$$((k-j) \mu(\tilde{C}) \|G_C^{-1}\|_{1,1} < 1) \implies (\hat{x}_\ell^{(k)} > 0). \quad (34)$$

To complete the proof, we will show that the LHS of (34) holds true. Let us start by bounding $\mu(\tilde{C})$. First, we note that $\tilde{\mathbf{c}}_p = \mathbf{0}$ for $p \in S_{j-1}$. Since $\mu(A) < \frac{1}{2K-1} < \frac{1}{j}$, Lemma 6.2 yields $\mu(\tilde{A}) \leq \mu_{j-1} \beta_{j-1}$ and $\mu_{j-1} < \frac{1}{2K-j}$. Since $\tilde{\mathbf{b}}_p = \frac{\tilde{\mathbf{a}}_p}{\|\tilde{\mathbf{a}}_p\|}$ for $p \notin S_{j-1}$, it follows from (8) that

$$\mu(\tilde{B}) \leq \frac{\mu(\tilde{A})}{\beta_{j-1}} < \frac{1}{2K-j}. \quad (35)$$

The cases of OMP and OLS are now treated separately.

OLS case. $\tilde{B}_{S_k \setminus S_j}$ is a column normalized matrix corresponding to a subset of columns of \tilde{B} . Moreover, $\mu(\tilde{B}_{S_k \setminus S_j}) \leq \mu(\tilde{B}) < \frac{1}{2K-j} \leq \frac{1}{k-j}$ according to (35). So, we can apply Lemma 6.1 to matrix $\tilde{B}_{S_k \setminus S_j}$: the Gram matrix $G_B = \tilde{B}_{S_k \setminus S_j}^t \tilde{B}_{S_k \setminus S_j}$ is invertible and

$$\|G_B^{-1}\|_{1,1} \leq \frac{1}{1 - (k-j-1)\mu(\tilde{B}_{S_k \setminus S_j})} \leq \frac{2K-j}{2K-k+1}, \quad (36)$$

where the last bound follows from (35). Combining (35) and (36), $\mu(\tilde{B}) \|G_B^{-1}\|_{1,1}$ is upper bounded by $\frac{1}{2K-k+1} < \frac{1}{k-j}$. So, the LHS of (34) holds true.

OMP case. Let Δ be the square diagonal matrix $\text{diag}(\|\tilde{\mathbf{a}}_i\|, i \in S_k \setminus S_j)$. Clearly, $\tilde{A}_{S_k \setminus S_j} = \tilde{B}_{S_k \setminus S_j} \Delta$ and $G_A^{-1} = \Delta^{-1} G_B^{-1} \Delta^{-1}$. The submultiplicative property of induced norms yields

$$\|G_A^{-1}\|_{1,1} \leq \|G_B^{-1}\|_{1,1} \|\Delta^{-1}\|_{1,1}^2 \quad (37)$$

where $\|\Delta^{-1}\|_{1,1} = \max\{\|\tilde{\mathbf{a}}_i\|^{-1}, i \in S_k \setminus S_j\} \leq \beta_{j-1}^{-1/2}$ by Lemma 6.2. Thus,

$$\mu(\tilde{A}) \|G_A^{-1}\|_{1,1} \leq \frac{\mu(\tilde{A})}{\beta_{j-1}} \|G_B^{-1}\|_{1,1} < \frac{1}{2K - k + 1} < \frac{1}{k - j}$$

according to (35) and (36). We conclude that the LHS of (34) is true.

References

- [1] Barbu, I., Herzet, C., 2016. A new approach for volume reconstruction in tomoPIV with the alternating direction method of multipliers. *Meas. Sci. Technol.* 27, 104002.
- [2] Ben-Haim, Z., Eldar, Y.C., Elad, M., 2010. Coherence-based performance guarantees for estimating a sparse vector under random noise. *IEEE Trans. Signal Process.* 58, 5030–5043.
- [3] Blumensath, T., Davies, M.E., 2007. On the difference between Orthogonal Matching Pursuit and Orthogonal Least Squares. Technical Report. University of Edinburgh.
- [4] Bourguignon, S., Ninin, J., Carfantan, H., Mongeau, M., 2016. Exact sparse approximation problems via mixed-integer programming: Formulations and computational performance. *IEEE Trans. Signal Process.* 64, 1405–1419.
- [5] Bruckstein, A.M., Elad, M., Zibulevsky, M., 2008. On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations. *IEEE Trans. Inf. Theory* 54, 4813–4820.
- [6] Cai, T.T., Wang, L., 2011. Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Trans. Inf. Theory* 57, 4680–4688.

- [7] Cai, T.T., Wang, L., Xu, G., 2010. Stable recovery of sparse signals using an oracle inequality. *IEEE Trans. Inf. Theory* 56, 3516–3522.
- [8] Chang, L.H., Wu, J.Y., 2014. An improved RIP-based performance guarantee for sparse signal recovery via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* 60, 5702–5715.
- [9] Chen, S., Billings, S.A., Luo, W., 1989. Orthogonal least squares methods and their application to non-linear system identification. *Int. J. Control* 50, 1873–1896.
- [10] Cherni, A., Chouzenoux, E., Delsuc, M.A., 2018. Fast dictionary-based approach for mass spectrometry data analysis, in: *Proc. IEEE ICASSP*, Calgary, Canada. pp. 816–820.
- [11] Cotter, S.F., Adler, J., Rao, B.D., Kreutz-Delgado, K., 1999. Forward sequential algorithms for best basis selection. *IEE Proc. Vision, Image and Signal Processing* 146, 235–244.
- [12] Davenport, M.A., Wakin, M.B., 2010. Analysis of orthogonal matching pursuit using the restricted isometry property. *IEEE Trans. Inf. Theory* 56, 4395–4401.
- [13] Donoho, D.L., Elad, M., Temlyakov, V.N., 2006. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theory* 52, 6–18.
- [14] Donoho, D.L., Tsaig, Y., 2008. Fast solution of ℓ_1 -norm minimization problems when the solution may be sparse. *IEEE Trans. Inf. Theory* 54, 4789–4812.
- [15] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. *Ann. Statist.* 32, 407–499.
- [16] Foucart, S., Rauhut, H., 2013. A mathematical introduction to compressive sensing. *Applied and Numerical Harmonic Analysis*, Birkhäuser, Basel.

- [17] Fuchs, J.J., 2004. On sparse representations in arbitrary redundant bases. *IEEE Trans. Inf. Theory* 50, 1341–1344.
- [18] Herzet, C., Drémeau, A., Soussen, C., 2016. Relaxed recovery conditions for OMP/OLS by exploiting both coherence and decay. *IEEE Trans. Inf. Theory* 62, 459–470.
- [19] Herzet, C., Soussen, C., Idier, J., Gribonval, R., 2013. Exact recovery conditions for sparse representations with partial support information. *IEEE Trans. Inf. Theory* 59, 7509–7524.
- [20] Iordache, M.D., Bioucas-Dias, J.M., Plaza, A., 2011. Sparse unmixing of hyperspectral data. *IEEE Trans. Geosci. Remote Sensing* 49, 2014–2039.
- [21] Kim, D., Haldar, J.P., 2016. Greedy algorithms for nonnegativity-constrained simultaneous sparse recovery. *Signal Process.* 125, 274–289.
- [22] Kim, J., Wang, J., Shim, B., 2021. Optimal restricted isometry condition of normalized sampling matrices for exact sparse recovery with orthogonal least squares. *IEEE Trans. Signal Process.* 69, 1521–1536.
- [23] Kumar, A., Sindhwani, V., Kambadur, P., 2013. Fast conical hull algorithms for near-separable non-negative matrix factorization, in: *Proc. Int. Conf. Mach. Learn.*, Atlanta, GA. pp. 231–239.
- [24] Li, B., Shen, Y., Wu, Z., Li, J., 2015. Sufficient conditions for generalized orthogonal matching pursuit in noisy case. *Signal Process.* 108, 111–123.
- [25] Miller, A.J., 2002. *Subset selection in regression*. 2nd ed., Chapman and Hall, London, UK.
- [26] Mo, Q., Shen, Y., 2012. A remark on the restricted isometry property in orthogonal matching pursuit. *IEEE Trans. Inf. Theory* 58, 3654–3656.
- [27] Mørup, M., Madsen, K.H., Hansen, L.K., 2008. Approximate L0 constrained non-negative matrix and tensor factorization, in: *IEEE Int. Symp. Circuits Syst.*, Seattle, WA. pp. 1328–1331.

- [28] Nadisic, N., Vandaele, A., Gillis, N., Cohen, J.E., 2020. Exact sparse nonnegative least squares, in: Proc. IEEE ICASSP, Barcelona, Spain. pp. 5395–5399.
- [29] Nguyen, T.T., Idier, J., Soussen, C., Djermoune, E.H., 2019. Non-negative orthogonal greedy algorithms. *IEEE Trans. Signal Process.* 67, 5643–5658.
- [30] Pati, Y.C., Rezaifar, R., Krishnaprasad, P.S., 1993. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in: Proc. ACSSC, pp. 40–44.
- [31] Rebollo-Neira, L., Lowe, D., 2002. Optimized orthogonal matching pursuit approach. *IEEE Signal Process. Lett.* 9, 137–140.
- [32] Soussen, C., Gribonval, R., Idier, J., Herzet, C., 2013. Joint k -step analysis of orthogonal matching pursuit and orthogonal least squares. *IEEE Trans. Inf. Theory* 59, 3158–3174.
- [33] Tibshirani, R., 1996. Regression shrinkage and selection via the Lasso. *J. R. Statist. Soc. B* 58, 267–288.
- [34] Tropp, J.A., 2004. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory* 50, 2231–2242.
- [35] Wen, J., Wang, J., Zhang, Q., 2017a. Nearly optimal bounds for orthogonal least squares. *IEEE Trans. Signal Process.* 65, 5347–5356.
- [36] Wen, J., Zhou, Z., Wang, J., Tang, X., Mo, Q., 2017b. A sharp condition for exact support recovery with orthogonal matching pursuit. *IEEE Trans. Signal Process.* 65, 1370–1382.
- [37] Yaghoobi, M., Davies, M.E., 2015. Fast non-negative orthogonal least squares, in: Proc. Eur. Sig. Proc. Conf., Nice, France. pp. 479–483.
- [38] Yaghoobi, M., Wu, D., Davies, M.E., 2015. Fast non-negative orthogonal matching pursuit. *IEEE Signal Process. Lett.* 22, 1229–1233.