



HAL
open science

Non-negative orthogonal greedy algorithms

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

► **To cite this version:**

Thi Thanh Nguyen, Jérôme Idier, Charles Soussen, El-Hadi Djermoune. Non-negative orthogonal greedy algorithms. *IEEE Transactions on Signal Processing*, 2019, 67 (21), pp.5643-5658. 10.1109/TSP.2019.2943225 . hal-02049424v3

HAL Id: hal-02049424

<https://hal.science/hal-02049424v3>

Submitted on 22 Aug 2019

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.

Non-Negative Orthogonal Greedy Algorithms

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

Abstract—Orthogonal greedy algorithms are popular sparse signal reconstruction algorithms. Their principle is to select atoms one by one. A series of unconstrained least-squares subproblems of gradually increasing size is solved to compute the approximation coefficients, which is efficiently performed using a fast recursive update scheme. When dealing with non-negative sparse signal reconstruction, a series of *non-negative* least-squares subproblems have to be solved. Fast implementation becomes tricky since each subproblem does not have a closed-form solution anymore. Recently, non-negative extensions of the classical orthogonal matching pursuit and orthogonal least squares algorithms were proposed, using slow (*i.e.*, non-recursive) or recursive but inexact implementations. In this paper, we revisit these algorithms in a unified way. We define a class of non-negative orthogonal algorithms and exhibit their structural properties. We propose a fast and exact implementation based on the active-set resolution of non-negative least-squares and exploiting warm start initializations. The algorithms are assessed in terms of accuracy and computational complexity for a sparse spike deconvolution problem. We also present an application to near-infrared spectra decomposition.

Index Terms—Orthogonal greedy algorithms; sparse reconstruction; non-negativity; non-negative least-squares; active-set algorithms.

I. INTRODUCTION

In the last decade, sparse approximation has received considerable attention in the signal and image processing community, in connection with compressive sensing. Greedy algorithms for sparse signal reconstruction are very popular iterative schemes. Their principle is to repeatedly (*i*) enrich the sparse support by selecting a new dictionary atom, and then (*ii*) update the sparse approximation coefficients. In orthogonal greedy algorithms, the sparse approximation signal is computed as the orthogonal projection of the data vector onto the subspace spanned by the selected atoms. Therefore, the coefficients can be estimated by solving an Unconstrained Least Squares (ULS) problem. Popular orthogonal greedy algorithms include Orthogonal Matching Pursuit (OMP) [1] and Orthogonal Least Squares (OLS) [2], also known as forward selection [3], Order Recursive Matching Pursuit (ORMP) [4], Optimized Orthogonal Matching Pursuit (OOMP) [5] and Pure

Orthogonal Matching Pursuit [6]. OMP and OLS differ in the way the new atom is selected. In both cases, the atom inducing the largest decrease of the residual norm is selected. However, all nonzero atom weights are optimally tuned in OLS whereas only the new atom weight is considered in OMP, which amounts to selecting the atom having the largest inner product with the current residual. The computational complexity of OLS is obviously higher, since the selection rule requires to solve as many ULS problems as the number of candidate atoms. Fortunately, the ULS solutions have a closed-form expression, which can be recursively (fastly) updated when the support is enriched by a new element, see *e.g.*, [3]. Specifically, both OMP and OLS implementations are recursive and make use of matrix factorization, such as Gram-Schmidt orthogonalization, the Cholesky factorization or techniques utilizing the matrix inversion lemma [7].

Many application fields such as geoscience and remote sensing [8], [9], audio [10], chemometrics [11], bioinformatics [12], astrophysics [13] and computed tomography [14], give rise to inverse problems where the signal or image of interest is sparse, but also non-negative. In such contexts, a common practice is to regularize the inverse problem in order to favor both sparsity and non-negativity of the sought signal, see, *e.g.*, [12], [14]–[16]. Some classical sparse algorithms can be straightforwardly adapted to deal with non-negativity constraints. This is the case of proximal splitting algorithms and the Alternating Direction Method of Multipliers (ADMM) for convex optimization [17], [18], and of the DC algorithm (Difference of Convex functions) for nonconvex optimization [19], [20]. On the contrary, the non-negative extension of greedy algorithms is a challenging issue since the unconstrained least-squares subproblems are replaced by non-negative least-squares (NNLS) subproblems which do not have closed-form solutions anymore, so a subroutine solving NNLS is needed. There are different methods for NNLS solving [21] such as active-set [22], interior-point [23], and gradient-projection [24]. The latter two families typically require to tune some stopping criteria in an empirical manner, resulting in approximate resolution of the NNLS problem. Here, we are focusing on active-set methods since such methods have a greedy structure and exactly solve NNLS problems after a finite number of iterations. Although our focus will be on extensions of orthogonal greedy schemes to the non-negative case, let us mention that several other non-negative sparse methods have been elaborated on the basis of the active-set NNLS algorithm, *e.g.*, hard-thresholded NNLS [25], [26] and Sparse NNLS or Reverse Sparse NNLS [27].

The classical CoSaMP [28], Subspace Pursuit (SP) [29] and Hard Thresholding Pursuit [6] algorithms were extended to the non-negative setting in [30]. Similar to orthogonal greedy

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

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

T. T. Nguyen and E.-H. Djermoune are with the Université de Lorraine and CNRS at the Centre de Recherche en Automatique de Nancy (CRAN), 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).

algorithms, the non-negative versions NN-CoSaMP, NN-SP and NN-HTP are based on straightforward replacements of ULS resolutions by NNLS resolutions. The principle of these algorithms is to update a support of fixed size at consecutive iterations using operations such as support expansion and support shrinkage. Since two consecutive supports are possibly substantially different, one can hardly exploit the knowledge of the current NNLS solution to compute the solution related to the new support in a very efficient manner. On the contrary, the supports found at consecutive iterations of orthogonal greedy algorithms are nested, allowing one to use warm start initialization. The nested property is therefore crucial to design fast implementations.

Several existing contributions deal with orthogonal greedy algorithms in the non-negative case. Non-Negative OMP (NNOMP) was first proposed by Bruckstein *et al.* [31] as a direct generalization of OMP. At each iteration, the atom having the maximum *positive* inner product with the current residual is selected. Contrary to OMP, negative inner products are discarded. Then, the sparse approximation coefficients are updated by solving the NNLS problem related to the augmented subset. The canonical (*i.e.*, non-recursive) NNOMP implementation of Bruckstein *et al.* [31] solves NNLS subproblems independently. Later, Yaghoobi *et al.* proposed an accelerated version named Fast Non-Negative OMP (FN-NOMP), which avoids solving NNLS subproblems but rather recursively approximates the sought solution using QR matrix factorization [32]. Although FNNOMP is much faster than canonical NNOMP, it is an approximate version likely to deliver a different output. In [33], Yaghoobi *et al.* introduced a canonical version of Non-Negative OLS (NNOLS), defined as a direct non-negative extension of OLS. The principle of NNOLS is to select the atom for which the positive residual (the residual corresponding to the NNLS solution) is minimum. This selection rule appears to be time-consuming since one needs to compute as many positive residuals as the number of candidate atoms, *i.e.*, $n - k$ NNLS problems have to be solved at iteration k , with n the size of the dictionary.

Yaghoobi *et al.* [33] further proposed two accelerated versions of NNOLS named Suboptimal NNOLS (SNNOLS) and Fast NNOLS (FNNOLS). These proposals can be understood from a dual interpretation of OLS in the unconstrained setting. It is well-known that the OLS selection rule can be formulated in two equivalent ways [5], [34]: OLS selects the atom (*i*) inducing the minimum value of the data residual in ℓ_2 -norm; (*ii*) whose orthogonal projection onto the subspace orthogonal to the active atoms has a maximum angle with the current data residual. When dealing with non-negativity, SNNOLS [33] selects the atom that is positively correlated with the current residual whose projection forms a maximum angle with the residual. However, the equivalence that holds for OLS is not true anymore since maximizing the angle for positive inner products is not equivalent to minimizing the ℓ_2 -norm of the positive residual. Therefore, SNNOLS iterates do not generally identify with NNOLS iterates, hence the *suboptimal* algorithm denomination. In SNNOLS, the projected atoms have to be recomputed whenever the support is updated. Moreover, an NNLS problem must be solved at each iteration to update

the sparse approximation coefficients. FNNOLS is a recursive implementation in the same spirit as FNNOMP, where no NNLS problem needs to be solved anymore. It shall be noticed that FNNOLS and SNNOLS do not necessarily deliver the same iterates, and that both can be viewed as approximate versions of NNOLS.

Generally speaking, the “orthogonal” denomination of NNOMP and NNOLS is somewhat abusive since when the support set S is updated, the related NNLS solution may not identify with the orthogonal projection of the data vector onto the span of atoms indexed by S , but rather with its projection onto their positive span. Both projected vectors differ as soon as the NNLS solution has zero entries, *i.e.*, when some non-negativity constraints become active. Therefore, in NNOMP, NNOLS and their derived versions [31]–[33], the support of the sparse vector at the current iteration is a subset of the current support set S (which is expanded at each iteration) and may not identify with it. In turn, more than K iterations may be necessary to reach a truly K -sparse representation.

Our contributions are twofold. First, non-negative orthogonal greedy algorithms are revisited in a unified way. The algorithms under study share four desirable properties:

- 1) The norm of the data residual is always decreasing when a new atom enters the solution support.
- 2) The algorithm does not stop while additional atom selections would make it decrease, unless an explicit stopping condition is reached.
- 3) A so-called compression step is included to shrink the support set by removing the atoms having zero coefficients, so the support set identifies to the support of the current NNLS solution.
- 4) The residual vector is orthogonal to the selected atoms. In other words, the sparse approximation vector identifies with the orthogonal projection of the data vector onto the span of the selected atoms.

These structural properties are exhibited and compared to those of existing non-negative greedy algorithms. The second contribution is a fast and exact implementation of non-negative orthogonal greedy algorithms exploiting the active-set algorithm [22] for NNLS solving, based on warm start initialization. Moreover, we elaborate on recursive implementations and design further types of accelerations.

The paper is organized as follows. Section II introduces the family of so-called *Non-negative Orthogonal Greedy* (NNOG) algorithms. The different members of the family differ by the selection rule to pick a new atom at each iteration. It includes NNOLS, SNNOLS and NNOLS up to a modification of their structure, namely the compression step mentioned above. In Section III, a fast implementation is proposed based on recursivity and on the use of warm starts for solving the NNLS subproblems. Section IV is devoted to NNOG acceleration. Sections V and VI propose numerical validations on a simulated example of sparse deconvolution and on the decomposition of real-world near-infrared (NIR) spectra. Finally, discussion and perspectives will be found in Section VII.

II. NON-NEGATIVE GREEDY ALGORITHMS

A. Basic definitions and notations

Given a data vector $\mathbf{y} \in \mathbb{R}^m$ and a dictionary $H \in \mathbb{R}^{m \times n}$, we are interested in finding a K -sparse non-negative weight vector $\mathbf{x} \in \mathbb{R}_+^n$ yielding an accurate approximation $\mathbf{y} \approx H\mathbf{x}$. This can be formulated as the constrained minimization program:

$$\min_{\mathbf{x} \geq \mathbf{0}} \|\mathbf{y} - H\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq K. \quad (\ell_0+)$$

where $\mathbf{x} \geq \mathbf{0}$ means that each entry of \mathbf{x} is nonnegative, $\|\mathbf{x}\|_0$ is the ℓ_0 -“norm” counting the number of nonzero entries in \mathbf{x} , and the quadratic fidelity-to-data term $\|\mathbf{y} - H\mathbf{x}\|_2^2$ measures the quality of approximation. The ℓ_2 -norm $\|\cdot\|_2$ will be also denoted $\|\cdot\|$. Without loss of generality, we assume that H is column-normalized. Each dictionary column \mathbf{h}_i , $i = 1, \dots, n$ is called an atom. We have the following useful identity for any two vectors \mathbf{r} , \mathbf{h} of same length, \mathbf{h} being normalized:

$$\min_{v \geq 0} \|\mathbf{r} - v\mathbf{h}\|^2 = \|\mathbf{r}\|^2 - (\max\{\mathbf{h}^t \mathbf{r}, 0\})^2 \quad (1)$$

where \cdot^t stands for the transpose operator.

We denote by $S = \text{supp}(\mathbf{x}) = \{i : \mathbf{x}(i) \neq 0\}$ the support of \mathbf{x} ($\mathbf{x}(i)$ being the i -th entry of \mathbf{x}), \bar{S} the complement of S , $|S|$ the cardinality of S , H_S and $\mathbf{x}(S)$ the subdictionary and subvector indexed by S , respectively. Finally, H^\dagger and $\text{span}(H)$ are the pseudo-inverse and the column space of H , respectively. Let $\tilde{\mathbf{h}}_{i,S} = \mathbf{h}_i - H_S H_S^\dagger \mathbf{h}_i$ stand for the orthogonal projection of \mathbf{h}_i onto the orthogonal complement $(\text{span}(H_S))^\perp$, which will be simply denoted $\tilde{\mathbf{h}}_i$ whenever unambiguous, and $\tilde{\mathbf{g}}_i = \tilde{\mathbf{h}}_i / \|\tilde{\mathbf{h}}_i\|$ denote the normalized projected atom if $\mathbf{h}_i \notin \text{span}(H_S)$, i.e., $\tilde{\mathbf{h}}_i \neq \mathbf{0}$. If $\mathbf{h}_i \in \text{span}(H_S)$, it will be convenient to set $\tilde{\mathbf{g}}_i = \mathbf{0}$.

For any support S , let us call an unconstrained least-squares (ULS) and a nonnegative least-squares (NNLS) solution corresponding to S , any vector \mathbf{x} in \mathbb{R}^n or \mathbb{R}_+^n , respectively, that minimizes $\|\mathbf{y} - H\mathbf{x}\|^2$ subject to $\text{supp}(\mathbf{x}) \subset S$. Such vectors will be denoted $\hat{\mathbf{x}}_S$ and $\hat{\mathbf{x}}_S^+$, respectively. The following notations will be also useful:

$$\begin{aligned} \mathbf{r}_S &= \mathbf{y} - H\hat{\mathbf{x}}_S, \\ \mathbf{r}_S^+ &= \mathbf{y} - H\hat{\mathbf{x}}_S^+. \end{aligned}$$

When H_S is full column rank, $\|\mathbf{y} - H_S \mathbf{z}\|^2$ is a strictly convex function of $\mathbf{z} \in \mathbb{R}^{|S|}$, so $\hat{\mathbf{x}}_S^+$ and $\hat{\mathbf{x}}_S$ are then uniquely defined. Throughout the paper, we will denote by $C \subset S$ the so-called compressed support, defined as the support of $\hat{\mathbf{x}}_S^+$. The NNLS optimal solutions can be characterized using the Karush-Kuhn-Tucker (KKT) conditions [22, Chap. 3], which are recalled next for completeness.

Lemma 1 Consider the NNLS problem related to support S :

$$\min_{\mathbf{x} \geq \mathbf{0}} \|\mathbf{y} - H\mathbf{x}\|^2 \quad \text{s.t.} \quad \text{supp}(\mathbf{x}) \subset S. \quad (2)$$

$\hat{\mathbf{x}}_S^+$ is a solution to (2) if and only if the KKT conditions are satisfied:

$$\begin{cases} H_C^t (\mathbf{y} - H\hat{\mathbf{x}}_S^+) = \mathbf{0} \\ H_{S \setminus C}^t (\mathbf{y} - H\hat{\mathbf{x}}_S^+) \leq \mathbf{0} \end{cases} \quad (3)$$

where $C := \text{supp}(\hat{\mathbf{x}}_S^+) \subset S$.

Proof: From the definition of C , it is clear that $\hat{\mathbf{x}}_S^+(C) > \mathbf{0}$, so the active constraints in (2) are indexed by \bar{C} . Let $\boldsymbol{\lambda} \in \mathbb{R}^n$ gather the Lagrange multipliers related to both equality and inequality constraints. The Lagrangian function induced by (2) is defined as $\mathcal{L}(\mathbf{x}; \boldsymbol{\lambda}) = \|\mathbf{y} - H\mathbf{x}\|^2 - \boldsymbol{\lambda}^t \mathbf{x}$ and the KKT conditions for optimal variables $(\hat{\mathbf{x}}_S^+, \hat{\boldsymbol{\lambda}})$ read:

$$\begin{cases} \nabla_{\mathbf{x}} \mathcal{L}(\hat{\mathbf{x}}_S^+; \hat{\boldsymbol{\lambda}}) = 2H^t(H\hat{\mathbf{x}}_S^+ - \mathbf{y}) - \hat{\boldsymbol{\lambda}} = \mathbf{0}, \\ \forall i \in S, \hat{\boldsymbol{\lambda}}(i)\hat{\mathbf{x}}_S^+(i) = 0 \text{ with } \hat{\mathbf{x}}_S^+(i) \geq 0, \hat{\boldsymbol{\lambda}}(i) \geq 0, \\ \forall i \notin S, \hat{\mathbf{x}}_S^+(i) = 0. \end{cases} \quad (4)$$

For quadratic programming problems involving positive semidefinite matrices, the KKT conditions are necessary and sufficient conditions of optimality [35, Chap. 16], so $\hat{\mathbf{x}}_S^+$ is a solution to (2) if and only if

$$\begin{cases} \hat{\boldsymbol{\lambda}} = -2H^t(\mathbf{y} - H\hat{\mathbf{x}}_S^+) \\ \hat{\boldsymbol{\lambda}}(C) = \mathbf{0} \\ \hat{\boldsymbol{\lambda}}(S \setminus C) \geq \mathbf{0} \end{cases}$$

that is, when (3) is satisfied. \blacksquare

Definition 1 Let us call a positive support related to the full-size NNLS problem

$$\min_{\mathbf{x} \geq \mathbf{0}} \|\mathbf{y} - H\mathbf{x}\|^2, \quad (5)$$

any index set S such that H_S is full column rank and $\hat{\mathbf{x}}_S^+(S) > \mathbf{0}$. By extension, the empty support $S = \emptyset$ will also be considered as a positive support.

Lemma 2 S is a positive support if and only if H_S is full rank and $\hat{\mathbf{x}}_S(S) > \mathbf{0}$. Moreover, when S is a positive support, $\hat{\mathbf{x}}_S^+ = \hat{\mathbf{x}}_S$, $\mathbf{r}_S^+ = \mathbf{r}_S$ and $H_S^t \mathbf{r}_S = \mathbf{0}$.

Proof: When S is a positive support, ULS and NNLS solutions $\hat{\mathbf{x}}_S$ and $\hat{\mathbf{x}}_S^+$ are uniquely defined and coincide. The orthogonality property $H_S^t \mathbf{r}_S = \mathbf{0}$ follows from (3). \blacksquare

Positive supports will play an important role in our specification of fast non-negative orthogonal greedy algorithms.

B. Non-negative orthogonal greedy algorithms

Let us define the class of *non-negative orthogonal greedy* (NNOG) algorithms, sharing the following general structure. We start from the empty support $S = \emptyset$. At each iteration, an atom is moved from \bar{S} to S . A new NNLS solution $\hat{\mathbf{x}}_S^+$ is then computed to optimally adapt the weights to the newly extended support. The algorithm stops when the desired cardinality K is reached or when the norm of the residual cannot decrease anymore. The general structure of NNOG algorithms is given by Algorithm 1. Some aspects will be made clear later, such as the role of the test $\mathbf{h}_i^t \mathbf{r}_S > 0$ with respect to the decrease of the norm of the residual.

The NNOG class is a direct adaptation of the family of orthogonal greedy algorithms from the unconstrained case to the nonnegative one. At first glance, the two families only differ by the fact that an NNLS solution is computed rather than a ULS one to update the weights at each iteration.

Algorithm 1: General structure of a non-negative orthogonal greedy algorithm to solve (ℓ_0+) .

input : \mathbf{y}, H, K
output: \mathbf{x}

```

1  $\mathbf{x} \leftarrow \mathbf{0}$  ;  $S \leftarrow \emptyset$  ;  $\mathbf{r}_S \leftarrow \mathbf{y}$  ;
2 while  $|S| < K$  and  $\max_{i \in \bar{S}} \mathbf{h}_i^t \mathbf{r}_S > 0$  do
3   Select an index  $\ell \in \bar{S}$  by a selection rule  $\mathcal{S}(\mathbf{y}, H, S)$  ;
4    $S \leftarrow S \cup \{\ell\}$  ;
5    $\mathbf{x} \leftarrow \hat{\mathbf{x}}_S^+$  ;
6    $S \leftarrow \text{supp}(\mathbf{x})$  ;
7    $\mathbf{r}_S = \mathbf{y} - H\mathbf{x}$  ;
8 end
```

However, some important features differ between the two cases, which require non trivial adaptations.

In both cases, the greedy character corresponds to the fact that a unique atom is added to the current support per iteration. However, a distinct feature of NNOG algorithms is that the support size may be smaller than the current iteration index, because some components of $\hat{\mathbf{x}}_S^+(S)$ may vanish due to the activation of the corresponding nonnegativity constraints. In the unconstrained case, some components of $\hat{\mathbf{x}}_S(S)$ may also vanish, but such events are fortuitous and do not need any specific consideration.

In the unconstrained case, \mathbf{r}_S is orthogonal to $\text{span}(H_S)$. This geometrical property does not necessarily hold for \mathbf{r}_S^+ because NNLS is an inequality constrained problem. Fortunately, provided that the indices of zero components of $\hat{\mathbf{x}}_S^+$ are moved to \bar{S} , it remains true that $\mathbf{y} - H\hat{\mathbf{x}}_S^+$ is orthogonal to $\text{span}(H_S)$. This is a direct consequence of the following lemma, which states that $\hat{\mathbf{x}}_S^+$ reads as a ULS solution related to the compressed version of support S .

Lemma 3 For any S , let $C = \text{supp}(\hat{\mathbf{x}}_S^+)$ (where neither $\hat{\mathbf{x}}_S^+$ nor C are necessarily unique if H_S is not full column rank). Then we have

$$\hat{\mathbf{x}}_S^+ = \hat{\mathbf{x}}_C = \hat{\mathbf{x}}_C^+. \quad (6)$$

Proof: Using Lemma 1, we have $H_C^t \mathbf{r}_S^+ = \mathbf{0}$. Since $\mathbf{r}_S^+ = \mathbf{y} - H\hat{\mathbf{x}}_S^+$, we get $H_C^t \mathbf{y} = H_C^t H \hat{\mathbf{x}}_S^+$. Thus, $\hat{\mathbf{x}}_S^+$ is a ULS solution (denoted by $\hat{\mathbf{x}}_C$) associated to support C . We have also $\hat{\mathbf{x}}_C = \hat{\mathbf{x}}_C^+$ since $\hat{\mathbf{x}}_C \geq \mathbf{0}$. ■

According to the above definition of NNOG algorithms, distinct algorithms can only differ by the selection rule used to select an atom at each iteration. The design of a selection rule corresponds to the definition of a function $\mathcal{S}(\mathbf{y}, H, S)$, taking values in \bar{S} . It is clear that some indices $\ell \in \bar{S}$ correspond to inappropriate choices, because their selection would produce $\hat{\mathbf{x}}_{S \cup \{\ell\}}^+(\ell) = 0$, and hence a useless iteration, and possibly an early stopping of the algorithm. In contrast, in the unconstrained case, any selection $\ell \in \bar{S}$ yields a decrease of $\|\mathbf{y} - H\mathbf{x}\|^2$ unless $\mathbf{h}_\ell \in \text{span}(H_S)$. The capacity of some selection rules to avoid inappropriate selections is examined in the next two subsections.

Finally, a practically important aspect is the computing cost of NNOG algorithms. It is computationally more demanding to

solve an NNLS problem than the corresponding ULS problem, so one must expect a larger computing cost. However, NNOG algorithms lend themselves to recursive implementations akin to usual orthogonal greedy schemes, as detailed in Section III.

C. Descending atoms and descent selection rules

Greedy algorithms can be interpreted as descent algorithms dedicated to the minimization of the residual norm using supports of growing size. Contrary to the unconstrained case, only the selection of some atoms in \bar{S} may produce a decrease of the residual at a given iteration of an NNOG algorithm. For the rest of the atoms in \bar{S} , the residual norm decrease is possible only if the nonnegativity constraint is violated. Such a specificity has both formal and practical consequences on the design of NNOG algorithms, as examined in this subsection.

Definition 2 For a given support S , let us define the set of indices corresponding to descending atoms as follows:

$$\mathcal{D}_S = \left\{ i \in \{1, \dots, n\}, \|\mathbf{r}_{S \cup \{i\}}^+\| < \|\mathbf{r}_S^+\| \right\}.$$

Obviously, we have $\mathcal{D}_S \subset \bar{S}$. In what follows, we focus on selection rules ensuring the selection of a descending atom at any iteration. The latter rules are referred to as *descent selection rules*.

Definition 3 A descent selection rule is a function $\mathcal{S}(\mathbf{y}, H, S)$ that takes its values in \mathcal{D}_S .

Clearly, NNOG algorithms relying on a descent selection rule are descent algorithms and we have the following property.

Lemma 4 NNOG algorithms relying on a descent selection rule terminate after a finite number of iterations.

Proof: The error norm decreases at each iteration, and there is a finite number of supports with a cardinality not exceeding K , and thus a finite number of solutions to visit. ■

The following proposition allows one to characterize whether an atom is descending.

Proposition 1 The descending atom condition $i \in \mathcal{D}_S$ is equivalent to

$$0 < \mathbf{h}_i^t \mathbf{r}_S^+. \quad (7)$$

When S is a positive support, it is also equivalent to each condition

$$0 < \mathbf{h}_i^t \mathbf{r}_S, \quad (8)$$

$$0 < \tilde{\mathbf{g}}_i^t \mathbf{r}_S^+, \quad (9)$$

$$0 < \tilde{\mathbf{g}}_i^t \mathbf{r}_S. \quad (10)$$

When $H_{S \cup \{i\}}$ is full column rank, it is also equivalent to

$$\hat{\mathbf{x}}_{S \cup \{i\}}^+(i) > 0. \quad (11)$$

Proof: See Appendix A-A. ■

Let us remark that from the first item of Proposition 1, the selection rule is invoked at Line 3 of Algorithm 1 only if $\mathcal{D}_S \neq \emptyset$, otherwise the stopping condition of Line 2 is

activated. Hence, we do not need to define $\mathcal{S}(\mathbf{y}, H, S)$ when $\mathcal{D}_S = \emptyset$.

The following lemmas have an interesting impact on the practical specification of NNOG algorithms.

Lemma 5 *If S is a positive support and $i \in \mathcal{D}_S$, then matrix $H_{S \cup \{i\}}$ is full column rank.*

Proof: Assume that S is positive, so H_S is full column rank, and $\hat{\mathbf{x}}_S^+ = \hat{\mathbf{x}}_S$. Let us also assume that $H_{S \cup \{i\}}$ is not full column rank. Then $\mathbf{h}_i \in \text{span}(H_S)$, so $\hat{\mathbf{x}}_S$ is a ULS solution related to $S \cup \{i\}$, and also an NNLS solution related to $S \cup \{i\}$ since $\hat{\mathbf{x}}_S \geq \mathbf{0}$. This implies that $i \notin \mathcal{D}_S$. ■

Lemma 6 *After each iteration of an NNOG algorithm relying on a descent selection rule, it holds that the support of the current solution is positive.*

Proof: The proof is immediate by recursive application of Lemmas 3 and 5, starting with the empty support. ■

Let us stress that Lemma 6 refers to NNOG algorithms strictly conforming to the scheme of Algorithm 1 (with an additional restriction to descent selection rules at Line 3). In particular, the *support compression* step performed at Line 6 is necessary to make Lemma 5 applicable. To our best knowledge, such a compression step has not been proposed in any previous contribution about nonnegative greedy schemes targeting ℓ_0 minimization¹.

According to Lemma 6, the restriction to a descent selection rule implies $\mathbf{r}_S^+ = \mathbf{r}_S$ at any iteration, which justifies that we have dropped the ‘+’ sign in Algorithm 1. This simplification is adopted in the rest of the paper. Moreover, the termination rule $\max_i \mathbf{h}_i^t \mathbf{r}_S \leq 0$ is used at Line 2 since in this case, there are no descending atoms anymore, so the residual cannot decrease by selection of a new atom.

D. Examples of descent selection rules

In what follows, selection rules are denoted $\mathcal{S}(S)$, the dependence on \mathbf{y} and H being implicit. Let us introduce three important selection rules by their distinct ways of picking an index in \mathcal{D}_S when $\mathcal{D}_S \neq \emptyset$.

- NNOMP rule [30]–[32], [38]:

$$\mathcal{S}_1(S) \in \arg \max_{i \notin S} \mathbf{h}_i^t \mathbf{r}_S \quad (12)$$

- Suboptimal NNOLS (SNNOLS, [33]) rule:

$$\mathcal{S}_2(S) \in \arg \max_{i \notin S} \tilde{\mathbf{g}}_i^t \mathbf{r}_S \quad (13)$$

- NNOLS rule [33]:

$$\mathcal{S}_3(S) \in \arg \min_{i \notin S} \|\mathbf{r}_{S \cup \{i\}}^+\|^2 \quad (14)$$

(14) is a descent selection rule by definition. The fact that (12) and (13) are descent selection rules is deduced from recursive application of Proposition 1 and Lemma 6. As regards the latter rules, \mathbf{r}_S is the current residual vector, *i.e.*, a readily

available quantity. On the other hand, projected atoms $\tilde{\mathbf{g}}_i$ enter rule (13), so we can expect the computing cost of (13) to be larger than that of (12). Rule (14) needs the solution of NNLS problems on supports $S \cup \{i\}$, which is even more demanding.

Note that [39] introduced another version of non-negative OMP named NN-OMP in which the selection rule is that of OMP. Clearly, this version does not rely on a descent selection rule. On the other hand, it is unclear whether the FNNOMP and FNNOLS algorithms proposed in [32], [33] rely on a descent selection rule. They will therefore not be further analyzed.

The following proposition makes it possible to compare the three rules (12)–(14) by relating them to the minimization of a residual norm.

Proposition 2 *Rules (12)–(14) are equivalent to*

$$\mathcal{S}_j(S) \in \arg \min_{i \in \mathcal{D}_S} \mu_j(S, i), \quad (15)$$

where μ_j are specific to each rule:

$$\mu_1(S, i) = \min_v \|\mathbf{y} - H_S \mathbf{x}_S - \mathbf{h}_i v\|^2, \quad (16)$$

$$\mu_2(S, i) = \min_{\mathbf{u}, v} \|\mathbf{y} - H_S \mathbf{u} - \mathbf{h}_i v\|^2, \quad (17)$$

$$\mu_3(S, i) = \min_{\mathbf{u} \geq \mathbf{0}, v \geq 0} \|\mathbf{y} - H_S \mathbf{u} - \mathbf{h}_i v\|^2. \quad (18)$$

Proof: Let us first emphasize that because (12)–(14) are descent selection rules, $i \notin S$ in (12)–(14) can be replaced by $i \in \mathcal{D}_S$. Clearly, (15) with (18) simply duplicate (14). Using identity (1) with $\mathbf{r} = \mathbf{r}_S$ and $\mathbf{h} = \mathbf{h}_i$, (12) can be rewritten as the argmin over $i \in \mathcal{D}_S$ of

$$\min_{v \geq 0} \|\mathbf{y} - H_S \mathbf{x}_S - \mathbf{h}_i v\|^2. \quad (19)$$

Likewise, applying (1) with $\mathbf{r} = \mathbf{r}_S$ and $\mathbf{h} = \tilde{\mathbf{g}}_i$ and using the fact that $\tilde{\mathbf{g}}_i$ is the normalized version of $\tilde{\mathbf{h}}_i$, (13) can be rewritten as the argmin over $i \in \mathcal{D}_S$ of

$$\min_{v \geq 0} \|\mathbf{r}_S - \tilde{\mathbf{h}}_i v\|^2 = \min_{\mathbf{u}, v \geq 0} \|\mathbf{y} - H_S \mathbf{u} - \mathbf{h}_i v\|^2. \quad (20)$$

(20) follows from explicit minimization with respect to \mathbf{u} and from the fact that \mathbf{r}_S and $\tilde{\mathbf{h}}_i$ read as the orthogonal projections of \mathbf{y} and \mathbf{h}_i onto $(\text{span}(H_S))^\perp$. Finally, the positivity constraint on v in (19)–(20) turns out to be inactive. Indeed, according to Proposition 1, we have that $\mathbf{h}_i^t \mathbf{r}_S > 0$ and $\tilde{\mathbf{h}}_i^t \mathbf{r}_S > 0$ for $i \in \mathcal{D}_S$. It is easy to see from (1) that the minimum error norm in (19)–(20) is then strictly lower than $\|\mathbf{r}_S\|$. Hence, the optimal variable v in (19)–(20) is nonzero.

Finally, (15) with (19) and (15) with (20) rewrite as (15) with (16) and (15) with (17), respectively. ■

For $j \in \{1, 2, 3\}$, an alternate way of viewing the descending character of each rule consists in noticing that $\|\mathbf{r}_{S \cup \{i\}}\|^2 \leq \mu_j(S, i) < \|\mathbf{r}_S\|^2$ for all $i \in \mathcal{D}_S$. It is interesting to see that, by restricting the selection to the set of descending atoms, the selection rules of NNOMP and SNNOLS rely on the same criteria (16)–(17) as those of OMP and OLS, respectively.

¹However, this kind of operation was introduced in [36], [37].

E. On the usefulness of support compression

As already mentioned, rules (12)-(14) were introduced in existing works devoted to problem (ℓ_0+) without support compression. Let us analyze the impact of removing the support compression step at Line 6 in Algorithm 1.

- The subset S then becomes an *extended* support, with $\mathbf{x} = \hat{\mathbf{x}}_S^+$ and $\text{supp}(\mathbf{x}) \subset S$. As a consequence, when $\text{supp}(\mathbf{x}) \neq S$, the current residual $\mathbf{r}_S^+ = \mathbf{y} - H\mathbf{x}$ does not identify to \mathbf{r}_S anymore since \mathbf{x} does not read as a ULS solution related to S .
- In the case of NNOMP and NNOLS (selection rules (12) and (14), with \mathbf{r}_S replaced by the current residual \mathbf{r}_S^+), Proposition 1 can still be applied, since condition $\mathbf{h}_i^t \mathbf{r}_S^+ > 0$ still characterizes descending atoms without support compression. As for SNNOLS, we cannot guarantee that $S(\mathbf{y}, H, S)$ defined in (13) is a descent selection rule when S is not positive, since the equivalence (9) in Proposition 1 does not hold anymore.
- Finally, for any NNOG algorithm, Lemma 5 is no more valid when S is not a positive support, so H_S might become rank deficient at any subsequent iteration, $\hat{\mathbf{x}}_S^+$ not being well defined anymore.

We consider the last point as the main justification to apply support compression. An open question would be to determine whether the assumptions of Lemma 5 can be weakened, the positivity of S being replaced by the full rankness of H_S . However, we lack concrete elements in this sense. Therefore, at this stage, NNOG with compression present formal guarantees that NNOG without compression do not.

III. ACTIVE-SET NNLS ALGORITHMS AND RECURSIVITY

Let us consider the NNLS problem related to support S in (2). When matrix H_S is full column rank, it is a special case of a strictly convex quadratic program. The successive resolution of possibly many NNLS problems for nested supports S is a basic ingredient of NNOG algorithms. NNLS problems do not admit closed-form solutions in general. However, *active-set* NNLS (AS-NNLS) algorithms are well-known schemes that solve NNLS problems in a finite number of iterations [22], [40]. Moreover, the computation of NNLS solutions can be efficiently accelerated by exploiting that warm start solutions are available and using classical recursive updates of matrix factorizations.

This section first contains a short reminder on active-set NNLS algorithms and on their efficient implementation. Then, we show how to preserve computational efficiency when NNLS subproblems are solved within an NNOG scheme. We also analyze the similarity between the structures of AS-NNLS algorithm and NNOMP, already pointed out in [27], [41].

A. Fast active-set algorithms

Among many numerical methods for solving (2), AS-NNLS algorithms correspond to greedy schemes, since the solution is found by incremental modifications of its support $V \subset S$, the *active set* being defined as the complementary set $S \setminus V$ (by reference to the active constraints). Such an incremental

Algorithm 2: Active-set algorithm to solve the NNLS problem related to T , starting from a positive support S .

input : \mathbf{y} , H , target set T , initial support $S \subset T$, $\hat{\mathbf{x}}_S$
output: $V := \text{supp}(\hat{\mathbf{x}}_T^+)$, $\hat{\mathbf{x}}_T^+ := \hat{\mathbf{x}}_V$

```

1  $\mathbf{x} \leftarrow \hat{\mathbf{x}}_S$ ;  $V \leftarrow S$ ;
2 while  $\max\{\mathbf{h}_i^t \mathbf{r}_V, i \in T \setminus V\} > 0$  do
3    $\ell^+ \leftarrow \arg \max\{\mathbf{h}_i^t \mathbf{r}_V, i \in T \setminus V\}$ ;
4    $V \leftarrow V \cup \{\ell^+\}$ ;
5   Update  $\hat{\mathbf{x}}_V$  (call Algorithm 4);
6   while  $\min(\hat{\mathbf{x}}_V) < 0$  do
7      $\ell^- \in \arg \min_{\{i \in V: \hat{\mathbf{x}}_V(i) < 0\}} \mathbf{x}(i)/(\mathbf{x}(i) - \hat{\mathbf{x}}_V(i))$ ;
8      $\alpha \leftarrow \mathbf{x}(\ell^-)/(\mathbf{x}(\ell^-) - \hat{\mathbf{x}}_V(\ell^-))$ ;
9      $\mathbf{x} \leftarrow \mathbf{x} + \alpha(\hat{\mathbf{x}}_V - \mathbf{x})$ ;
10     $V \leftarrow V \setminus \{\ell^-\}$ ;
11    Update  $\hat{\mathbf{x}}_V$  (call Algorithm 4);
12  end
13   $\mathbf{x} \leftarrow \hat{\mathbf{x}}_V$ ;
14   $\mathbf{r}_V \leftarrow \mathbf{y} - H\mathbf{x}$ ;
15 end

```

structure is an essential element to obtain practically fast implementations [21], [22], [35]. Whenever V is modified, the corresponding ULS solution $\hat{\mathbf{x}}_V$ is updated. Each iteration requires at least one ULS solution of the *selection* type, *i.e.*, $H_{V \cup \{\ell\}}^\dagger \mathbf{y}$. Some iterations also need to compute *deselection* type ULS solutions $H_{V \setminus \{\ell\}}^\dagger \mathbf{y}$. For the sake of computational efficiency, it is crucial to compute both types of solutions recursively, given that $\hat{\mathbf{x}}_V$ is already available. In this respect, the situation is identical to that of bi-directional greedy algorithms in the unconstrained case (*e.g.*, Bayesian OMP [42], SBR [43] or FoBa [44]), also called stepwise algorithms [3, Chapter 3], for which selections and deselections are implemented recursively. Fast recursive implementations require specific computations and storage of quantities related to the Gram matrix $G_V = H_V^\dagger H_V$. Efficient selection steps can be obtained using QR or Cholesky matrix factorizations applied to G_V , or the Matrix Inversion Lemma (MIL), with roughly the same cost [7]. MIL consists in storing and updating the inverse of G_V . It appears to be the cheapest concerning deselections, so our default choice here is based on the MIL.

B. Warm start and full recursivity

Since NNOG algorithms are based on iterated calls to an active-set scheme, we must carefully consider the way we initialize the latter. The starting point of AS-NNLS is usually defined as the zero solution, associated with the empty support. This is the case in the Lawson-Hanson algorithm, which is the reference AS-NNLS scheme [22]. In [35, Chap. 16], Nocedal and Wright proposed an AS-NNLS algorithm with any feasible vector as a possible starting point. Algorithm 2 is a generalized version of the Lawson-Hanson algorithm to address the NNLS problem related to an augmented target set $T \supset S$, where the initial point is not restricted to be the zero vector (the rest of the scheme being unaltered). Specifically, the initial point is

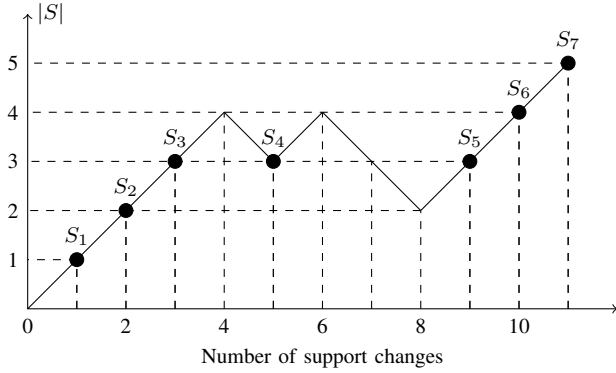


Fig. 1. Step-by-step illustration of NNOG after each change of support. Bullets represent supports corresponding to the first seven NNOG iterates whereas other intermediate supports found during the calls to AS-NNLS are represented without bullets.

set as the ULS solution $\hat{x}_S \geq \mathbf{0}$, S being a positive support. For this specific initial point, it can easily be checked that Algorithm 2 identifies with Nocedal and Wright’s scheme.

At any iteration of an NNOG algorithm, the current solution \hat{x}_S^+ can be used as the initial point to compute $\hat{x}_{S \cup \{\ell\}}^+$ using Algorithm 2 with $T = S \cup \{\ell\}$. In this way, the initial support is $V = S$, so $T \setminus V = \{\ell\}$ and the first iteration of AS-NNLS begins by selecting ℓ . In practice, the NNLS algorithm is expected to terminate after a single iteration (if no deselection is performed in the second part of it), or otherwise after very few iterations. Algorithm 3 is a global view of the active-set implementation of NNOG obtained by integrating the calls to the AS-NNLS solver (Algorithm 2) in the NNOG framework (Algorithm 1). Whenever a new atom ℓ is selected, AS-NNLS starts by computing $\hat{x}_{S \cup \{\ell\}}$. If $\hat{x}_{S \cup \{\ell\}} \geq \mathbf{0}$, then $\hat{x}_{S \cup \{\ell\}}^+ = \hat{x}_{S \cup \{\ell\}}$ and AS-NNLS stops after one support change. Otherwise, AS-NNLS deselects at least one atom from $S \cup \{\ell\}$ (Algorithm 2, Line 7) and then alternates between atom selections and series of deselections. This mechanism is illustrated by a simple example in the next subsection.

A reduced number of iterations is obtained because we use Algorithm 2 with a warm start. To further improve the overall numerical efficiency, we also need to reduce the computing cost of the ULS solution at Lines 5 and 11. These are selection and deselection-type ULS problems, respectively, that can be solved recursively provided that the inverse of the Gram matrix G_V be both an input and an output quantity of the NNLS algorithm. In Appendix B, Algorithm 4 is a pseudo-code to implement ULS updates in the forward ($V \leftarrow V \cup \{\ell\}$) and backward scenarios ($V \leftarrow V \setminus \{\ell\}$). This implementation enables us to obtain a fully recursive implementation of AS-NNLS as well by updating the Gram matrix inverse at each call to ULS in Algorithm 2 (Lines 4-5 and 10-11).

C. Step-by-step illustrative example of NNOG

Fig. 1 displays a schematic step-by-step illustration of NNOG. NNOG iterates S_k are represented with bullets. NNOG starts with the empty support. In this example, it turns out that during the first three iterations, NNOG yields a positive support $S_{k-1} \cup \{\ell\}$ ($\hat{x}_{S_{k-1} \cup \{\ell\}} \geq \mathbf{0}$), hence $S_k \leftarrow$

Algorithm 3: NNOG with active-set implementation.

input : \mathbf{y}, H, K
output: $\mathbf{x} := \hat{x}_S^+$ (with S a positive support)

- 1 $\mathbf{x} \leftarrow \mathbf{0}$; $S \leftarrow \emptyset$; $\mathbf{r}_S \leftarrow \mathbf{y}$;
- 2 **while** $|S| < K$ and $\max_{i \in \bar{S}} \mathbf{h}_i^t \mathbf{r}_S > 0$ **do**
- 3 Select an index $\ell \in \bar{S}$ by a selection rule $\mathcal{S}(\mathbf{y}, H, S)$;
- 4 Call $[C, \mathbf{x}] = \text{AS_NNLS}(\mathbf{y}, H, S \cup \{\ell\}, S, \mathbf{x})$;
- 5 $S \leftarrow C$;
- 6 $\mathbf{r}_S = \mathbf{y} - H\mathbf{x}$;
- 7 **end**

$S_{k-1} \cup \{\ell\}$. Therefore, $S_1 \subset S_2 \subset S_3$ are of cardinalities 1, 2 and 3, respectively. At iteration 4, $S_3 \cup \{\ell\}$ is not positive so AS-NNLS performs two support changes, namely the selection of ℓ and a deselection. The next NNOG iterate reads $S_4 \leftarrow S_3 \cup \{\ell\} \setminus \{\ell_1\}$. Iteration 5 is more tricky (and unlikely). Here again, $S_4 \cup \{\ell\}$ is not a positive support. The first deselection does not yield a positive support either, so another deselection is carried out, yielding $V \leftarrow S_4 \cup \{\ell\} \setminus \{\ell_1, \ell_2\}$, V being a positive support. However, the stopping condition of AS-NNLS (Line 2 of Algorithm 2, with $T \leftarrow S_4 \cup \{\ell\}$) is not met since \mathbf{h}_{ℓ_1} is a descending atom. Therefore, ℓ_1 is reselected: $V \leftarrow V \cup \{\ell_1\}$. Since V is now a positive support and there are no descending atoms anymore in $T \setminus V$, the convergence of AS-NNLS is reached. In the last two NNOG iterations, $S_{k-1} \cup \{\ell\}$ are positive supports, so single selection moves are done within AS-NNLS.

D. Connection between AS-NNLS and NNOMP

In Algorithm 3, the structure of the NNOG main iteration consists of one atom selection followed by a variable number of updates (selections or deselections) of the support S when the ULS solution \hat{x}_S has some negative entries. The first of these updates is necessarily an atom deselection. Interestingly, [27] and [41] pointed out that the AS-NNLS algorithm initialized with the zero vector, has a structure similar to NNOMP: each main NNLS iteration consists of an atom selection followed by a series of atom deselections (respectively, Line 3 and Lines 6-12 of Algorithm 2). This connection led Peharz and Pernkopf to propose AS-NNLS with an early stopping rule $|S| = K$ as a stand-alone NNOG algorithm (called Sparse NNLS in [27]). Given the strong similarity of Sparse NNLS and NNOMP, an interesting question is to determine whether their iterates always coincide. It turns out that this is not always true. However, as long as Sparse NNLS performs only simple support changes, both algorithms yield the same iterates, according to the following proposition.

Proposition 3 *Let us consider Sparse NNLS, i.e., Algorithm 2 initialized with the empty support together with the early stopping rule $|S| = K$. When each iteration produces at most one removal in Lines 6–12, the output of Sparse NNLS identifies with that of NNOMP (i.e., Algorithm 1 with rule S_1 in (12)).*

Proof: See Appendix A-B. ■

IV. ACCELERATION OF NNOG ALGORITHMS

The NNOMP, SNNOLS and NNOLS selection rules (12)-(14) all read as the optimization of a criterion with respect to the candidate index $i \notin S$. Since all three ensure the selection of an atom in \mathcal{D}_S , an obvious acceleration of SNNOLS and NNOLS consists of pre-selecting the descending atoms according to $\mathbf{h}_i^\dagger \mathbf{r}_S > 0$ (see Proposition 1) to carry out the optimization tasks (13) and (14) over $i \in \mathcal{D}_S$ only. This operation is referred to as *type-I pruning* of the dictionary². Testing the sign of $\mathbf{h}_i^\dagger \mathbf{r}_S$ requires $\mathcal{O}(m)$ operations. This is much less than the (recursive) computation of the criteria $\hat{\mathbf{g}}_i^\dagger \mathbf{r}_S$ and $\|\mathbf{r}_{S \cup \{i}^+}\|^2$, which costs at least $\mathcal{O}(|S|^2 + km)$ operations.

A. Atom selection

The atom selection step of both NNOMP and SNNOLS can be efficiently implemented using vectorized computations, which allow one to benefit from the inherent parallelism of SIMD processors. The NNOMP case is the simplest, since $H_S^\dagger \mathbf{r}_S$ directly yields the expected set of inner products $\mathbf{h}_i^\dagger \mathbf{r}_S$. Indeed, the selection steps of OMP and NNOMP are both based on the minimization of (16), so they share the same possibilities of parallel computations. Likewise, OLS and SNNOLS being both based on the minimization of (17), they share the same possibilities of parallel implementation. In coherence with our choice of recursive implementation of ULS solutions (see Appendix B), we have adopted a MIL based solution to solve (17) in a vectorized way (see Matlab code in supplementary material).

In contrast, the NNOLS selection rule (14) does not lend itself to fully vectorized computations, since we have as many NNLS subproblems to solve as candidate atoms, with a variable number of subiterations of AS-NNLS for each of them. Fortunately, the structure of the NNOLS rule can be made closer to that of SNNOLS, with the benefit of vectorized computations for the largest part. The key point is that for each candidate atom, the initial step of AS-NNLS corresponding to Lines 1-5 of Algorithm 2 yields the same unconstrained minimizer $\hat{\mathbf{x}}_{S \cup \{i}$ as the one involved in the SNNOLS rule (17). Hence, these vectors can be obtained using vectorized computations that exactly identify to the main step of SNNOLS atom selection. The extra computations induced by NNOLS reside in the additional AS-NNLS iterations required for each non-positive support $S \cup \{i\}$. According to our empirical tests, only a small minority of atoms needs more than one iteration. Moreover, a lot of them can be pruned without actually performing any additional AS-NNLS iterations. Let us denote by e_{opt} the smallest residual error produced by atoms i for which $S \cup \{i\}$ is a positive support. Since $\|\mathbf{r}_{S \cup \{i}^+}\| \geq \|\mathbf{r}_{S \cup \{i}\|}$ for all i , one can immediately ignore the atoms i for which $\|\mathbf{r}_{S \cup \{i}\| \geq e_{\text{opt}}$. Moreover, since we sequentially visit the remaining ones, the threshold e_{opt} can be gradually lowered whenever a new atom is found to improve the smallest current residual error. This operation called *type-II pruning* is specific to NNOLS implementation.

²Contrary to screening techniques, see *e.g.*, [45], note that the atoms indexed by $i \notin \mathcal{D}_S$ are pruned from the dictionary for the current NNOG iteration only, but they are considered again in further iterations.

B. Coefficient update

Once an atom i is selected, NNOG algorithms need to update the coefficients by solving an NNLS subproblem (Algorithm 1, Line 5). However, in the case of NNOLS, the update is already being performed in the selection step. In the case of NNOMP, one needs to call AS-NNLS (Algorithm 2) from the initial set S to the target set $S \cup \{\ell\}$. For SNNOLS, a call to Algorithm 2 is needed only when $S \cup \{\ell\}$ is not positive.

C. Software

An open source Matlab implementation of the acceleration strategies detailed above is provided on CodeOcean [46] (<https://doi.org/10.24433/CO.2445681.v1>). This software contains a fully recursive, vectorized version of NNOMP, NNOLS and SNNOLS together with test programs.

V. SIMULATED SPARSE DECONVOLUTION

A. Data generation

In order to assess the behavior of NNOG algorithms, we consider a convolution problem with a Gaussian kernel \mathbf{h} of standard deviation σ . \mathbf{h} is approximated by a finite impulse response of length 6σ by thresholding the smallest values. The corresponding normalized dictionary H is a Toeplitz matrix. Simulated data are generated according to $\mathbf{y} = H\mathbf{x}^* + \mathbf{n}$ where \mathbf{x}^* and \mathbf{n} stand for the ground truth and white Gaussian noise, respectively. The support S^* of \mathbf{x}^* , of cardinality K , is randomly generated with a uniform distribution whereas the non-zero coefficients of \mathbf{x}^* are either set to a positive constant or randomly generated from an i.i.d. folded normal distribution. The signal-to-noise ratio is defined by $\text{SNR} = 10 \log_{10}(P_{H\mathbf{x}^*}/P_{\mathbf{n}})$ where $P_{H\mathbf{x}^*} = \|H\mathbf{x}^*\|^2/m$ is the average power of the noise-free data and $P_{\mathbf{n}}$ is the noise variance. The results presented below are achieved on a macOS X system with 16 GB RAM and Intel Core i7 processor at 2.7 GHz.

Hereafter, we first evaluate the computational cost of NNOG algorithms with the proposed implementations. In subsections V-B and V-C, the computational complexity of NNOMP, SNNOLS and NNOLS is compared with that of OMP and OLS. Subsection V-D contains a comparison of NNOG algorithms with other non-negative sparse algorithms proposed in the literature including FNNOMP, FNNOLS [32], [33], NN-CoSaMP, and NN-SP [30]. Algorithms are evaluated in terms of sparse reconstruction accuracy and reconstruction time.

B. Validation of NNOG accelerations

We generate a convolution dictionary of size 1200×1140 with $\sigma = 10$. For each $K \in \{20, 40, 60, 80\}$, 200 trials are carried out in which the support S^* is generated. The non-zero coefficients of \mathbf{x}^* are set to 1 and the SNR is set to 30 dB. OMP, OLS, NNOMP, SNNOLS and NNOLS are run until the support cardinality equals K . We consider the recursive implementation of OMP and OLS utilizing the matrix inversion lemma. We chose this implementation to make it consistent with our proposed NNOG implementation. Note that NNOG algorithms may need more than K iterations

TABLE I
ACCELERATION GAIN OF NNOG ALGORITHMS FOR A SPARSE DECONVOLUTION PROBLEM WITH GAUSSIAN KERNEL ($\sigma = 10$). MEAN OVER 200 TRIALS. THE DICTIONARY SIZE IS 1200×1140 .

K	Acceleration gain			Non-negativity loss			Iterations		
	NNOMP	SNNOLS	NNOLS	NNOMP	SNNOLS	NNOLS	NNOMP	SNNOLS	NNOLS
20	2.7	1.2	9	1.2	1.0	1.5	20	22	21
40	4.5	2.1	59	1.2	1.2	1.9	41	51	43
60	6.4	3.4	120	1.1	1.2	2.1	65	97	73
80	8.5	4.0	128	1.1	1.3	4.1	95	152	121

because of support compression. The following quantities are computed and averaged out over the number of trials:

- *Acceleration gain*: CPU time ratio between the canonical and accelerated implementations of NNOG algorithms.
- *Non-negativity loss*: CPU time ratio between accelerated implementation of NNOG algorithms and the corresponding unconstrained versions.
- *Iterations*: average number of iterations of NNOG algorithms needed to yield a support of cardinality K . This number is larger than K when support compression occurs.

Before going further, let us clarify that the so-called ‘‘canonical implementations’’ refer to the following settings. The AS-NNLS algorithm is called from scratch from the initial zero solution. Moreover, obvious accelerations are taken into account such as type-I pruning, computation of the ULS solutions for candidate supports $S \cup \{i\}$ so as to avoid calling AS-NNLS when the augmented support is positive, and recursive computation of the AS-NNLS iterates. On the other hand, advanced accelerations such as type-II pruning and warm start initialization are not included. Support compression is included in such a way that both canonical and accelerated versions of a given NNOG algorithm yield the same iterates.

The scores can be found in Tab. I. Accelerated implementations yield a gain in time that increases with K . Since NNOLS needs to solve many NNLS subproblems per iteration, the gain is much larger. The time gain is intermediate for NNOMP. We further notice that using accelerated implementations, the cost of NNOMP and SNNOLS becomes comparable with that of OMP and OLS, respectively, the non-negativity loss remaining below 1.3. Regarding NNOLS vs OLS, the non-negativity loss remains lower than 5 in these simulations. At last, the fact that the number of iterations is often larger than K reveals that support compression is happening quite often.

C. Computation burden of SNNOLS and NNOLS

The purpose of this subsection is to examine thoroughly the computation burden of SNNOLS and NNOLS so as to evaluate the accelerations proposed in Section IV. Four indicators are computed at each iteration:

- 1) $\rho_{\downarrow} = |\mathcal{D}_S|/|\bar{S}|$: rate of descending atoms. The rate of discarded atoms after type-I pruning reads $1 - \rho_{\downarrow}$.
- 2) ρ_{cand}^+ : rate of descending candidate atoms for which $S \cup \{i\}$ is a positive support.
- 3) ρ_{II} : rate of candidate atoms discarded by type-II pruning among all atoms i for which $S \cup \{i\}$ is not a positive support.

- 4) ρ_{sel}^+ : rate of selected atoms yielding a positive support $S \cup \{\ell\}$.

The computational cost of an NNOG iteration is closely related to the values of these ratios. Indeed, large scores indicate that the cost of testing candidate atoms is dramatically reduced. Specifically, ρ_{\downarrow} is the rate of candidates that are truly considered in the selection rule (15) of NNOG algorithms. Both ratios ρ_{cand}^+ and ρ_{II} quantify the computational burden of the NNOLS selection step: large values of ρ_{cand}^+ and ρ_{II} indicate that AS-NNLS has to be run for a few candidate atoms only, since other atoms are either pruned or yield a non-negative ULS solution. ρ_{sel}^+ is defined similar to ρ_{cand}^+ . However, ρ_{sel}^+ applies to the selected atoms, which brings information on the computational burden of the coefficient update stage of SNNOLS.

Using the same setting as in the previous subsection ($H \in \mathbb{R}^{1200 \times 1140}$ and SNR = 30 dB) with $K = 80$, the computational burden indicators of SNNOLS and NNOLS are shown in Fig. 2. It is noticeable that the number of iterations L required to reach a support of cardinality K is larger than K because of support compression. Specifically, the histograms of Fig. 2(b) show that on average, L is larger than K for NNOLS and even larger for SNNOLS (the average values are given in the last columns of Tab. I). This is consistent with the fact that the NNOLS selection rule is more involved but more reliable. Moreover, the standard deviation of L corresponding to the histograms of Fig. 2(b) is 17 and 12 for SNNOLS and NNOLS, respectively, which indicates that the size of the support found after k iterations may significantly vary between trials. In order to get meaningful evaluations, we choose to compute the average values of each indicator over the last t iterations, with $t \in \{0, \dots, K - 1\}$. When $t = 0$, only the last iteration is taken into account, so the current support is of size K . For larger values of t , the supports found during the last t iterations have varying sizes but the averaging operation remains meaningful, especially for the last iterations, which are the most costly. The curves displaying the average of each indicator over 200 trials and over the last t iterations are shown in Fig. 2(a). One can observe from the curve $(1 - \rho_{\downarrow})$ that the rate of non-descending atoms gradually increases for both SNNOLS and NNOLS. Therefore, type-I pruning is more effective at late iterations, where half of the atoms are discarded.

The efficiency of NNOLS implementation is measured by ρ_{cand}^+ and ρ_{II} in Fig. 2(a). Large values of ρ_{cand}^+ and ρ_{II} are obtained, which indicates that the computing cost of NNOLS is dramatically reduced. The score of ρ_{cand}^+ always remains above 70%. This means that the computation of the NNLS

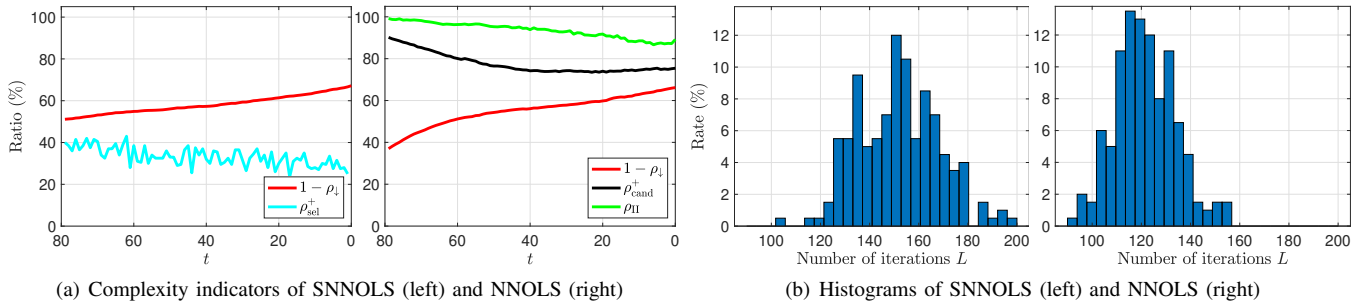


Fig. 2. Complexity analysis of SNNOLS and NNOLS for a sparse deconvolution problem with Gaussian kernel ($\sigma = 10$) and a SNR of 30 dB. The dictionary size is 1200×1140 , $K = 80$ and 200 trials are performed. (a) Evolution of complexity factors during the last t iterations (from $L - t$ to L). ($1 - \rho_I$): atoms discarded by type-I pruning; ρ_{sel}^+ : positive supports found by SNNOLS; ρ_{cand}^+ : candidate atoms yielding a positive support in NNOLS; ρ_{II} : atoms discarded by type-II pruning in NNOLS. (b) Histogram of the average number of iterations L required to reach a support of cardinality $K = 80$.

solutions reduces to a single ULS update for more than 70% of candidate atoms. Moreover, many of the remaining atoms can be discarded due to type-II pruning. For instance, in the last iterations, around 25% of descending atoms do not yield a positive support and type-II pruning eliminates 90% of them. As a result, only 3% of descending atoms require to proceed with a complete NNLS solving. Let us stress that the extra cost of NNOLS as compared with SNNOLS essentially comes from the number of atoms related to a complete NNLS solving, so it directly depends on the efficiency of type-II pruning.

Regarding SNNOLS, the ratio ρ_{sel}^+ remains above 20%, which indicates that the computation of the SNNOLS iterate reduces to a single ULS update for at least 20% of the trials. Generally speaking, the slight decrease of ρ_{sel}^+ , ρ_{cand}^+ and ρ_{II} shows that both SNNOLS and NNOLS call AS-NNLS more often at the late iterations. A possible reason for this behavior might be that the average correlation between selected atoms increases with the cardinality of the support.

D. Performance comparison between NNOG and competing algorithms

NNOG algorithms NNOMP, SNNOLS and NNOLS, integrating support compression and the proposed acceleration strategies, are compared with the corresponding unconstrained versions OMP and OLS, where a post-processing is applied to obtain non-negative coefficients. NNOG algorithms are also compared to the fast versions FNNOMP and FNNOLS [32], [33], as well as the non-negative extensions of CoSaMP and Subspace Pursuit proposed in [30]. All comparisons are based on reconstruction accuracy and CPU time. Reconstruction accuracy is quantified by three factors:

- *Support recovery*: ratio of true positives to K ;
- *Coefficient inaccuracy*: relative error for the recovered coefficients ($\|\mathbf{x} - \mathbf{x}^*\| / \|\mathbf{x}^*\|$);
- *Residual norm*: Euclidean norm of the data residual ($\|\mathbf{y} - H\mathbf{x}\|$).

The curves shown in Fig. 3 are obtained for a convolution dictionary of size 2400×2388 , with $\sigma = 2$ and for various settings of $K \in [2, 80]$. For each K , 400 trials are carried out in which the support S^* is drawn according to the uniform distribution and the non-zero coefficients of \mathbf{x}^* are drawn from

a Gamma distribution with shape and scale parameters set to 1 and 2, respectively. Note that for increasing K , the density of spikes increases, and hence the difficulty of the problem. Working at a given level of SNR would have even increased the difficulty at large values of K , so we have preferred to keep the noise level P_n constant; it is set to $P_n = 10^{-2}$. In this simulation condition, it is noticeable in Fig. 3(b,f) that the coefficient inaccuracy $\|\mathbf{x} - \mathbf{x}^*\|$ becomes somewhat large when the spike density is high. For easier problems (lower value of σ and lower density of spikes), many algorithms would provide good results, so it would be difficult to observe differences in terms of estimation quality.

1) *Comparison with unconstrained greedy algorithms*: We first perform a comparison of the NNOG algorithms with (unconstrained) OMP and OLS to demonstrate the usefulness of working with NNOG algorithms. In order to obtain non-negative coefficients from the OMP (respectively, OLS) outputs, a simple possibility is to apply a post-processing step using the solution support yielded after K iterations of OMP (resp., OLS). This post-processing consists of solving the NNLS problem related to the support found by OMP/OLS, with possible annealing of coefficients when non-negativity constraints are activated. The resulting schemes are denoted by OMP+ and OLS+.

Fig. 3(a-d) includes a comparison of NNOMP and NNOLS with OMP, OLS, OMP+ and OLS+. Their support recovery and coefficient accuracy performances are weaker than those of NNOG algorithms. In Fig. 3(a), the scores of OMP+ (OLS+) are identical to those of OMP (OLS). This is because the support recovery measure considers the rate of true positives, which is unchanged since the post-processing step of OMP+ (OLS+) essentially removes false positives. On the contrary, the coefficient inaccuracy ratio is improved due to the latter removal. In Fig. 3(c-d), the time and error scores of OMP (resp., OLS) are both lower than those of NNOMP (resp., SNNOLS and NNOLS). This is not a surprise since unconstrained algorithms are simpler, and the obtained solutions are expected to reach lower values of the residual since they do not satisfy the non-negativity constraint. The simple post-processing in OMP+/OLS+ does not allow to yield a residual norm as small as the one obtained using NNOG. This shows that NNOG algorithms do improve the

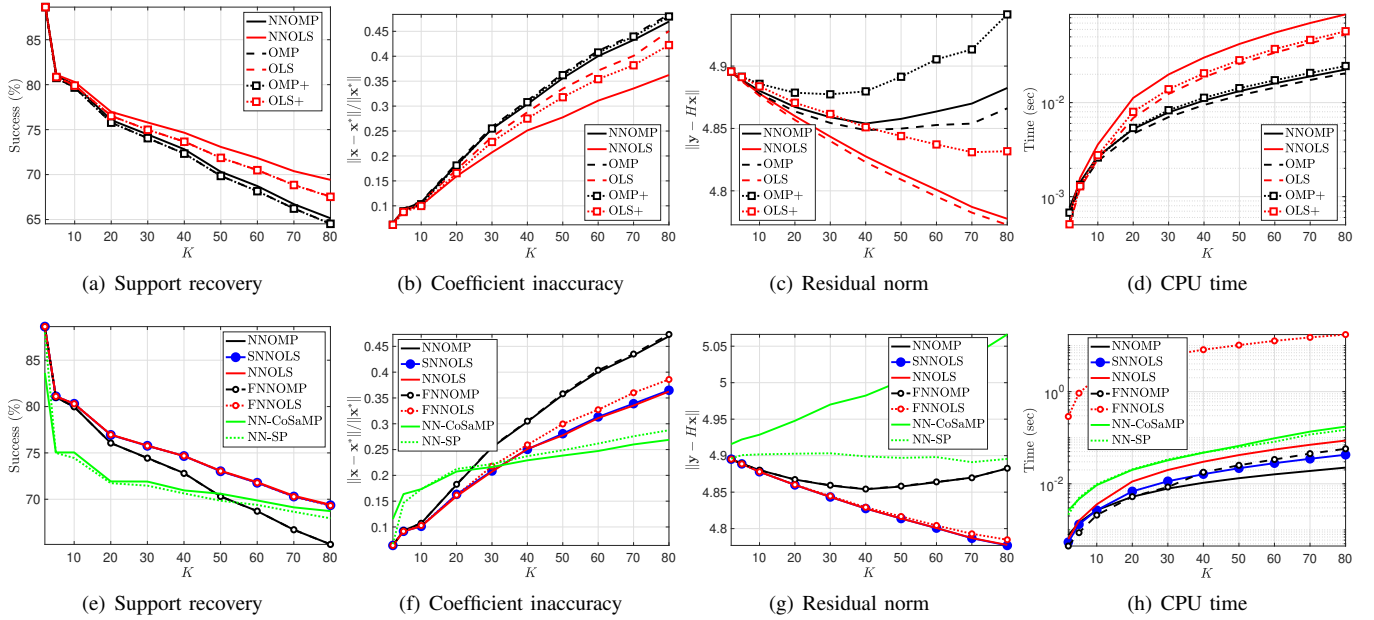


Fig. 3. Comparison of NNOG algorithms with unconstrained greedy algorithms and other non-negative sparse solvers. Performances are averaged over 400 trials in a simulated sparse deconvolution problem with Gaussian kernel with $\sigma = 2$, $P_n = 10^{-2}$, and a dictionary of size 2400×2388 . (a)–(d) NNOG vs unconstrained greedy algorithms; (e)–(h) NNOG vs other sparse non-negative algorithms.

performance of (unconstrained) OMP and OLS.

2) *Comparison with approximate fast non-negative greedy algorithms:* Fig. 3(e-h) includes a comparison with the FNNOMP and FNNOLS implementations³ of Yaghoobi *et al.* [32], [33]. FNNOLS turns out to be much slower and less accurate than our fast implementations of SNNOLS and NNOLS. On the other hand, the statistical performances of NNOMP and FNNOMP are very close both in terms of computational time and accuracy. However, FNNOMP may return some negative coefficients in the reconstructed sparse vector. FNNOMP was introduced as a fast but approximate version of the canonical NNOMP algorithm by Bruckstein *et al.* In contrast, our fast implementation of NNOMP is exact and the non-negativity constraint is always satisfied.

3) *Comparison with non-negative versions of CoSaMP and SP:* In Fig. 3(e-h), NNOG implementations are also compared with NN-CoSaMP and NN-SP [30]. NN-HTP [30] yields the weakest performance, so it is not further considered here.

The principle of NN-CoSaMP and NN-SP is to maintain an estimated support of size K . To do so, three kinds of operations are performed at each iteration including support merging, NNLS estimation of coefficients and support pruning. In NN-CoSaMP, the current support of size K is merged with a support of size $2K$, and an NNLS problem is solved with this augmented support. Then, thresholding is performed by keeping the K coefficients having the largest magnitudes. The structure of NN-SP is similar, but two NNLS problems (of size roughly $2K$ and K) are solved before and after thresholding. The NN-SP output thus contains at most K nonzero elements depending on the activation of nonnegativity constraints.

In Fig. 3(e-h), the NNOG algorithms yield competitive accuracy performance compared to NN-SP and NN-CoSaMP.

³kindly provided by Dr. Mehrdad Yaghoobi.

In particular, NNOG algorithms have a better ability to find a low value of the residual norm for a given cardinality K . This said, no algorithm outperforms all competitors in terms of coefficient accuracy for all scenarios. Nonetheless, the times of computation of NNOG with our recursive implementations are always lower than those of NN-CoSaMP and NN-SP. Let us stress that the structure of the latter algorithms does not easily lend itself to recursive implementations. Indeed, there is no nested property between the supports found at consecutive iterations. Consequently, solving the NNLS problems in a very efficient manner does not seem obvious anymore. For instance, NN-CoSaMP needs to solve NNLS problems for augmented supports of size roughly $3K$. The current K -sparse vector obtained after thresholding may be used as a warm start (although it is usually not an NNLS solution), but the number of inner NNLS iterations is expected to be much larger than for NNOG algorithms.

VI. SPARSE DECOMPOSITION OF NIR SPECTRA

A. Real data and generated dictionary

More than 300 wood samples with different compositions (raw wood, plywood, particle boards, MDF-HDF) and finitions (raw, painted, varnished) were collected on a wood waste park and scanned using a Nicolet 8700 FTIR spectrometer. The resulting reflectance spectra are composed of 1647 wavenumbers covering the NIR range $3600\text{--}10000\text{ cm}^{-1}$ (which corresponds to wavelengths in the interval $[1, 2.8]\text{ }\mu\text{m}$). The aim is to design a binary classifier based on the identification of a subset of informative wavelengths for detecting the so-called non recyclable samples, *i.e.*, MDF-HDF [47], [48]. As the final objective is to design an industrial (fast) sorting system, the number of selected wavelengths has to be as small as possible (typically between 16 and 32). Sparse modeling of the

TABLE II
CPU TIME AND NORMALIZED APPROXIMATION ERROR OF GREEDY ALGORITHMS FOR SPARSE DECOMPOSITION OF NIR SPECTRA. AVERAGE OVER 50 SPECTRA. SYMBOL ** INDICATES THAT THE CONSIDERED ALGORITHM DOES NOT ENFORCE THE NON-NEGATIVE CONSTRAINT.

Algorithm	**OMP	OMP+	NNOMP	FNNOMP	Sparse NNLS	**OLS	OLS+	SNNOLS	NNOLS	FNNOLS	NN-CoSaMP	NN-SP
Time (ms)	23	24	28	25	25	25	26	30	40	2720	45	57
$\ \mathbf{y} - H\mathbf{x}\ /\ \mathbf{y}\ \times 10^2$	6.6	16.4	9.6	9.6	9.6	5.7	16.6	6.4	6.4	7.0	10.1	8.2

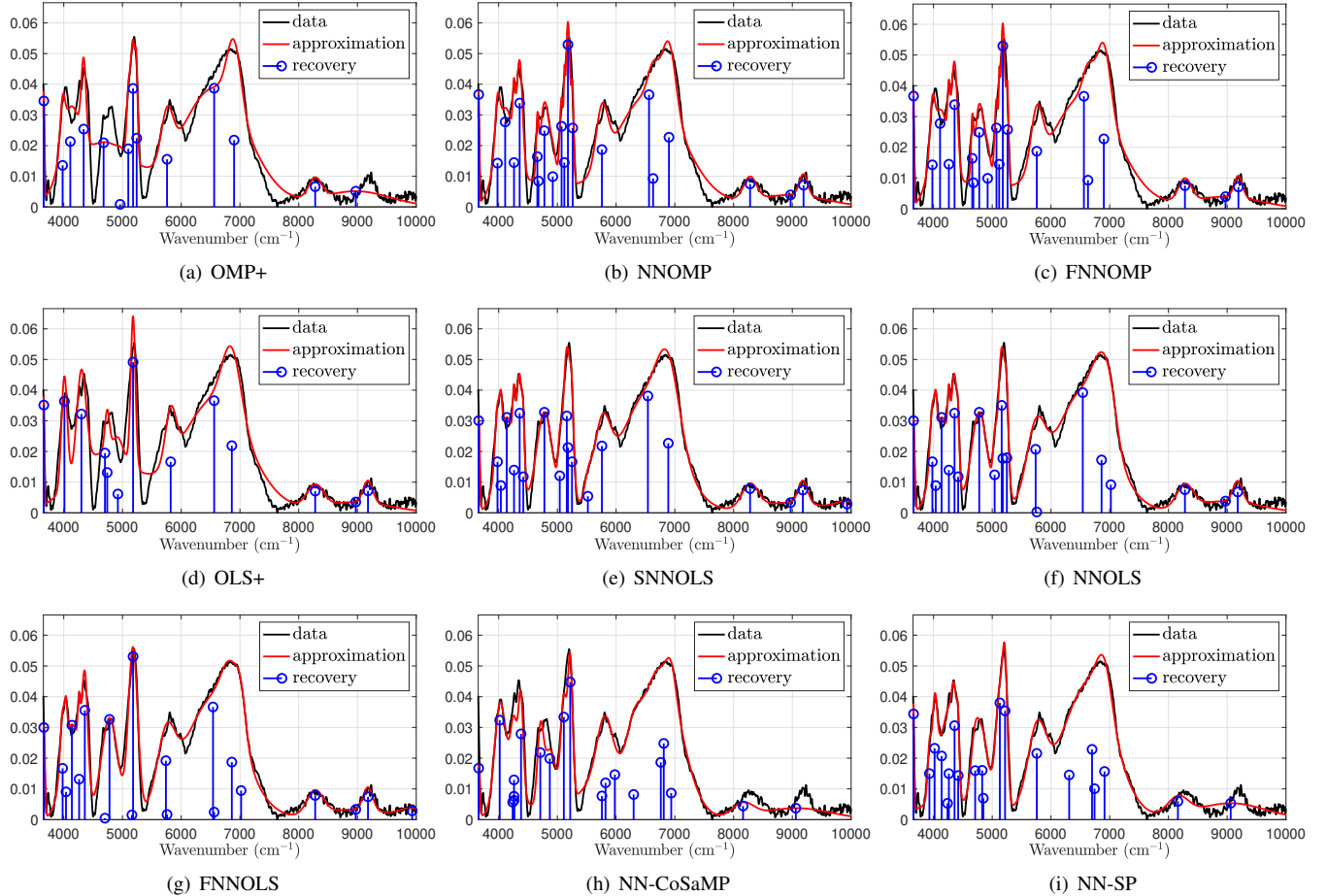


Fig. 4. Data approximation and sparse recovery of an NIR spectrum using various algorithms. The Gaussian dictionary contains 2998 atoms and the sparsity level is set to $K = 20$.

NIR spectra is a common approach to select the informative wavelengths. For instance, Turlach *et al.* [49] proposed an algorithm for variable selection based on an extension of the LASSO and applied it to near infrared measurements.

Hereafter, we consider the decomposition of 50 NIR spectra, seen as data vectors \mathbf{y} of length 1647. Data pre-processing includes baseline removal, offset correction ensuring zero lower bound, and unit energy normalization. To decompose the spectra, a dictionary H is built with Gaussian-shaped columns obtained by discretizing the parameters of a Gaussian function (centers and widths). It is formed by appending the columns of the convolution dictionaries (corresponding to a fixed width σ) used in Section V for 60 equally spaced values of $\sigma \in [10, 600]$ cm^{-1} . The generated dictionary is composed of 2998 atoms. The centers of Gaussian atoms of same width σ are sampled with a step equal to σ , whereas the sampling step of the input signals \mathbf{y} equals 4 cm^{-1} .

B. Decomposition results

For each spectrum, NNOMP, SNNOLS, NNOLS, Sparse NNLS, OMP, OLS are run until a support of cardinality $K = 20$ is found. The simple OMP+ and OLS+ algorithms introduced in § V-D are also considered, together with FNNOMP, FNNOLS, NN-CoSaMP, and NN-SP. For each algorithm, the CPU time and the relative residual norm $\|\mathbf{y} - H\mathbf{x}\|/\|\mathbf{y}\|$, averaged over 50 spectra, are displayed in Tab. II.

First, the time and approximation errors of OMP (resp., OLS) are both lower than those of NNOMP (resp., SNNOLS and NNOLS). This is not a surprise since unconstrained algorithms are expected to reach lower criterion values. When one performs a post-processing by removing these negative spikes such as with OMP+/OLS+, the error scores become larger than those of NNOMP algorithms, and the sparse approximation accuracy is weaker. This confirms the weak capacity of OMP and OLS to reconstruct correct supports from non-negative sparse representations. NNOMP, Sparse NNLS, and FNNOMP

have roughly the same cost and performance. Furthermore, SNNOLS and NNOLS yield the lowest approximation errors.

Similar to the simulations of Section V, the computing times of NN-CoSaMP and NN-SP are greater than those of NNOG implementations. NNOLS is significantly slower than NNOMP and SNNOLS, while SNNOLS has the same performance as NNOLS and a computing time closer to that of NNOMP. Finally, the CPU times of NNOG algorithms *vs* their related unconstrained versions given in Tab. II are consistent with the non-negativity losses gathered in Tab. I.

These approximation and sparse recovery results are further illustrated for a given spectrum in Fig. 4. Note that the Sparse NNLS result (not shown) identifies with that of NNOMP. OMP and OLS (not shown) yield 7 and 9 negative peaks. Besides, the OMP+ and OLS+ approximations around 4500–7000 cm^{-1} and 5000 cm^{-1} , respectively, are rather poor. SNNOLS and NNOLS outputs almost coincide. They outperform that of NNOMP, in particular, around 4200 and 7000 cm^{-1} and those of NN-CoSaMP and NN-SP around 9000 cm^{-1} . FNNOLS is more costly and slightly less accurate than SNNOLS and NNOLS.

VII. CONCLUSION

Until now, greedy algorithms dedicated to non-negative sparse signal reconstruction have been considered as slow schemes, requiring the repeated resolution of constrained least square problems. In order to accelerate the computation, approximate schemes have been proposed [32], [33] at the price of some loss of control on the algorithmic behavior, and possibly degraded performance. Another commonly found option has been to replace the ℓ_0 “norm” by the ℓ_1 -norm and to solve the resulting convex programming problem, with a possible loss in terms of performance (see [27] for an interesting case of sparse NMF).

The first contribution of this paper is to provide a unified framework to define non-negative orthogonal greedy algorithms in a formal way, ensuring well-defined iterations. The second and probably most important one in terms of practical impact, is to show that the additional cost of non-negative greedy algorithms to handle the sign constraint can be strongly reduced using three ingredients. The main one is that non-negative greedy algorithms can be made fully recursive. Moreover, several pruning strategies can be combined to reduce the number of tests at the atom selection stage. Finally, the latter step can benefit from vectorized computations. According to our practical tests with Gaussian dictionaries, the computing time of NNOMP becomes comparable to that of OMP. Hence, we believe that the additional cost should not prevent users from skipping from OMP to NNOMP in nonnegative sparse problem. The performances of NNOMP and FNNOMP are very close both in terms of computational time and accuracy. Since FNNOMP was introduced as a fast but approximate version of NNOMP, it does not seem mandatory to work on approximate schemes to derive efficient implementations. On the other hand, we have obtained a dramatic acceleration of NNOLS compared to both canonical and fast versions proposed in [33]. The computing cost of

NNOLS remains larger than that of OLS. However, whenever OLS can be used, our exact implementation of NNOLS is a realistic option to handle nonnegativity constraints, given that the potential of performance gain between NNOMP and NNOLS is comparable to the one between OMP and OLS in the unconstrained case. We have also studied SNNOLS, which is a suboptimal version of NNOLS originally introduced in [33]. Our conclusion is that SNNOLS represents a good trade-off between NNOMP and NNOLS: it is structurally simpler than NNOLS and possibly faster, with very similar performance in terms of estimation error. Likewise, we have compared NNOMP with Sparse NNLS and concluded that Sparse NNLS is structurally simpler than NNOMP with almost the same performance as NNOMP, and a slightly reduced computing cost. Let us point out that the computation time and accuracy found for different algorithms might potentially follow different trends with different kinds of dictionaries. The latter conclusions (*e.g.*, comparisons between NNOLS and SNNOLS) found with Gaussian dictionaries could therefore be different when working with other problems. We conjecture though that the performance gain between SNNOLS/NNOLS with respect to NNOMP would be observed as well for other problems involving ill-posed dictionaries, since this conclusion is consistent with the practitioner knowledge of OLS and OMP.

Our comparisons with NN-SP and NN-CoSaMP show that NNOG algorithms are very competitive in terms of accuracy, although no algorithm outperforms the others in all scenarios. The proposed recursive implementations yield a substantial reduction of computation compared to NN-SP and NN-CoSaMP. An attractive feature of NNOG algorithms is their versatility. Indeed, the implementations presented in this paper can be used to address both forms of (ℓ_0+) problems with constraints $\|\mathbf{x}\|_0 \leq K$ (as addressed in this paper) or $\|\mathbf{y} - H\mathbf{x}\|_2^2 \leq \varepsilon^2$, where ε^2 is related to the noise variance. In the latter case, one simply needs to replace the stopping condition $|S| = K$ in Algorithm 1 by $\|r_S\|_2^2 \leq \varepsilon^2$. NN-SP, NN-CoSaMP, and NN-HTP are apparently less versatile, although the proposal of novel versions dedicated to the constraint $\|\mathbf{y} - H\mathbf{x}\|_2^2 \leq \varepsilon^2$ could be addressed as a perspective to the work in [30].

Our contributions can be extended in several directions. A straightforward generalization can be made to deal with nonnegativity-constrained simultaneous sparse decomposition, which is useful in several applications such as hyperspectral imaging [39], dynamic PET [50], and diffusion MRI [30]. On the other hand, other greedy algorithms such as BOMP [42] and SBR [43] could also be extended to the non-negative setting using similar principles and using a recursive implementation. Finally, let us mention that [51] establishes the first K -step recovery analysis of NNOMP, NNOLS and SNNOLS under the Mutual Incoherence Property condition.

ACKNOWLEDGMENT

The authors would like to thank Dr. Mehrdad Yaghoobi (University of Edinburgh) for kindly sharing his Matlab implementation of FNNOLS.

APPENDIX A
PROOF OF TECHNICAL RESULTS

A. Proof of Proposition 1

Let us first prove that $\mathbf{h}_i^t \mathbf{r}_S^+ > 0$ implies $i \in \mathcal{D}_S$. Let $\mathbf{h}_i^t \mathbf{r}_S^+ > 0$. According to (1) for $\mathbf{r} = \mathbf{r}_S^+$, we deduce that $i \in \mathcal{D}_S$ since

$$\|\mathbf{r}_{S \cup \{i\}}^+\|^2 \leq \min_{v \geq 0} \|\mathbf{r}_S^+ - v \mathbf{h}_i\|^2 < \|\mathbf{r}_S^+\|^2.$$

Conversely, let $i \in \mathcal{D}_S$. Define $f(\mathbf{z}) = \|\mathbf{y} - H_{S \cup \{i\}} \mathbf{z}\|^2$ where $\mathbf{z} \in \mathbb{R}^{|\mathcal{S}|+1}$. Let us also define the subvectors $\mathbf{z}_S := \hat{\mathbf{x}}_S^+(S \cup \{i\})$ and $\mathbf{z}_{S \cup \{i\}} := \hat{\mathbf{x}}_{S \cup \{i\}}^+(S \cup \{i\})$, $\hat{\mathbf{x}}_S^+$ and $\hat{\mathbf{x}}_{S \cup \{i\}}^+$ being two NNLS solutions related to S and $S \cup \{i\}$. Condition $i \in \mathcal{D}_S$ reads

$$f(\mathbf{z}_{S \cup \{i\}}) < f(\mathbf{z}_S).$$

Since f is convex, one has

$$(\mathbf{z}_{S \cup \{i\}} - \mathbf{z}_S)^t \nabla f(\mathbf{z}_S) \leq f(\mathbf{z}_{S \cup \{i\}}) - f(\mathbf{z}_S) < 0 \quad (21)$$

where the gradient of f is defined by

$$\nabla f(\mathbf{z}_S) = 2H_{S \cup \{i\}}^t (H_{S \cup \{i\}} \mathbf{z}_S - \mathbf{y}) = -2H_{S \cup \{i\}}^t \mathbf{r}_S^+.$$

Denoting by $C := \text{supp}(\hat{\mathbf{x}}_S^+)$ the compressed support, we have from Lemma 1 that $H_{S \setminus C}^t \mathbf{r}_S^+ \leq \mathbf{0}$ and $H_C^t \mathbf{r}_S^+ = \mathbf{0}$. Since $\hat{\mathbf{x}}_S^+$ is supported by C , the latter equality implies that $\mathbf{z}_S^t \nabla f(\mathbf{z}_S) = 0$. (21) yields $(\mathbf{z}_{S \cup \{i\}})^t H_{S \cup \{i\}}^t \mathbf{r}_S^+ > 0$, i.e.,

$$(\hat{\mathbf{x}}_{S \cup \{i\}}^+(S \cup \{i\}))^t H_{S \cup \{i\}}^t \mathbf{r}_S^+ > 0. \quad (22)$$

Since $H_C^t \mathbf{r}_S^+ = \mathbf{0}$, (22) rereads:

$$(\hat{\mathbf{x}}_{S \cup \{i\}}^+(S \setminus C))^t H_{S \setminus C}^t \mathbf{r}_S^+ + (\mathbf{h}_i^t \mathbf{r}_S^+) \hat{\mathbf{x}}_{S \cup \{i\}}^+(i) > 0 \quad (23)$$

and since $H_{S \setminus C}^t \mathbf{r}_S^+ \leq \mathbf{0}$ and $\hat{\mathbf{x}}_{S \cup \{i\}}^+ \geq \mathbf{0}$, (23) implies that

$$(\mathbf{h}_i^t \mathbf{r}_S^+) \hat{\mathbf{x}}_{S \cup \{i\}}^+(i) > 0 \quad (24)$$

and thus $\mathbf{h}_i^t \mathbf{r}_S^+ > 0$.

Let us now assume that S is a positive support. According to Lemma 2, we have $\hat{\mathbf{x}}_S^+ = \hat{\mathbf{x}}_S$ and $\mathbf{r}_S^+ = \mathbf{r}_S$, so (7) and (8) are identical, as well as (9) and (10). To show that (7)-(8) are equivalent to (9)-(10), we first notice that $\mathbf{r}_S \in (\text{span}(H_S))^\perp$, thus $\tilde{\mathbf{h}}_i^t \mathbf{r}_S = \mathbf{h}_i^t \mathbf{r}_S$ since $\tilde{\mathbf{h}}_i - \mathbf{h}_i \in \text{span}(H_S)$. Therefore, (8) rereads $0 < \mathbf{h}_i^t \mathbf{r}_S$, which implies that $\tilde{\mathbf{h}}_i \neq \mathbf{0}$ and $\tilde{\mathbf{g}}_i = \tilde{\mathbf{h}}_i / \|\tilde{\mathbf{h}}_i\| \neq \mathbf{0}$. Hence, (8) rereads $0 < \tilde{\mathbf{g}}_i^t \mathbf{r}_S$, which identifies with (10).

Finally, let us show that $i \in \mathcal{D}_S$ is equivalent to condition (11). Consider the function $f(\mathbf{z}) = \|\mathbf{y} - H_{S \cup \{i\}} \mathbf{z}\|^2$ and the notations \mathbf{z}_S and $\mathbf{z}_{S \cup \{i\}}$ defined above. Assuming that $H_{S \cup \{i\}}$ is full column rank, we have that f is strictly convex, so f admits a unique minimizer $\mathbf{z}_{S \cup \{i\}}$. If $\hat{\mathbf{x}}_{S \cup \{i\}}^+(i) > 0$, then $\mathbf{z}_{S \cup \{i\}} \neq \mathbf{z}_S$ and

$$\|\mathbf{r}_{S \cup \{i\}}^+\|^2 = f(\mathbf{z}_{S \cup \{i\}}) < \|\mathbf{r}_S^+\|^2 = f(\mathbf{z}_S),$$

that is, $i \in \mathcal{D}_S$. Conversely, $\hat{\mathbf{x}}_{S \cup \{i\}}^+(i) = 0$ implies $\hat{\mathbf{x}}_{S \cup \{i\}}^+ = \hat{\mathbf{x}}_S^+$, hence $\mathbf{r}_{S \cup \{i\}}^+ = \mathbf{r}_S^+$ and $i \notin \mathcal{D}_S$.

Algorithm 4: Recursive ULS [21].
Format: ULS($\mathbf{y}, H, V, \mathbf{f}\mathbf{w}, \ell, \hat{\mathbf{x}}, \Theta, e^2$)

```

1 if fw then
2    $\phi \leftarrow H_V^t \mathbf{h}_\ell$ ;
3    $\delta \leftarrow (1 - \phi^t \Theta \phi)^{-1}$ ;
4    $\beta \leftarrow \phi^t \hat{\mathbf{x}}(V) - \mathbf{h}_\ell^t \mathbf{y}$ ;
5    $e^2 \leftarrow e^2 - \delta \beta^2$ ;
6    $\hat{\mathbf{x}}(V \cup \{\ell\}) \leftarrow \hat{\mathbf{x}}(V \cup \{\ell\}) + \delta \beta \begin{bmatrix} \Theta \phi \\ -1 \end{bmatrix}$ ;
7    $\Theta \leftarrow \begin{bmatrix} \Theta & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} + \delta \begin{bmatrix} \Theta \phi \\ -1 \end{bmatrix} \begin{bmatrix} \Theta \phi \\ -1 \end{bmatrix}^t$ ;
8    $V \leftarrow V \cup \{\ell\}$ ;
9 else
10   $j \leftarrow \text{index of } \ell \text{ in } V$ ;
11   $e^2 \leftarrow e^2 + (\hat{\mathbf{x}}(\ell))^2 / \theta_j(j)$ ;
12   $\hat{\mathbf{x}}(V) \leftarrow \hat{\mathbf{x}}(V) - \hat{\mathbf{x}}(\ell) \theta_j / \theta_j(j)$ ;
13   $\Theta \leftarrow \Theta(-j, -j) - \theta_j(-j) \theta_j(-j)^t / \theta_j(j)$ ;
14   $V \leftarrow V \setminus \{\ell\}$ ;
15 end
```

B. Proof of Proposition 3

Any iteration of AS-NNLS starts with the addition of a new atom to the current support (Line 4 of Algorithm 2). Then, a variable number of atoms are removed from it one after the other (Lines 6 to 12). Let r denote the number of removals at the current AS-NNLS iteration. Let also $V \subset \{1, \dots, n\}$ and $\hat{\mathbf{x}}_V$ respectively stand for the current support and solution obtained at Lines 4 and 5, and let $V' \subset V$ and $\hat{\mathbf{x}}_{V'}$ denote the corresponding quantities after r removals. Let us show that any AS-NNLS iteration for which $r = 0$ or $r = 1$ yields a solution of the NNLS problem restricted to the support V .

If $\hat{\mathbf{x}}_V \geq \mathbf{0}$, then $r = 0$, so $V' = V$ and $\hat{\mathbf{x}}_{V'} = \hat{\mathbf{x}}_V^+ \geq \mathbf{0}$. Otherwise, we have $\min(\hat{\mathbf{x}}_V) < 0$ and $r > 0$. If $r = 1$, a single index ℓ^- is removed at Lines 6-12, so that $V' = V \setminus \{\ell^-\}$, and $\hat{\mathbf{x}}_{V'} \geq \mathbf{0}$. Let us remark that we have $\hat{\mathbf{x}}_{V'}(\ell^-) < 0$ according to Line 7. Let us then prove that $\hat{\mathbf{x}}_{V'} = \hat{\mathbf{x}}_V^+$ by showing that KKT conditions are satisfied at $\hat{\mathbf{x}}_{V'}$ for the NNLS problem related to support V . Note that the NNLS solution is unique since the supports generated by AS-NNLS are such that H_V is full column rank, as pointed out in subsection III-B. According to Lemma 1, the KKT conditions read:

$$H_{V'}^t (\mathbf{y} - H \hat{\mathbf{x}}_{V'}) = \mathbf{0}, \quad (25a)$$

$$\mathbf{h}_{\ell^-}^t (\mathbf{y} - H \hat{\mathbf{x}}_{V'}) \leq 0. \quad (25b)$$

(25a) is obviously satisfied. On the other hand, remark that $\hat{\mathbf{x}}_{V'} = \hat{\mathbf{x}}_V^+$, since $\hat{\mathbf{x}}_{V'} \geq \mathbf{0}$ and that according to Proposition 1, $\hat{\mathbf{x}}_{V'}(\ell^-) < 0$ implies that $\mathbf{h}_{\ell^-}^t \mathbf{r}_{V'} \leq \mathbf{0}$, which identifies with (25b). This concludes the proof.

APPENDIX B
RECURSIVE IMPLEMENTATION OF ULS

Algorithm 4 recalls the recursive ULS computation using MIL [21] for selection (forward move $V \leftarrow V \cup \{\ell\}$) and de-selection (backward move $V \leftarrow V \setminus \{\ell\}$) operations. The ULS solution $\hat{\mathbf{x}} := \hat{\mathbf{x}}_V$ is updated. Moreover, $\Theta := (H_V^t H_V)^{-1}$

refers to the inverse of the Gram matrix related to subset V . The Boolean entry `fw` is set to `true` and `false` for selection and deselection updates, respectively. Finally, e^2 stands for the squared residual error $\|\mathbf{r}_V\|^2$. All these factors are updated in Algorithm 4. Notation $-j$ refers to all indices except j , and θ_j stands for the j -th column of Θ .

The calls to Algorithm 4 for updating ULS solutions at Lines 4-5 and 10-11 of Algorithm 2 take the respective forms:

$$\begin{aligned} & \text{ULS}(\mathbf{y}, H, V, 1, \ell^+, \hat{\mathbf{x}}_V, \Theta, \|\mathbf{r}_V\|^2), \\ & \text{ULS}(\mathbf{y}, H, V, 0, \ell^-, \hat{\mathbf{x}}_V, \Theta, \|\mathbf{r}_V\|^2). \end{aligned}$$

REFERENCES

- [1] Y. C. Pati, R. Rezaifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition", in *Proc. 27th Asilomar Conf. on Signals, Sys. and Comp.*, 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] A. J. Miller, *Subset selection in regression*, Chapman and Hall, London, UK, 2nd edition, Apr. 2002.
- [4] 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.
- [5] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach", *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 137–140, Apr. 2002.
- [6] 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 Proc. in Mathematics & Statistics.
- [7] B. L. Sturmfels and M. G. Christensen, "Comparison of orthogonal matching pursuit implementations", in *Proc. Eur. Sig. Proc. Conf.*, Bucharest, Romania, Aug. 2012, pp. 220–224.
- [8] 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.
- [9] E. Esser, Y. Lou, and J. Xin, "A method for finding structured sparse solutions to nonnegative least squares problems with applications", *SIAM J. Imaging Sci.*, vol. 6, no. 4, pp. 2010–2046, Oct. 2013.
- [10] T. Virtanen, J. F. Gemmeke, and B. Raj, "Active-set Newton algorithm for overcomplete non-negative representations of audio", *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 11, pp. 2277–2289, Nov. 2013.
- [11] R. Bro and S. De Jong, "A fast non-negativity-constrained least squares algorithm", *Journal of Chemometrics*, vol. 11, no. 5, pp. 393–401, 1997.
- [12] M. Slawski, R. Hussong, A. Tholey, T. Jakoby, B. Gregorius, A. Hildebrandt, and M. Hein, "Isotope pattern deconvolution for peptide mass spectrometry by non-negative least squares/least absolute deviation template matching", *BMC Bioinformatics*, vol. 13, no. 291, pp. 1–18, Nov. 2012.
- [13] J. A. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines", *Astron. Astrophys. Suppl.*, vol. 15, pp. 417–426, 1974.
- [14] S. Petra and C. Schnörr, "Average case recovery analysis of tomographic compressive sensing", *Linear Alg. Appl.*, vol. 441, pp. 168–198, 2014.
- [15] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints", *J. Mach. Learn. Res.*, vol. 5, pp. 1457–1469, 2004.
- [16] J. Rapin, J. Bobin, A. Larue, and J.-L. Starck, "Sparse and non-negative BSS for noisy data", *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5620–5632, Nov. 2013.
- [17] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, chapter 10, pp. 185–212, Springer-Verlag, New York, 2011.
- [18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers", *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, July 2011.
- [19] G. Gasso, A. Rakotomamonjy, and S. Canu, "Recovering sparse signals with a certain family of nonconvex penalties and DC programming", *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4686–4698, Dec. 2009.
- [20] H. A. Le Thi, B. T. Nguyen Thi, and H. M. Le, "Sparse signal recovery by difference of convex functions algorithms", in *Intelligent Information and Database Systems*, A. Selamat, N. T. Nguyen, and H. Haron, Eds., Berlin, 2013, vol. 7803 of *Lect. Notes Comput. Sci.*, pp. 387–397, Springer Verlag.
- [21] A. Björck, *Numerical Methods for Least Squares Problems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, Apr. 1996.
- [22] C. L. Lawson and R. J. Hanson, *Solving least squares problems*, pp. 149–199, Society for Industrial and Applied Mathematics, 1974.
- [23] M. H. Wright, "Interior methods for constrained optimization", *Acta Numerica*, vol. 1, pp. 341–407, Jan. 1992.
- [24] F. Benvenuto, R. Zanella, L. Zanni, and M. Bertero, "Nonnegative least-squares image deblurring: Improved gradient projection approaches", *Inverse Probl.*, vol. 26, no. 2, pp. 1–18, Feb. 2010.
- [25] M. Slawski and M. Hein, "Sparse recovery by thresholded non-negative least squares", *Adv. Neural Inf. Process. Syst.*, vol. 24, pp. 1926–1934, 2011.
- [26] M. Slawski and M. Hein, "Non-negative least squares for high-dimensional linear models: Consistency and sparse recovery without regularization", *Electron. J. Stat.*, vol. 7, pp. 3004–3056, 2013.
- [27] R. Peharz and F. Pernkopf, "Sparse nonnegative matrix factorization with ℓ^0 constraints", *Neurocomputing*, vol. 80, pp. 38–46, Mar. 2012.
- [28] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples", *Appl. Comp. Harmonic Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.
- [29] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction", *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.
- [30] D. Kim and J. P. Haldar, "Greedy algorithms for nonnegativity-constrained simultaneous sparse recovery", *Signal Process.*, vol. 125, pp. 274–289, 2016.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] T. Blumensath and M. E. Davies, "On the difference between Orthogonal Matching Pursuit and Orthogonal Least Squares", Tech. Rep., Univ. Edinburgh, Mar. 2007.
- [35] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer texts in Operations Research and Financial Engineering. Springer Verlag, New York, 2nd edition, July 2006.
- [36] S. Leichter, G. Dantzig, and J. Davis, "A strictly improving linear programming Phase I algorithm", *Ann. Oper. Res.*, vol. 47, pp. 409–430, 1993.
- [37] M. Mørup, K. H. Madsen, and L. K. Hansen, "Approximate L0 constrained non-negative matrix and tensor factorization", in *IEEE Int. Symp. Circuits Syst.*, Seattle, WA, May 2008, pp. 1328–1331.
- [38] K. N. Ramamurthy, J. J. Thiagarajan, and A. Spanias, "Recovering non-negative and combined sparse representations", *Digital Signal Process.*, vol. 26, no. 1, pp. 21–35, Mar. 2014.
- [39] Z. Wang, R. Zhu, K. Fukui, and J. Xue, "Cone-based joint sparse modelling for hyperspectral image classification", *Signal Process.*, vol. 144, pp. 417–429, 2018.
- [40] X. Chen, F. Xu, and Y. Ye, "Lower bound theory of nonzero entries in solutions of ℓ_2 - ℓ_p minimization", *SIAM J. Sci. Comput.*, vol. 32, no. 5, pp. 2832–2852, 2010.
- [41] S. Foucart and D. Koslicki, "Sparse recovery by means of nonnegative least squares", *IEEE Signal Process. Lett.*, vol. 21, no. 4, pp. 498–502, Apr. 2014.
- [42] C. Herzet and A. Drémeau, "Bayesian pursuit algorithms", in *Proc. Eur. Sig. Proc. Conf.*, Aalborg, Denmark, Aug. 2010, pp. 1474–1478.
- [43] C. Soussen, J. Idier, D. Brie, and J. Duan, "From Bernoulli-Gaussian deconvolution to sparse signal restoration", *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4572–4584, Oct. 2011.
- [44] T. Zhang, "Sparse recovery with orthogonal matching pursuit under RIP", *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6215–6221, Sept. 2011.
- [45] A. Bonnefoy, V. Emiya, L. Ralaivola, and R. Gribonval, "Dynamic screening: Accelerating first-order algorithms for the lasso and group-lasso", *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5121–5132, Oct. 2015.

- [46] T. T. Nguyen, C. Soussen, J. Idier, and E.-H. Djermoune, “Non-negative greedy algorithms: Matlab implementation”, <https://www.codeocean.com/>, Mar. 2019.
- [47] L. Belmerhnia, E.-H. Djermoune, C. Carteret, and D. Brie, “Simultaneous regularized sparse approximation for wood wastes NIR spectra features selection”, in *Proc. CAMSAP*, Cancun, Mexico, Dec. 2015.
- [48] K. Wagner, T. Schnabel, M.-C. Barbu, and A. Petutschnigg, “Analysis of selected properties of fibreboard panels manufactured from wood and leather using the near infrared spectroscopy”, *Int J. Spectrosc.*, vol. 2015, pp. 691796, 2015.
- [49] B. A. Turlach, W. N. Venables, and S. J. Wright, “Simultaneous variable selection”, *Technometrics*, vol. 47, no. 3, pp. 349–363, 2005.
- [50] Y. Lin, J. P. Haldar, Q. Li, P. Conti, and R. M. Leahy, “Sparsity constrained mixture modeling for the estimation of kinetic parameters in dynamic PET”, *IEEE Trans. Med. Imag.*, vol. 33, no. 1, pp. 173–185, Jan. 2014.
- [51] T. T. Nguyen, C. Soussen, J. Idier, and E.-H. Djermoune, “Sign preservation analysis of orthogonal greedy algorithms”, Tech. rep., Univ. Lorraine, CentraleSupélec, LS2N, <https://hal.archives-ouvertes.fr/hal-01971697>, Nancy, France, Jan. 2019.