Sparse recovery conditions for Orthogonal Least Squares

Charles Soussen, Rémi Gribonval, Jérôme Idier, Cédric Herzet

Sparse recovery conditions for Orthogonal Least Squares

Charles Soussen*, Rémi Gribonval, Jérôme Idier, and Cédric Herzet

Abstract

We extend Tropp's analysis of Orthogonal Matching Pursuit (OMP) using the Exact Recovery Condition (ERC) [1] to a first exact recovery analysis of Orthogonal Least Squares (OLS). We show that when ERC is met, OLS is guaranteed to exactly recover the unknown support. Moreover, we provide a closer look at the analysis of both OMP and OLS when ERC is not fulfilled. We show that there exist dictionaries for which some subsets are never recovered with OMP. This phenomenon, which also appears with ℓ_1 minimization, does not occur for OLS. Finally, numerical experiments based on our theoretical analysis show that none of the considered algorithms is uniformly better than the other.

Index Terms

ERC exact recovery condition; Orthogonal Matching Pursuit; Orthogonal Least Squares; Order Recursive Matching Pursuit; Optimized Orthogonal Matching Pursuit; forward selection.

C. Soussen is with the Centre de Recherche en Automatique de Nancy (CRAN, UMR 7039, Nancy-University, CNRS). Campus Sciences, B.P. 70239, F-54506 Vandœuvre-lès-Nancy, France. Tel: (+33)-3 83 68 44 71, Fax: (+33)-3 83 68 44 62 (e-mail: Charles.Soussen@cran.uhp-nancy.fr.) This work was carried out in part while C. Soussen was visiting IRCCyN during the academic year 2010-2011 with the financial support of CNRS.

R. Gribonval and C. Herzet are with INRIA Rennes - Bretagne Atlantique, Campus de Beaulieu, F-35042 Rennes Cedex, France Tel: (+33)-2 99 84 25 06/73 50, Fax: (+33)-2 99 84 71 71 (e-mail: Remi.Gribonval@inria.fr; Cedric.Herzet@inria.fr). R. Gribonval acknowledges the partial support of the European Union's FP7-FET program, SMALL project, under grant agreement n° 225913.

J. Idier is with the Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN, UMR CNRS 6597), BP 92101, 1 rue de la Noë, 44321 Nantes Cedex 3, France. Tel: (+33)-2 40 37 69 09, Fax: (+33)-2 40 37 69 30 (e-mail: Jerome.Idier@irccyn.ec-nantes.fr).

CONTENTS

I. Introduction

Classical greedy subset selection algorithms include, by increasing order of complexity: Matching Pursuit (MP) [2], Orthogonal Matching Pursuit (OMP) [3] and Orthogonal Least Squares (OLS) [4, 5]. OLS is indeed relatively expensive in comparison with OMP since OMP performs one linear inversion per iteration whereas OLS performs as many linear inversions as there are non-active atoms. We refer the reader to the technical report [6] for a comprehensive review on the difference between OMP and OLS.

OLS is referred to using many other names in the literature. It is known as forward selection in statistical regression [7] and as the greedy algorithm [5], Order Recursive Matching Pursuit (ORMP) [8] and Optimized Orthogonal Matching Pursuit (OOMP) [9] in the signal processing literature, all these algorithms being actually the same. It is worth noticing that the above-mentioned algorithms were introduced by following either an optimization [4, 7] or an orthogonal projection methodology [5], or both [8, 9]. In the optimization viewpoint, the atom yielding the largest decrease of the approximation error is selected. This leads to a greedy sub-optimal algorithm dedicated to the minimization of the approximation error. In the orthogonal projection viewpoint, the atom selection rule is defined as an extension of the OMP rule: the data vector and the dictionary atoms are being projected onto the subspace

that is orthogonal to the span of the active atoms, and the *normalized* projected atom having the largest inner product with the data residual is selected. As the number of active atoms increases by one at any iteration, the projections are done on a subspace whose dimension is decreasing.

A. Main objective of the paper

Our primary goal is to address the OLS exact recovery analysis from noise-free data and to investigate the connection between the OMP and OLS exact recovery conditions. In the literature, much attention was paid to the exact recovery analysis of sparse algorithms that are faster than OLS, *e.g.,* thresholding algorithms and simpler greedy algorithms like OMP [10]. But to the best of our knowledge, no exact recovery result is available for OLS. In their recent paper [11], Davies and Eldar mention this issue and state that the relation between OMP and OLS remains unclear.

B. Existing results for OMP

Our starting point is the existing analysis of OMP whose structure is somewhat close to OLS. Exact recovery studies rely on alternate methodologies.

Tropp's Exact Recovery Condition (ERC) [1] is a necessary and sufficient condition of exact recovery in a worst case analysis. On the one hand, if a subset of k atoms satisfies the ERC, then it can be recovered from any linear combination of the k atoms in at most k steps. On the other hand, when the ERC is not satisfied, one can generate a counterexample (*i.e.,* a specific combination of the k atoms) for which OMP fails, *i.e.,* OMP selects a wrong atom during its first k iterations. Specifically, the atom selected in the *first* iteration is a wrong one.

Davenport and Wakin [12] used another analysis to show that OMP yields exact support recovery under certain Restricted Isometry Property (RIP) assumptions. Actually, the ERC necessarily holds when Davenport and Wakin's condition is fulfilled since ERC is a necessary and sufficient condition of exact recovery.

C. Generalization of Tropp's condition

We propose to extend Tropp's condition to OLS. We remark that the very first iteration of OLS is identical to that of OMP: the first selected atom is the one whose inner product with the input vector is maximal. Therefore, when ERC does not hold, the counterexample for which the first iteration of OMP fails also yields a failure of the first iteration of OLS. Hence one cannot expect to derive an exact recovery

condition for OLS which is weaker than ERC at the first iteration. We show that the ERC indeed ensures the success of OLS.

We further address the case where ERC does not hold, *i.e.*, the first iteration of OMP/OLS[1] is not guaranteed to succeed but nevertheless succeeds "by chance". We derive weaker conditions which guarantee that an exact support recovery occurs in the subsequent iterations. These extended recovery conditions coincide with ERC at the first iteration but differ from it from the second iteration.

In summary, our main results state that:

- Tropp's ERC is a sufficient condition of exact recovery for OLS (Theorem 2).
- When the early iterations of Oxx have all succeeded, we derive two sufficient conditions, named ERC-OMP and ERC-OLS, for the recovery of the remaining true atoms (Theorem 3).
- Moreover, we show that our conditions are, in some sense, necessary (Theorems 4 and 5).

D. Organization of the paper

In Section II, we recall the principle of OMP and OLS and their interpretation in terms of orthogonal projections. Then, we properly define the notions of successful support recovery and support recovery failure. Section III is dedicated to the analysis of OMP and OLS at any iteration where the most technical developments and proofs are omitted for readability reasons. These important elements can be found in the appendix section A. In Section IV, we show using Monte Carlo simulations that there is no systematic implication between the ERC-OMP and ERC-OLS conditions but we exhibit some elements of discrimination between OMP and OLS.

II. NOTATIONS AND PREREQUISITES

The following notations will be used in this paper. $\langle\,.\,,\,.\,\rangle$ refers to the inner product between vectors, and $\|\,.\,\|$ and $\|\,.\,\|_1$ stand for the Euclidean norm and the ℓ_1 norm, respectively. $.^\dagger$ denotes the pseudo-inverse of a matrix. For a full rank and undercomplete matrix, we have $\boldsymbol{X}^\dagger = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t$ where $.^t$ stands for the matrix transposition. When $\boldsymbol{X}$ is overcomplete, $\mathrm{spark}(\boldsymbol{X})$ denotes the minimum number of columns from $\boldsymbol{X}$ that are linearly dependent [13]. The letter $\mathcal{Q}$ denotes some subset of the column indices, and $\boldsymbol{X}_\mathcal{Q}$ is the submatrix of $\boldsymbol{X}$ gathering the columns indexed by $\mathcal{Q}$. Finally, $\boldsymbol{P}_\mathcal{Q} = \boldsymbol{X}_\mathcal{Q}\boldsymbol{X}_\mathcal{Q}^\dagger$ and $\boldsymbol{P}_\mathcal{Q}^\perp = \boldsymbol{I} - \boldsymbol{P}_\mathcal{Q}$ denote the orthogonal projection operators on $\mathrm{span}(\boldsymbol{X}_\mathcal{Q})$ and $\mathrm{span}(\boldsymbol{X}_\mathcal{Q})^\perp$, where $\mathrm{span}(\boldsymbol{X})$ stands for the column span of $\boldsymbol{X}$, $\mathrm{span}(\boldsymbol{X})^\perp$ is the orthogonal complement of $\mathrm{span}(\boldsymbol{X})$ and $\boldsymbol{I}$ is the identity matrix whose dimension is equal to the number of rows in $\boldsymbol{X}$.

[1]In the rest of the paper, we will use the notation Oxx when referring to properties that apply to both OMP and OLS.

A. Subset selection

Let $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ denote the dictionary gathering unitary atoms $\boldsymbol{a}_i \in \mathbb{R}^m$. $\boldsymbol{A}$ is a matrix of size $m \times n$. Assuming that the atoms are unitary is actually not necessary for OLS as the behavior of OLS is unchanged whether the atoms are normalized or not [6]. On the contrary, OMP is highly sensitive to the normalization of atoms since its selection rule involves the inner products between the current residual and the non-selected atoms.

We consider a subset $\mathcal{Q}^\star$ of $\{1, \ldots, n\}$ of cardinality $k \triangleq \mathrm{Card}\,[\mathcal{Q}^\star] < \min(m, n)$ and study the behavior of OMP and OLS *for all* inputs $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$, *i.e.,* for any combination $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}^\star}\boldsymbol{t}$ where the submatrix $\boldsymbol{A}_{\mathcal{Q}^\star}$ is of size $m \times k$ and the weight vector $\boldsymbol{t} \in \mathbb{R}^k$. The k atoms $\{\boldsymbol{a}_i,\, i \in \mathcal{Q}^\star\}$ indexed by $\mathcal{Q}^\star$ will be referred to as the "true" atoms while for the remaining ("wrong") atoms $\{\boldsymbol{a}_i,\, i \notin \mathcal{Q}^\star\}$, we will use the generic notation $\boldsymbol{a}_{\mathrm{bad}}$. The forward greedy algorithms considered in this paper start from the empty support and select a new atom per iteration. At intermediate iterations $j \in \{0, \ldots, k-1\}$, we denote by $\mathcal{Q}$ the current support (with $\mathrm{Card}\,[\mathcal{Q}] = j$).

Throughout the paper, we make the general assumption that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. It is important to mention that this assumption does not guarantee that the representation $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}^\star}\boldsymbol{t}$ is unique, *i.e.,* there may be another k-term representation $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}'}\boldsymbol{t}'$ where $\boldsymbol{A}_{\mathcal{Q}'}$ includes some wrong atoms $\boldsymbol{a}_{\mathrm{bad}}$. The stronger assumption $\mathrm{spark}(\boldsymbol{A}) > 2k$ is a necessary and sufficient condition for uniqueness of any k-term representation [13]. Therefore, when $\mathrm{spark}(\boldsymbol{A}) > 2k$, the selection of a wrong atom by a greedy algorithm disables a k-term representation of $\boldsymbol{y}$ in k steps [1]. We make the weak assumption that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank because it is sufficient to elaborate our exact recovery conditions under which no wrong atom is selected in the first k iterations.

B. OMP and OLS algorithms

The common feature between OMP and OLS is that they both perform an orthogonal projection whenever the support $\mathcal{Q}$ is updated: the data approximation reads $\boldsymbol{P}_{\mathcal{Q}}\boldsymbol{y}$ and the residual error is defined by

$$\boldsymbol{r}_{\mathcal{Q}} \triangleq \boldsymbol{y} - \boldsymbol{P}_{\mathcal{Q}}\boldsymbol{y} = \boldsymbol{P}_{\mathcal{Q}}^{\perp}\boldsymbol{y}.$$

Let us now recall how the selection rule of OLS differs from that of OMP.

At each iteration of OLS, the atom $\boldsymbol{a}_\ell$ yielding the minimum least-square error $\|\boldsymbol{r}_{\mathcal{Q}\cup\{\ell\}}\|^2$ is selected:

$$\ell^{\mathrm{OLS}} \in \operatorname*{arg\,min}_{i \notin \mathcal{Q}} \|\boldsymbol{r}_{\mathcal{Q}\cup\{i\}}\|^2$$

and $n - \mathrm{Card}\,[\mathcal{Q}]$ least-square problems are being solved to compute $\|\boldsymbol{r}_{\mathcal{Q}\cup\{i\}}\|^2$ for all $i \notin \mathcal{Q}$ ([2]) [4]. On the contrary, OMP adopts the simpler rule

$$\ell^{\mathrm{OMP}} \in \arg\max_{i\notin\mathcal{Q}} |\langle \boldsymbol{r}_{\mathcal{Q}}, \boldsymbol{a}_i\rangle|$$

to select the new atom $\boldsymbol{a}_\ell$ and then solves only one least-square problem to compute $\|\boldsymbol{r}_{\mathcal{Q}\cup\{\ell\}}\|^2$ [6]. Depending on the application, the OMP and OLS stopping rules can involve a maximum number of atoms and/or a residual threshold. Note that when the data are noise-free (they read as $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}^\star}\boldsymbol{t}$) and no wrong atom is selected, the squared error $\|\boldsymbol{r}_{\mathcal{Q}}\|^2$ is equal to 0 after at most k iterations. Therefore, we will consider no more than k iterations in the following.

C. Geometric interpretation

A geometric interpretation in terms of orthogonal projections will be useful for deriving recovery conditions. It is essentially inspired by the technical report of Blumensath and Davies [6] and by Davenport and Wakin's analysis of OMP under the RIP assumption [12].

We introduce the notation $\tilde{\boldsymbol{a}}_i = \boldsymbol{P}_{\mathcal{Q}}^{\perp}\boldsymbol{a}_i$ for the projected atoms onto $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})^{\perp}$ where for simplicity, the dependence upon $\mathcal{Q}$ is omitted. When there is a risk of confusion, we will use $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}}$ instead of $\tilde{\boldsymbol{a}}_i$. Notice that $\tilde{\boldsymbol{a}}_i = \boldsymbol{0}$ *if and only if* $\boldsymbol{a}_i \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})$. In particular, $\tilde{\boldsymbol{a}}_i = \boldsymbol{0}$ for $i \in \mathcal{Q}$. Finally, we define the normalized vectors

$$\tilde{\boldsymbol{b}}_i = \begin{cases} \tilde{\boldsymbol{a}}_i/\|\tilde{\boldsymbol{a}}_i\| & \text{if } \tilde{\boldsymbol{a}}_i \neq \boldsymbol{0}, \\ \boldsymbol{0} & \text{otherwise.} \end{cases}$$

Again, we will use $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}}$ when there is a risk of confusion.

We now emphasize that the projected atoms $\tilde{\boldsymbol{a}}_i$ (or $\tilde{\boldsymbol{b}}_i$) play a central role in the analysis of both OMP and OLS. Because the residual $\boldsymbol{r}_{\mathcal{Q}} = \boldsymbol{P}_{\mathcal{Q}}^{\perp}\boldsymbol{y}$ lays in $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})^{\perp}$, $\langle \boldsymbol{r}_{\mathcal{Q}}, \boldsymbol{a}_i\rangle = \langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{a}}_i\rangle$ and the OMP selection rule rereads:

$$\ell^{\mathrm{OMP}} \in \arg\max_{i\notin\mathcal{Q}} |\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{a}}_i\rangle| \tag{1}$$

whereas for OLS, minimizing $\|\boldsymbol{r}_{\mathcal{Q}\cup\{i\}}\|^2$ with respect to $i \notin \mathcal{Q}$ is equivalent to maximizing $\|\boldsymbol{r}_{\mathcal{Q}}\|^2 - \|\boldsymbol{r}_{\mathcal{Q}\cup\{i\}}\|^2 = \langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{b}}_i\rangle^2$ (see *e.g.,* [9] for a complete calculation):

$$\ell^{\mathrm{OLS}} \in \arg\max_{i\notin\mathcal{Q}} |\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{b}}_i\rangle|. \tag{2}$$

[2]Our purpose is not to focus on the OLS implementation. However, let us just mention that in the typical implementation, the least-square problems are solved recursively using the Gram Schmidt orthonormalization procedure [4].

We notice that (1) and (2) only rely on the vectors $\boldsymbol{r}_{\mathcal{Q}}$ and $\tilde{\boldsymbol{a}}_i$ belonging to the subspace $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})^{\perp}$. OMP maximizes the inner product $|\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{a}}_i \rangle|$ whereas OLS minimizes the angle between $\boldsymbol{r}_{\mathcal{Q}}$ and $\tilde{\boldsymbol{a}}_i$ (this difference was already stressed and graphically illustrated in [6]). When the dictionary is close to orthogonal, *e.g.*, for dictionaries satisfying the RIP assumption, this does not make a strong difference since $\|\tilde{\boldsymbol{a}}_i\|$ is close to 1 for all atoms [12]. But in the general case, $\|\tilde{\boldsymbol{a}}_i\|$ may have wider variations between 0 and 1 leading to substantial differences between the behavior of OMP and OLS.

D. Definition of successful recovery and failure

Throughout the paper, we will use the unifying notation

$$\tilde{\boldsymbol{c}}_i \triangleq \begin{cases} \tilde{\boldsymbol{a}}_i & \text{for OMP,} \\ \tilde{\boldsymbol{b}}_i & \text{for OLS} \end{cases}$$

for statements that are common to OMP and OLS.

We first stress that in special cases where the Oxx selection rule yields multiple solutions including a wrong atom, *i.e.*, when

$$\max_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}} |\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle| = \max_{i \notin \mathcal{Q}^\star} |\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle|, \tag{3}$$

we consider that Oxx automatically makes the wrong decision. Tropp used this convention for OMP and showed that in the limit case where the upper bound on his ERC condition (see Section III-A) is reached, the limit situation (3) occurs, hence a wrong atom is selected at the first iteration [1].

Let us now properly define the notions of successful support recovery and support recovery failure.

Definition 1 ***[Successful recovery]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. Oxx with $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ as input succeeds if and only if there exists $j \leqslant \mathrm{Card}\,[\mathcal{Q}^\star]$ such that all first j iterations of Oxx select atoms in $\mathcal{Q}^\star$ and the residual $\boldsymbol{r}_{\mathcal{Q}}$ is equal to $\mathbf{0}$ after the j-th iteration.*

In other words, when a successful recovery occurs, the subset yielded by Oxx satisfies $\mathcal{Q}_{\boldsymbol{y}} \subseteq \mathcal{Q} \subseteq \mathcal{Q}^\star$ where $\mathcal{Q}_{\boldsymbol{y}}$ is the "sparsest subset", *i.e.*, the subset of $\mathcal{Q}^\star$ corresponding to the nonzero weights t_i's in the decomposition $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}^\star}\boldsymbol{t}$. When all t_i's are nonzero, $\mathcal{Q}_{\boldsymbol{y}}$ identifies with $\mathcal{Q}^\star$ and a successful recovery coincides with the exact recovery of $\mathcal{Q}^\star$ in k iterations.

The word "failure" refers to the exact contrary of successful recovery.

Definition 2 ***[Failure]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. Oxx with $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ as input fails when at least one wrong atom is selected during the first k iterations. In particular, Oxx fails when* (3) *occurs with $\boldsymbol{r}_{\mathcal{Q}} \neq \mathbf{0}$.*

III. Overview of our recovery analysis of OMP and OLS

In this section, we present our main concepts and results regarding the sparse recovery guarantees with OLS, their connection with the existing OMP results and the new results regarding OMP. For clarity reasons, we place the technical analysis including most of the proofs in the main appendix section A.

Let us first recall Tropp's ERC condition for OMP which is our starting point.

A. Tropp's ERC condition for OMP

Theorem 1 ***[ERC is a sufficient recovery condition for OMP and a necessary condition at the first iteration [1, Theorems 3.1 and 3.10]]*** *If $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank and*

$$F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\text{bad}}) \triangleq \max_{\boldsymbol{a}_{\text{bad}}} \left\| \boldsymbol{A}_{\mathcal{Q}^\star}^\dagger \boldsymbol{a}_{\text{bad}} \right\|_1 < 1, \qquad \text{ERC}(\boldsymbol{A}, \mathcal{Q}^\star)$$

then OMP succeeds for any input $\boldsymbol{y} \in \text{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$. Furthermore, when ERC($\boldsymbol{A}, \mathcal{Q}^\star$) does not hold, there exists $\boldsymbol{y} \in \text{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which some $\boldsymbol{a}_{\text{bad}}$ is selected at the first iteration of OMP. When $\text{spark}(\boldsymbol{A}) > 2k$, this implies that OMP cannot recover the (unique) k-term representation of $\boldsymbol{y}$.

Note that ERC($\boldsymbol{A}, \mathcal{Q}^\star$) only involves the dictionary atoms since it results from a worst case analysis: if ERC($\boldsymbol{A}, \mathcal{Q}^\star$) holds, then a successful recovery occurs with $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}^\star}\boldsymbol{t}$ *whatever* $\boldsymbol{t} \in \mathbb{R}^k$.

B. Main theorem

A theorem similar to Theorem 1 applies to OLS. This is our main contribution.

Theorem 2 ***[ERC is a sufficient recovery condition for OLS and a necessary condition at the first iteration]*** *If $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank and ERC($\boldsymbol{A}, \mathcal{Q}^\star$) holds, then OLS succeeds for any input $\boldsymbol{y} \in \text{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$. Furthermore, when ERC($\boldsymbol{A}, \mathcal{Q}^\star$) does not hold, there exists $\boldsymbol{y} \in \text{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which some $\boldsymbol{a}_{\text{bad}}$ is selected at the first iteration of OLS. When $\text{spark}(\boldsymbol{A}) > 2k$, this implies that OLS cannot recover the (unique) k-term representation of $\boldsymbol{y}$.*

The necessary condition result is obvious since the very first iteration of OLS coincides with that of OMP and ERC is a necessary condition for OMP. The core of our contribution is the proof that ERC is a sufficient condition for the exact recovery with OLS. We now introduce the main concepts on which our OLS analysis relies. They also lead to a more precise analysis of OMP from the second iteration.

C. Main concepts

Let us keep in mind that ERC is a worst case necessary condition *at the first iteration*. But what happens when the ERC is not met but nevertheless, the first j iterations of Oxx select j true atoms ($j < k$)? Can we characterize the exact recovery conditions at the $(j+1)$-th iteration? We will answer to these questions and provide:

1) an extension of the ERC condition to the j-th iteration of OMP;
2) a new necessary and sufficient condition dedicated to the j-th iteration of OLS.

This will allow us to prove Theorem 2 as a special case of the latter condition when $j = 0$.

In the following two paragraphs, we introduce useful notations for a single wrong atom $\boldsymbol{a}_{\mathrm{bad}}$ and then define our new exact recovery conditions by considering all the wrong atoms together. $\mathcal{Q}$ plays the role of the subset found by Oxx after the first j iterations.

1) Notations related to a single wrong atom: For $\mathcal{Q} \subsetneq \mathcal{Q}^\star$, we define:

$$F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}}) \triangleq \sum_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}} \left| \left(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star} \boldsymbol{a}_{\mathrm{bad}} \right)(i) \right| \tag{4}$$

$$F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}}) \triangleq \sum_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}} \frac{\|\tilde{\boldsymbol{a}}_i\|}{\|\tilde{\boldsymbol{a}}_{\mathrm{bad}}\|} \left| \left(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star} \boldsymbol{a}_{\mathrm{bad}} \right)(i) \right| \tag{5}$$

when $\tilde{\boldsymbol{a}}_{\mathrm{bad}} \neq \boldsymbol{0}$ and $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}}) = 0$ when $\tilde{\boldsymbol{a}}_{\mathrm{bad}} = \boldsymbol{0}$ (we recall that $\tilde{\boldsymbol{a}}_i = \boldsymbol{P}^\perp_{\mathcal{Q}} \boldsymbol{a}_i$ and $\tilde{\boldsymbol{a}}_{\mathrm{bad}} = \boldsymbol{P}^\perp_{\mathcal{Q}} \boldsymbol{a}_{\mathrm{bad}}$ depend on $\mathcal{Q}$). Up to some manipulations on orthogonal projections, (4) and (5) can be rewritten as follows.

Lemma 1 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. For $\mathcal{Q} \subsetneq \mathcal{Q}^\star$, $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$ and $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$ also read*

$$F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}}) = \|\tilde{\boldsymbol{A}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{a}}_{\mathrm{bad}}\|_1 \tag{6}$$

$$F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}}) = \|\tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}_{\mathrm{bad}}\|_1 \tag{7}$$

where the matrices $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} = \{\tilde{\boldsymbol{a}}_i,\, i \in \mathcal{Q}^\star \setminus \mathcal{Q}\}$ and $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} = \{\tilde{\boldsymbol{a}}_i,\, i \in \mathcal{Q}^\star \setminus \mathcal{Q}\}$ of size $m \times (k-j)$ are full rank.

Lemma 1 is proved in Appendix B.

2) ERC-Oxx conditions for the whole dictionary: We define four binary conditions by considering all the wrong atoms together:

$$\max_{\boldsymbol{a}_{\text{bad}}} F^{\text{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}}) < 1 \qquad \text{ERC-OMP}(\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q})$$

$$\max_{\boldsymbol{a}_{\text{bad}}} F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}}) < 1 \qquad \text{ERC-OLS}(\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q})$$

$$\max_{\substack{\mathcal{Q}\subsetneq\mathcal{Q}^\star \\ \text{Card}[\mathcal{Q}]=j}} \max_{\boldsymbol{a}_{\text{bad}}} F^{\text{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}}) < 1 \qquad \text{ERC-OMP}(\boldsymbol{A}, \mathcal{Q}^\star, j)$$

$$\max_{\substack{\mathcal{Q}\subsetneq\mathcal{Q}^\star \\ \text{Card}[\mathcal{Q}]=j}} \max_{\boldsymbol{a}_{\text{bad}}} F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}}) < 1 \qquad \text{ERC-OLS}(\boldsymbol{A}, \mathcal{Q}^\star, j)$$

We will use the common notations $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}})$, ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) and ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, j$) for statements that are common to both OMP and OLS.

Remark 1 *$F^{\text{OMP}}_{\mathcal{Q}^\star,\emptyset}(\boldsymbol{a}_{\text{bad}})$ and $F^{\text{OLS}}_{\mathcal{Q}^\star,\emptyset}(\boldsymbol{a}_{\text{bad}})$ both reread $F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\text{bad}}) = \left\|\boldsymbol{A}^\dagger_{\mathcal{Q}^\star}\boldsymbol{a}_{\text{bad}}\right\|_1$ since $\tilde{\boldsymbol{a}}^\emptyset_i$ reduces to $\boldsymbol{a}_i$ which is unitary. Thus, ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \emptyset$) and ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, 0$) all identify with ERC($\boldsymbol{A}, \mathcal{Q}^\star$).*

D. Sufficient conditions of exact recovery at any iteration

The sufficient conditions of Theorems 1 and 2 reread as the special case of the following theorem where $\mathcal{Q} = \emptyset$.

Theorem 3 ***[Sufficient recovery condition for Oxx after*** j **successful *iterations]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. If Oxx with $\boldsymbol{y} \in \text{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ as input selects $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ and ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) holds, then Oxx succeeds in the sense of Definition 1.*

The following corollary is a straightforward adaptation of Theorem 3 to ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, j$).

Corollary 1 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. If Oxx with $\boldsymbol{y} \in \text{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ as input selects true atoms during the first $j \geqslant 0$ iterations and ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, j$) holds, then Oxx succeeds.*

The key element which enables us to establish Theorem 3 is a recursive relation linking $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}})$ with $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_{\text{bad}})$ when $\mathcal{Q}$ is increased by one element of $\mathcal{Q}^\star \backslash \mathcal{Q}$, resulting in subset $\mathcal{Q}'$. This leads to the main technical novelty of the paper, stated in Lemma 7 (see Appendix A-A). From the thorough analysis of this recursive relation, we elaborate the following lemma which guarantees the monotonic decrease of $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}})$ when $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ is growing.

Lemma 2 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. Let $\mathcal{Q} \subsetneq \mathcal{Q}' \subsetneq \mathcal{Q}^\star$. For any $\boldsymbol{a}_{\rm bad}$,*

$$F^{\rm OMP}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_{\rm bad}) \leqslant F^{\rm OMP}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\rm bad}) \tag{8}$$

$$F^{\rm OLS}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\rm bad}) < 1 \Rightarrow F^{\rm OLS}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_{\rm bad}) \leqslant F^{\rm OLS}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\rm bad}) \tag{9}$$

We refer the reader to Appendix A-A for the proof of Lemmas 7 and 2, and then Theorem 3.

E. Necessary conditions of exact recovery at any iteration

We recall that ERC is a worst case necessary condition guaranteed for the selection of a true atom by OMP and OLS in their very first iteration. We provide extended results stating that ERC-Oxx are worst case necessary conditions when the first iterations of Oxx have succeeded, up to a "reachability assumption" defined hereafter, for OMP.

Definition 3 ***[Reachability]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}}$ is full rank. $\mathcal{Q}$ is reachable if and only if there exists an input $\boldsymbol{y} = \boldsymbol{A}_{\mathcal{Q}}\boldsymbol{t}$ where $t_i \neq 0$ for all i, for which Oxx recovers $\mathcal{Q}$ in $\mathrm{Card}\,[\mathcal{Q}]$ iterations. Specifically, the selection rule* (1)-(2) *always yields a unique maximum.*

We start with the OLS condition which is simpler.

1) OLS necessary condition:

Theorem 4 ***[Necessary condition for OLS after j iterations]*** *Let $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ be a subset of cardinality j. Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank and $\mathrm{spark}(\boldsymbol{A}) \geqslant (j+2)$. If ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) does not hold, then there exists $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which OLS selects $\mathcal{Q}$ in the first j iterations and then a wrong atom $\boldsymbol{a}_{\rm bad}$ in the $(j+1)$-th iteration.*

Theorem 4 is proved in Appendix A-B. An obvious corollary can be obtained by replacing $\mathcal{Q}$ with j akin to the derivation of Corollary 1 from Theorem 3. From now on, such obvious corollaries will not be explicitly stated.

2) Reachability issues: The reader may have noticed that Theorem 4 implies that $\mathcal{Q}$ can be reached by OLS at least for some input $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$. In Appendix A-B, we establish a stronger result:

Lemma 3 (Reachability by OLS) *Any subset $\mathcal{Q}$ with $\mathrm{Card}\,[\mathcal{Q}] \leqslant \mathrm{spark}(\boldsymbol{A}) - 2$ can be reached by OLS with some input $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})$.*

The assumption $\mathrm{Card}\,[\mathcal{Q}] \leqslant \mathrm{spark}(\boldsymbol{A}) - 2$ enables us to guarantee that the OLS selection rule (2) always yields a unique maximum (see Appendix A-B).

Perhaps surprisingly, the result of Lemma 3 does not remain valid for OMP although it holds under certain RIP assumptions [12, Theorem 4.1]. As shown in Example 1 hereafter, there are counterexamples where $\mathcal{Q}$ cannot be reached by OMP not only for $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})$ but also for any $\boldsymbol{y} \in \mathbb{R}^m$. The same somewhat surprising phenomenon of non-reachability also occurs with ℓ_1 minimization, associated to certain k-faces of the ℓ_1 ball in $\mathbb{R}^n$ whose projection through $\boldsymbol{A}$ yields interior faces. This result is a direct consequence of the Null Space Property [14].

Example 1 *Consider the simple dictionary*

$$\boldsymbol{A} = \begin{bmatrix} \cos\theta_1 & \cos\theta_1 & 0 & 0 \\ -\sin\theta_1 & \sin\theta_1 & \cos\theta_2 & \cos\theta_2 \\ 0 & 0 & \sin\theta_2 & -\sin\theta_2 \end{bmatrix}$$

with $\mathcal{Q} = \{1, 2\}$. Set θ_2 to an arbitrary value in $(0, \pi/2)$. When $\theta_1 \neq 0$ is close enough to 0, OMP can never reach $\mathcal{Q}$ in two iterations (specifically, when $\boldsymbol{y} \in \mathbb{R}^3$ is proportional to neither $\boldsymbol{a}_1$ nor $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ or $\boldsymbol{a}_4$ is selected in the first two iterations).

This result is proved in Section A-B3. Although in Example 1, a subset of cardinality 2 can never be reached, we remark that for undercomplete dictionaries, any subset of cardinality 2 can be reached for some $\boldsymbol{y} \in \mathbb{R}^m$.

3) OMP necessary conditions including reachability assumptions: Our necessary condition for OMP is somewhat tricky because we must assume that $\mathcal{Q}$ is reachable by OMP using some input in $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})$.

Theorem 5 ***[Necessary condition for OMP after j iterations]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank and $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ is reachable. If ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) does not hold, then there exists $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which OMP selects $\mathcal{Q}$ in the first j iterations and then a wrong atom $\boldsymbol{a}_{\mathrm{bad}}$ in the $(j+1)$-th iteration.*

Theorem 5 is proved together with Theorem 4 in Appendix A-B. Setting aside the reachability issues, the principle of the proof is common to OMP and OLS. We proceed the proof of the sufficient condition (Theorem 3) backwards, as was done in [1, Theorem 3.10] in the case $\mathcal{Q} = \emptyset$.

In the special case where $j = 1$, Theorem 5 simplifies to a corollary similar to the OLS necessary condition (Theorem 4) because any subset $\mathcal{Q}$ of cardinality 1 is obviously reachable using the atom indexed by $\mathcal{Q}$ as input vector.

Corollary 2 ***[Necessary condition for OMP in the second iteration]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank and let $i \in \mathcal{Q}^\star$. If ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \{i\}$) does not hold, then there exists $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which OMP selects $\boldsymbol{a}_i$ and then a wrong atom $\boldsymbol{a}_{\mathrm{bad}}$ in the first two iterations.*

4) Discrimination between OMP and OLS at the k-th iteration: We provide an element of discrimination between OMP and OLS when their first $k-1$ iterations have selected true atoms, so that there is one remaining true atom which has not been chosen. Let us first observe that in Example 1, OMP is not guaranteed to select the second true atom when $\boldsymbol{a}_1$ or $\boldsymbol{a}_2$ has already been chosen. This is actually a major difference with OLS.

Theorem 6 ***[Guaranteed success of the k-th iteration of OLS]*** *If $[\boldsymbol{A}_{\mathcal{Q}^\star}, \boldsymbol{a}_{\mathrm{bad}}]$ is full rank for any $\boldsymbol{a}_{\mathrm{bad}}$, then ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, k-1$) is true. Thus, if the first $k-1$ iterations of OLS select true atoms, the last true atom is necessarily selected in the k-th iteration.*

This result is straightforward from the definition of OLS in the optimization viewpoint: "OLS selects the new atom yielding the least possible residual" and the remark that in the k-th iteration, the last true atom yields a zero valued residual. Another (analytical) proof of Theorem 6, given below, is based on the definition of ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, k-1$). It will enable us to understand why the statement of Theorem 6 is not valid for OMP.

Proof: Assume that OLS yields a subset $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ after $k-1$ iterations. Let $\boldsymbol{a}_{\mathrm{last}}$ denote the last true atom so that $\boldsymbol{A}_{\mathcal{Q}^\star} = [\boldsymbol{A}_{\mathcal{Q}}, \boldsymbol{a}_{\mathrm{last}}]$ up to some permutation of columns. Since $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}$ reduces to $\tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{last}}$ and because $\tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{last}}$ is unitary, the pseudo-inverse $\tilde{\boldsymbol{B}}^{\dagger}_{\mathcal{Q}^\star \setminus \mathcal{Q}}$ takes the form $\left[\tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{last}}\right]^t$. Finally, (7) simplifies to:

$$F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}}) = |\langle \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{last}}, \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{bad}} \rangle| \leqslant 1 \tag{10}$$

since both vectors in the inner product are either unitary or equal to $\boldsymbol{0}$. Apply Lemma 8 in Appendix B: since $[\boldsymbol{A}_{\mathcal{Q}^\star}, \boldsymbol{a}_{\mathrm{bad}}]$ is full rank, $\left[\tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{last}}, \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\mathrm{bad}}\right]$ is full rank, thus (10) is a strict inequality. ■

Similar to the calculation in the proof above, we rewrite $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$ defined in (6):

$$F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}}) = \frac{|\langle \tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}, \tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{bad}} \rangle|}{\|\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}\|^2}. \tag{11}$$

However, we cannot ensure that $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}}) \leqslant 1$ since $\tilde{\boldsymbol{a}}^{\mathcal{Q}}_i$ are not unitary vectors.

To further distinguish OMP and OLS, we elaborate a "bad recovery condition" under which OMP is guaranteed to fail in the sense that $\mathcal{Q}^\star$ is not reachable.

Theorem 7 ***[Sufficient condition for bad recovery with OMP]*** *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. If*

$$\min_{\substack{\mathcal{Q}\subsetneq\mathcal{Q}^\star \\ \mathrm{Card}[\mathcal{Q}]=k-1}} \left[\max_{\boldsymbol{a}_{\mathrm{bad}}} F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})\right] \geqslant 1, \qquad \text{BRC-OMP}(\boldsymbol{A},\mathcal{Q}^\star)$$

then $\mathcal{Q}^\star$ cannot be reached by OMP using any input in $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$.

Specifically, BRC-OMP($\boldsymbol{A},\mathcal{Q}^\star$) guarantees that a wrong selection occurs at the k-th iteration when the previous iterations have succeeded.

Proof: Assume that for some $\boldsymbol{y}\in\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$, the first $k-1$ iterations of OMP succeed, *i.e.*, they select $\mathcal{Q}\subsetneq\mathcal{Q}^\star$ of cardinality $k-1$. Let $\boldsymbol{a}_{\mathrm{last}}$ denote the last true atom ($\boldsymbol{A}_{\mathcal{Q}^\star}=[\boldsymbol{A}_{\mathcal{Q}},\boldsymbol{a}_{\mathrm{last}}]$ up to some permutation of columns). The residual $\boldsymbol{r}_{\mathcal{Q}}$ yielded by OMP after $k-1$ iterations is obviously proportional to $\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}$.

BRC-OMP($\boldsymbol{A},\mathcal{Q}^\star$) implies that ERC-OMP($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$) is false, thus there exists $\boldsymbol{a}_{\mathrm{bad}}\notin\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})$ such that $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})\geqslant 1$. According to (11), $|\langle\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}},\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{bad}}\rangle|\geqslant\|\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}\|^2$ thus $|\langle\boldsymbol{r}_{\mathcal{Q}},\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{bad}}\rangle|\geqslant|\langle\boldsymbol{r}_{\mathcal{Q}},\tilde{\boldsymbol{a}}^{\mathcal{Q}}_{\mathrm{last}}\rangle|$. We conclude that $\boldsymbol{a}_{\mathrm{last}}$ cannot be chosen in the k-th iteration of OMP. ■

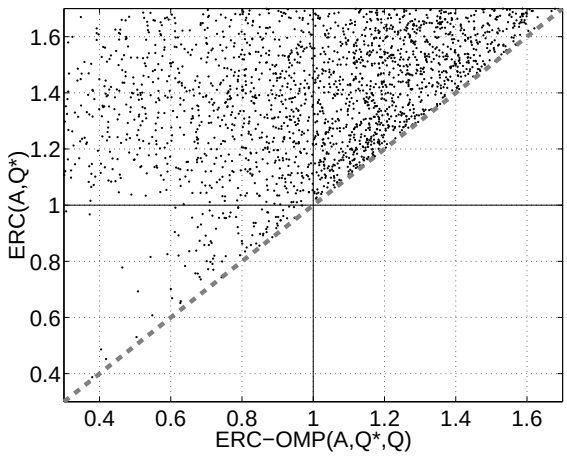
Although BRC-OMP($\boldsymbol{A},\mathcal{Q}^\star$) may appear restrictive (as a minimum is involved in the left-hand side), we will see in Section IV that it may frequently be met, even when the atoms of $\boldsymbol{A}$ are not strongly correlated.

IV. Empirical comparison of the OMP and OLS exact recovery conditions

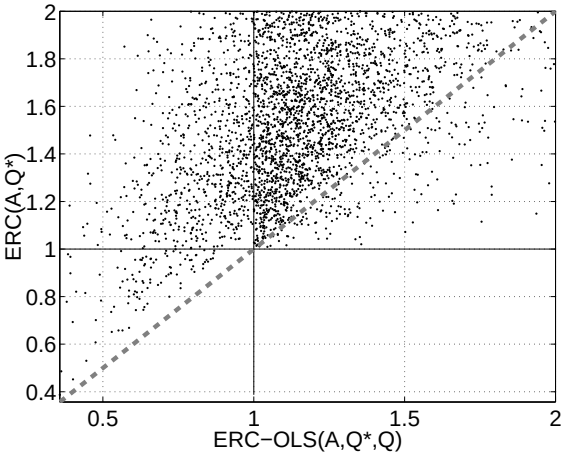
The purpose of this section is to test whether there is some systematic implication between the conditions ERC-OMP($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$) and ERC-OLS($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$), and between ERC-OMP($\boldsymbol{A},\mathcal{Q}^\star,j$) and ERC-OLS($\boldsymbol{A},\mathcal{Q}^\star,j$). We set $j=\mathrm{Card}\,[\mathcal{Q}]=1$. Additionally, we will emphasize the distinction between OMP and OLS by evaluating the bad recovery condition for OMP. These empirical comparisons involve Matlab simulations with random dictionaries.

A. Comparison of the ERC-Oxx conditions

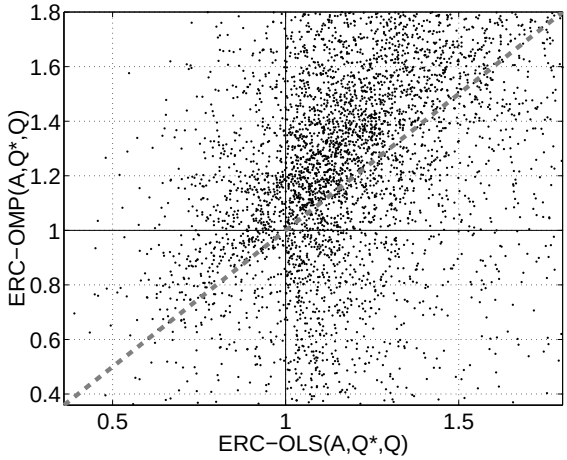
We compare ERC-OMP($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$) and ERC-OLS($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$) for a common dictionary and a given pair of subsets where $\mathcal{Q}\subsetneq\mathcal{Q}^\star$ is of cardinality 1. As the recovery conditions take the form "for all $\boldsymbol{a}_{\mathrm{bad}}$, $F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})<1$", it is sufficient to just consider the case where there is one wrong atom $\boldsymbol{a}_{\mathrm{bad}}$. Therefore, we consider dictionaries $\boldsymbol{A}$ with $k+1$ atoms, with $k=\mathrm{Card}\,[\mathcal{Q}^\star]$. Evaluating ERC($\boldsymbol{A},\mathcal{Q}^\star$), ERC-OMP($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$) and ERC-OLS($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$) amounts to computing $F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\mathrm{bad}})$, $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$ and $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$ and to testing whether they are lower than 1.



(a) $F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\mathrm{bad}})$ *vs* $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$.



(b) $F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\mathrm{bad}})$ *vs* $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$.



(c) $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$ *vs* $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$.

Fig. 1. Comparison of the OMP and OLS exact recovery conditions. We draw 10.000 Gaussian dictionaries of size 100×11 and set $k=10$ so that there is only one wrong atom $\boldsymbol{a}_{\mathrm{bad}}$. $\mathcal{Q}$ is always set to the first atom ($\mathrm{Card}\,[\mathcal{Q}]=1$). Plot of (a) $F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\mathrm{bad}})$ *vs* $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$; (b) $F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\mathrm{bad}})$ *vs* $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$; (c) $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$ *vs* $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$. For the last subfigure, we keep the trials for which $F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\mathrm{bad}})\geqslant 1$.

Fig. 1 is a scatter plot of the three criteria for 10.000 Gaussian dictionaries $\boldsymbol{A}$ of size 100×11, where the elements of $\boldsymbol{A}$ are drawn according to an i.i.d. Gaussian distribution. The subset $\mathcal{Q}=\{1\}$ is systematically chosen as the first atom of $\boldsymbol{A}$. Figs. 1(a,b) are in good agreement with Lemma 2: we verify that $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})\leqslant F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\mathrm{bad}})$ whether ERC holds or not, and that $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})\leqslant F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\mathrm{bad}})$ systematically occurs only when $F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\mathrm{bad}})<1$. On Fig. 1(c) displaying $F^{\mathrm{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$ versus $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}})$, we only keep the trials for which $F_{\mathcal{Q}^\star}(\boldsymbol{a}_{\mathrm{bad}})\geqslant 1$, *i.e.*, ERC($\boldsymbol{A},\mathcal{Q}^\star$) does not hold. Since both south-east and north-west quarter planes are populated, we conclude that neither OMP nor OLS is uniformly better. To

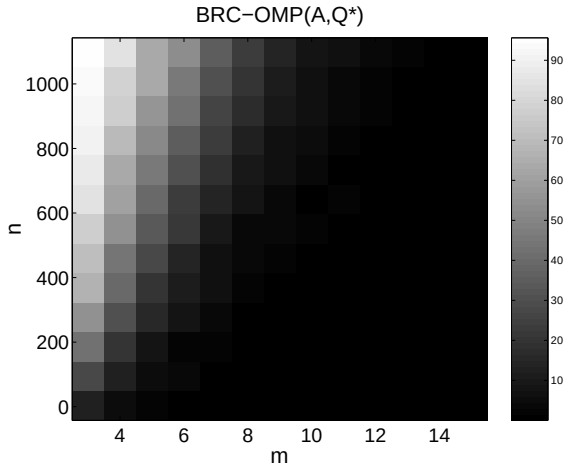


Fig. 2. Computation of the bad recovery condition BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) for Gaussian dictionaries of various sizes (m, n). 1,000 trials are performed for each size, and $\mathcal{Q}^\star$ is always set to the first two atoms ($k = 2$). The grey levels in the image correspond to the rate of guaranteed failure, *i.e.*, the proportion of trials where BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) holds.

be more specific, when ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) holds but ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) does not, there exists an input $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ for which OLS selects $\mathcal{Q} = \{1\}$ and then a wrong atom in the first two iterations (Theorem 4). On the contrary, OMP is guaranteed perform an exact recovery with this input according to Theorem 3. The same situation can occur when inverting the roles of OMP and OLS according to Corollary 2 and Theorem 3.

We have compared ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, 1$) and ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, 1$) which take into account all the possible subsets of $\mathcal{Q}^\star$ of cardinality 1. Again, we found that when ERC($\boldsymbol{A}, \mathcal{Q}^\star$) is not met, it can occur that ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, 1$) holds while ERC-OLS($\boldsymbol{A}, \mathcal{Q}^\star, 1$) does not and *vice versa*.

Note that this analysis becomes more complex when $\mathrm{Card}\,[\mathcal{Q}] \geqslant 2$ since ERC-OMP($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) alone is not a necessary condition for OMP anymore (Theorem 5 also involves the assumption that $\mathcal{Q}$ is reachable).

B. Discrimination at the second iteration

Because the above simulation cannot discriminate OMP and OLS, we consider the bad recovery condition BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) under which OMP is guaranteed to fail when k iterations are performed. Meanwhile, OLS recovers $\mathcal{Q}^\star$ at least for some input in $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$. Moreover, the k-th iteration of OLS is guaranteed to succeed provided that the first $k-1$ iterations have succeeded according to Theorem 6.

We compute BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) in the case $k = 2$ for various dictionary sizes (see Fig. 2). We

perform 1,000 trials per size (m, n) in which the elements of $\boldsymbol{A}$ are drawn according to an i.i.d. Gaussian distribution and $\mathcal{Q}^\star$ is always set to the first two atoms. We notice that BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) may frequently be met for overcomplete dictionaries, especially when m is low and $n \gg m$. Because $k = 2$, OLS performs at least as good as OMP: when the first iteration (common to both algorithms) has succeeded, OLS cannot fail according to Theorem 6 while OMP is guaranteed to fail in cases where the BRC holds.

This simulation can naturally be extended to the case $k > 2$ but the conclusions differ. OLS is not guaranteed to outperform OMP for any $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$, but when BRC-OMP($\boldsymbol{A}, \mathcal{Q}^\star$) is not met, OLS recovers $\mathcal{Q}^\star$ for some inputs while OMP cannot for any input.

V. Conclusions

Our first contribution is an original analysis of OLS based on the extension of Tropp's ERC condition. We showed that when ERC holds, OLS is guaranteed to yield an exact support recovery. Although OLS has been acknowledged in several communities for two decades, such a theoretical analysis was lacking. Our second contribution is a parallel study of OMP and OLS when a number of iterations have been performed and true atoms have been selected. We found that neither OMP nor OLS is uniformly better. In particular, we showed using simulated dictionaries that when the ERC is not met but the first iteration (which is common to OMP and OLS) selects a true atom, there are counter-examples for which OMP is guaranteed to yield an exact support recovery while OLS does not, and *vice versa*.

Finally, a few elements of analysis suggest that OLS behaves better than OMP. First, any subset $\mathcal{Q}$ can be reached by OLS using some input in $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})$ while for some dictionaries, it may occur that some subsets are never reached by OMP for any $\boldsymbol{y} \in \mathbb{R}^m$. In other words, OLS has a stronger capability of exploration. Secondly, when all true atoms except one have been found by OLS and no wrong selection occurred, OLS is guaranteed to find the last true atom in the following iteration while OMP may fail.

For realistic problems where the data are noisy and the dictionary is far from orthogonal, empirical studies report that OLS usually outperforms OMP for a larger numerical cost [9, 11]. In our experience, OLS yields a residual error which may be by far lower than that of OMP after the same number of iterations [15]. Moreover, it performs better support recoveries in terms of ratio between the number of good detections and of false alarms [16]. We believe that the reason why our exact recovery analysis does not corroborate this trend is that it is essentially based on a worst case analysis. An interesting perspective will consist of a theoretical study in the average case in order to evaluate more thoroughly the difference between OMP and OLS.

APPENDIX A
NECESSARY AND SUFFICIENT CONDITIONS OF EXACT RECOVERY FOR OMP AND OLS

This appendix includes the complete analysis of our OMP and OLS recovery conditions.

A. Sufficient conditions

We show that when Oxx happens to select true atoms during its early iterations, it is guaranteed to recover the whole unknown support in the subsequent iterations when the ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) condition is fulfilled. We establish Theorem 3 whose direct consequence is Theorem 2 stating that when ERC($\boldsymbol{A}, \mathcal{Q}^\star$) holds, OLS is guaranteed to succeed.

1) ERC-Oxx are sufficient recovery conditions at a given iteration: We follow the analysis of [1, Theorem 3.1] to extend Tropp's exact recovery condition to a sufficient condition dedicated to the $(j+1)$-th iteration of Oxx.

Lemma 4 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. If Oxx with $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ as input selects j true atoms $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ and ERC-Oxx($\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$) holds, then the $(j+1)$-th iteration of Oxx selects a true atom.*

Proof: According to the selection rule (1)-(2), Oxx selects a true atom at iteration $(j+1)$ *if and only if*

$$\phi(\boldsymbol{r}_{\mathcal{Q}}) \triangleq \frac{\max_{i \notin \mathcal{Q}^\star} |\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle|}{\max_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}} |\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle|} < 1. \tag{12}$$

Let us gather the vectors $\tilde{\boldsymbol{c}}_i$ indexed by $i \notin \mathcal{Q}^\star$ and $i \in \mathcal{Q}^\star \setminus \mathcal{Q}$ in two matrices $\tilde{\boldsymbol{C}}_{\mathrm{bad}}$ and $\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}$ of dimensions $m \times (n-k)$ and $m \times (k-j)$, respectively. The condition (12) rereads:

$$\phi(\boldsymbol{r}_{\mathcal{Q}}) = \frac{\|\tilde{\boldsymbol{C}}_{\mathrm{bad}}^t \boldsymbol{r}_{\mathcal{Q}}\|_\infty}{\|\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}^t \boldsymbol{r}_{\mathcal{Q}}\|_\infty} < 1.$$

Following Tropp's analysis, we re-arrange the vector $\boldsymbol{r}_{\mathcal{Q}}$ occurring in the numerator. Since $\boldsymbol{r}_{\mathcal{Q}} = \boldsymbol{P}_{\mathcal{Q}}^{\perp} \boldsymbol{y}$ and $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$, $\boldsymbol{r}_{\mathcal{Q}} \in \mathrm{span}(\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}) = \mathrm{span}(\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star \setminus \mathcal{Q}})$. We rewrite $\boldsymbol{r}_{\mathcal{Q}}$ as $\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \boldsymbol{r}_{\mathcal{Q}}$ where $\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}$ stands for the orthogonal projection on $\mathrm{span}(\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star \setminus \mathcal{Q}})$: $\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} = \tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}^t = \left(\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{C}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}^\dagger\right)^t$. $\phi(\boldsymbol{r}_{\mathcal{Q}})$ rereads

$$\phi(\boldsymbol{r}_{\mathcal{Q}}) = \frac{\|\left(\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}^\dagger \tilde{\boldsymbol{C}}_{\mathrm{bad}}\right)^t \tilde{\boldsymbol{C}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}^t \boldsymbol{r}_{\mathcal{Q}}\|_\infty}{\|\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}^t \boldsymbol{r}_{\mathcal{Q}}\|_\infty}.$$

This expression can obviously be majorized using the matrix norm:

$$\phi(\boldsymbol{r}_{\mathcal{Q}}) \leqslant \|\left(\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}^\dagger \tilde{\boldsymbol{C}}_{\mathrm{bad}}\right)^t\|_{\infty,\infty}. \tag{13}$$

Since the ℓ_∞ norm of a matrix is equal to the ℓ_1 norm of its transpose and $\|\,.\,\|_{1,1}$ equals the maximum column sum of the absolute value of its argument [1, Theorem 3.1], the upper bound of (13) rereads

$$\|\tilde{\boldsymbol{C}}^{\dagger}_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\boldsymbol{C}}_{\rm bad}\|_{1,1} = \max_{\tilde{\boldsymbol{c}}_{\rm bad}} \|\tilde{\boldsymbol{C}}^{\dagger}_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\boldsymbol{c}}_{\rm bad}\|_1 = \max_{\boldsymbol{a}_{\rm bad}} F^{\rm Oxx}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\rm bad})$$

according to Lemma 1.

By definition of ERC-Oxx($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$), this upper bound is lower than 1 thus $\phi(\boldsymbol{r}_\mathcal{Q}) < 1$. According to (12), Oxx selects a true atom. ■

2) Recursive expression of the ERC-Oxx formulas: We elaborate recursive expressions of $F^{\rm Oxx}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\rm bad})$ when $\mathcal{Q}$ is increased by one element resulting in the new subset $\mathcal{Q}' \subsetneq \mathcal{Q}^\star$ (here, we do not consider the case where $\mathcal{Q}' = \mathcal{Q}^\star$ since $F^{\rm Oxx}_{\mathcal{Q}^\star,\mathcal{Q}^\star}(\boldsymbol{a}_{\rm bad})$ is not properly defined, (4) and (5) being empty sums). We will use the notations $\mathcal{Q}' = \mathcal{Q}\cup\{i_{\rm new}\}$ where $i_{\rm new}\in\mathcal{Q}^\star\setminus\mathcal{Q}$ and $\boldsymbol{a}_{\rm new} \triangleq \boldsymbol{a}_{i_{\rm new}}$. To avoid any confusion, $\tilde{\boldsymbol{a}}_i$ will be systematically replaced by $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}}$ and $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'}$ to express the dependence upon $\mathcal{Q}$ and $\mathcal{Q}'$, respectively. In the same way, $\tilde{\boldsymbol{b}}_i$ will be replaced by $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}}$ or $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'}$ but for simplicity, we will keep the matrix notations $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star\setminus\mathcal{Q}}$ and $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star\setminus\mathcal{Q}'}$ without superscript, $\tilde{}$ referring to $\mathcal{Q}$ and $\mathcal{Q}'$, respectively.

Let us first link $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}}$ to $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'}$ when $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'} \neq \boldsymbol{0}$.

Lemma 5 *Assume that $\boldsymbol{A}_{\mathcal{Q}'}$ is full rank and $\mathcal{Q}' = \mathcal{Q}\cup\{i_{\rm new}\}$. Then, $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})^\perp$ is the orthogonal direct sum of the subspaces $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$ and $\mathrm{span}(\tilde{\boldsymbol{a}}_{\rm new}^{\mathcal{Q}})$, and the normalized projection of any atom $\boldsymbol{a}_i \notin \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'})$ takes the form:*

$$\tilde{\boldsymbol{b}}_i^{\mathcal{Q}} = \eta_i^{\mathcal{Q},\mathcal{Q}'}\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'} + \chi_i^{\mathcal{Q},\mathcal{Q}'}\tilde{\boldsymbol{b}}_{\rm new}^{\mathcal{Q}} \tag{14}$$

where

$$\eta_i^{\mathcal{Q},\mathcal{Q}'} = \frac{\|\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'}\|}{\|\tilde{\boldsymbol{a}}_i^{\mathcal{Q}}\|} \in (0,1], \tag{15}$$

$$\chi_i^{\mathcal{Q},\mathcal{Q}'} = \langle\tilde{\boldsymbol{b}}_i^{\mathcal{Q}}, \tilde{\boldsymbol{b}}_{\rm new}^{\mathcal{Q}}\rangle, \tag{16}$$

$$\left(\eta_i^{\mathcal{Q},\mathcal{Q}'}\right)^2 + \left(\chi_i^{\mathcal{Q},\mathcal{Q}'}\right)^2 = 1. \tag{17}$$

Proof: Since $\mathcal{Q}\subsetneq\mathcal{Q}'$, we have $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp \subseteq \mathrm{span}(\boldsymbol{A}_\mathcal{Q})^\perp$. Because $\boldsymbol{A}_{\mathcal{Q}'}$ is full rank, $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$ and $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})^\perp$ are of consecutive dimensions. Moreover, $\tilde{\boldsymbol{a}}_{\rm new}^{\mathcal{Q}} = \boldsymbol{a}_{\rm new} - \boldsymbol{P}_\mathcal{Q}\boldsymbol{a}_{\rm new} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'}) \cap \mathrm{span}(\boldsymbol{A}_\mathcal{Q})^\perp$, and $\tilde{\boldsymbol{a}}_{\rm new}^{\mathcal{Q}} \neq \boldsymbol{0}$ since $\boldsymbol{A}_{\mathcal{Q}'}$ is full rank. As a vector of $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'})$, $\tilde{\boldsymbol{a}}_{\rm new}^{\mathcal{Q}}$ is orthogonal to $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$. It follows that $\mathrm{span}(\tilde{\boldsymbol{a}}_{\rm new}^{\mathcal{Q}})$ is the orthogonal complement of $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$ in $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})^\perp$.

The orthogonal decomposition of $\tilde{\boldsymbol{a}}_i = \boldsymbol{P}_\mathcal{Q}^\perp\boldsymbol{a}_i$ reads:

$$\tilde{\boldsymbol{a}}_i^{\mathcal{Q}} = \tilde{\boldsymbol{a}}_i^{\mathcal{Q}'} + \langle\tilde{\boldsymbol{a}}_i^{\mathcal{Q}}, \tilde{\boldsymbol{b}}_{\rm new}^{\mathcal{Q}}\rangle\tilde{\boldsymbol{b}}_{\rm new}^{\mathcal{Q}}$$

since $\tilde{\boldsymbol{b}}_{\text{new}}^{\mathcal{Q}}$ is unitary. Replacing $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}} = \|\tilde{\boldsymbol{a}}_i^{\mathcal{Q}}\| \, \tilde{\boldsymbol{b}}_i^{\mathcal{Q}}$ and $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'} = \|\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'}\| \, \tilde{\boldsymbol{b}}_i^{\mathcal{Q}'}$ yields (14)-(16). Pythagoras' theorem yields (17). The assumption $\boldsymbol{a}_i \notin \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})$ implies that $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'} \neq \boldsymbol{0}$, then $\eta_i^{\mathcal{Q},\mathcal{Q}'} > 0$. ■

Lemma 6 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. Let $\mathcal{Q} \subsetneq \mathcal{Q}' \subsetneq \mathcal{Q}^\star$ with $\mathcal{Q}' = \mathcal{Q} \cup \{i_{\text{new}}\}$. Then,* $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}})$ *is the orthogonal direct sum of* $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'})$ *and* $\operatorname{span}(\tilde{\boldsymbol{b}}_{\text{new}}^{\mathcal{Q}})$.

Proof: According to Corollary 8 in Appendix B, $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}$ and $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'}$ are full rank matrices, thus their column spans are of consecutive cardinalities. Lemma 5 states that $\tilde{\boldsymbol{b}}_{\text{new}}^{\mathcal{Q}}$ is orthogonal to $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$, thus it is orthogonal to $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'} \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})^\perp$ for all $i \in \mathcal{Q}^\star \setminus \mathcal{Q}'$. ■

In the following lemma, we establish a link between $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}})$ and $F^{\text{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_{\text{bad}})$. It is a simple recursive relation in the case of OMP. For OLS, we cannot directly relate the two quantities but we express $F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}}) = \big\|\tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\text{bad}}\big\|_1$ with respect to $\tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}'} \tilde{\boldsymbol{b}}^{\mathcal{Q}'}_{\text{bad}}$.

Lemma 7 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. Let $\mathcal{Q} \subsetneq \mathcal{Q}' \subsetneq \mathcal{Q}^\star$ with $\mathcal{Q}' = \mathcal{Q} \cup \{i_{\text{new}}\}$. When $\boldsymbol{a}_{\text{bad}} \notin \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})$,*

$$F^{\text{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}}) = F^{\text{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_{\text{bad}}) + \left|\big(\boldsymbol{A}^\dagger_{\mathcal{Q}^\star} \boldsymbol{a}_{\text{bad}}\big)(i_{\text{new}})\right| \tag{18}$$

$$F^{\text{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}}) = \left|\chi^{\mathcal{Q},\mathcal{Q}'}_{\text{bad}} - \eta^{\mathcal{Q},\mathcal{Q}'}_{\text{bad}} \sum_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}'} \frac{\boldsymbol{\beta}^{\mathcal{Q}^\star \setminus \mathcal{Q}'}_{\text{bad}}(i) \chi^{\mathcal{Q},\mathcal{Q}'}_i}{\eta^{\mathcal{Q},\mathcal{Q}'}_i}\right| + \eta^{\mathcal{Q},\mathcal{Q}'}_{\text{bad}} \sum_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}'} \frac{\big|\boldsymbol{\beta}^{\mathcal{Q}^\star \setminus \mathcal{Q}'}_{\text{bad}}(i)\big|}{\eta^{\mathcal{Q},\mathcal{Q}'}_i} \tag{19}$$

where $\eta_i^{\mathcal{Q},\mathcal{Q}'}$ and $\chi_i^{\mathcal{Q},\mathcal{Q}'}$ are defined in (15)-(16) *and $\boldsymbol{\beta}^{\mathcal{Q}^\star \setminus \mathcal{Q}'}_{\text{bad}} \triangleq \tilde{\boldsymbol{B}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}'} \tilde{\boldsymbol{b}}^{\mathcal{Q}'}_{\text{bad}}$.*

Proof: (18) straightforwardly follows from the definition (4) of $F^{\text{OMP}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\text{bad}})$.

Let us now establish (19). We denote by $\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}$ and $\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'}$ the orthogonal projectors on $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}})$ and $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'})$. Because $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}})$ is the orthogonal direct sum of $\operatorname{span}(\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'})$ and $\operatorname{span}(\tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\text{new}})$ (Lemma 6), we have the orthogonal decomposition:

$$\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\text{bad}} = \tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'} \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\text{bad}} + \chi^{\mathcal{Q},\mathcal{Q}'}_{\text{bad}} \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\text{new}}.$$

(14) yields

$$\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\text{bad}} = \eta^{\mathcal{Q},\mathcal{Q}'}_{\text{bad}} \tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'} \tilde{\boldsymbol{b}}^{\mathcal{Q}'}_{\text{bad}} + \chi^{\mathcal{Q},\mathcal{Q}'}_{\text{bad}} \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\text{new}}$$

($\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}'} \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\text{new}} = \boldsymbol{0}$ according to Lemma 6) and then

$$\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\text{bad}} = \eta^{\mathcal{Q},\mathcal{Q}'}_{\text{bad}} \sum_{i \in \mathcal{Q}^\star \setminus \mathcal{Q}'} \boldsymbol{\beta}^{\mathcal{Q}^\star \setminus \mathcal{Q}'}_{\text{bad}}(i) \tilde{\boldsymbol{b}}^{\mathcal{Q}'}_i + \chi^{\mathcal{Q},\mathcal{Q}'}_{\text{bad}} \tilde{\boldsymbol{b}}^{\mathcal{Q}}_{\text{new}}$$

by definition of $\boldsymbol{\beta}_{\rm bad}^{\mathcal{Q}^\star\backslash\mathcal{Q}'}$. In the latter equation, we re-express $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'}$ with respect to $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}}$ using (14):

$$\tilde{\boldsymbol{P}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}\tilde{\boldsymbol{b}}_{\rm bad}^{\mathcal{Q}} = \eta_{\rm bad}^{\mathcal{Q},\mathcal{Q}'}\sum_{i\in\mathcal{Q}^\star\backslash\mathcal{Q}'}\frac{\boldsymbol{\beta}_{\rm bad}^{\mathcal{Q}^\star\backslash\mathcal{Q}'}(i)}{\eta_i^{\mathcal{Q},\mathcal{Q}'}}\,\tilde{\boldsymbol{b}}_i^{\mathcal{Q}} + \left\{\chi_{\rm bad}^{\mathcal{Q},\mathcal{Q}'} - \eta_{\rm bad}^{\mathcal{Q},\mathcal{Q}'}\sum_{i\in\mathcal{Q}^\star\backslash\mathcal{Q}'}\frac{\boldsymbol{\beta}_{\rm bad}^{\mathcal{Q}^\star\backslash\mathcal{Q}'}(i)\chi_i^{\mathcal{Q},\mathcal{Q}'}}{\eta_i^{\mathcal{Q},\mathcal{Q}'}}\right\}\tilde{\boldsymbol{b}}_{\rm new}^{\mathcal{Q}}.$$

We conclude that $F_{\mathcal{Q}^\star,\mathcal{Q}}^{\rm OLS}(\boldsymbol{a}_{\rm bad}) = \big\|\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}^{\dagger}\tilde{\boldsymbol{b}}_{\rm bad}^{\mathcal{Q}}\big\|_1$ reads (19). ■

3) ERC is a sufficient recovery condition for OLS: The key result of Lemma 2 (see Section III-D) states that $F_{\mathcal{Q}^\star,\mathcal{Q}}^{\rm OLS}(\boldsymbol{a}_{\rm bad})$ is decreasing when $\mathcal{Q}\subsetneq\mathcal{Q}^\star$ is growing provided that $F_{\mathcal{Q}^\star,\mathcal{Q}}^{\rm OLS}(\boldsymbol{a}_{\rm bad})<1$, and that $F_{\mathcal{Q}^\star,\mathcal{Q}}^{\rm OMP}(\boldsymbol{a}_{\rm bad})$ is always decreasing.

Proof of Lemma 2: It is sufficient to prove the result when $\mathrm{Card}\,[\mathcal{Q}'] = \mathrm{Card}\,[\mathcal{Q}]+1$. The case $\mathrm{Card}\,[\mathcal{Q}'] > \mathrm{Card}\,[\mathcal{Q}]+1$ obviously deduces from the former case by recursion.

Let $\mathcal{Q}\subsetneq\mathcal{Q}'\subsetneq\mathcal{Q}^\star$ with $\mathrm{Card}\,[\mathcal{Q}'] = \mathrm{Card}\,[\mathcal{Q}]+1$. The result is obvious when $\boldsymbol{a}_{\rm bad}\in\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'})$: $\tilde{\boldsymbol{a}}_{\rm bad}=\boldsymbol{0}$ then $F_{\mathcal{Q}^\star,\mathcal{Q}'}^{\rm Oxx}(\boldsymbol{a}_{\rm bad})=0$. When $\boldsymbol{a}_{\rm bad}\notin\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}'})$, (8) obviously deduces from (18). The proof of (9) relies on the study of function $\varphi(\eta)=|\sqrt{1-\eta^2}-C\eta|+D\eta$ which is fully defined in (25), (26) and (27) in Appendix C. Because this study is rather technical, we place it in Appendix C.

We notice that $F_{\mathcal{Q}^\star,\mathcal{Q}}^{\rm OLS}(\boldsymbol{a}_{\rm bad})$ given in (19) takes the form $\varphi\big(\eta_{\rm bad}^{\mathcal{Q},\mathcal{Q}'}\big)$ where the variables occurring in C and D (see (26) and (27)) are set to $N\leftarrow\mathrm{Card}\,[\mathcal{Q}^\star\backslash\mathcal{Q}']$, $\eta_i\leftarrow\eta_i^{\mathcal{Q},\mathcal{Q}'}$, $\chi_i\leftarrow\chi_i^{\mathcal{Q},\mathcal{Q}'}$, and $\boldsymbol{\beta}\leftarrow\mathrm{sgn}\big(\chi_{\rm bad}^{\mathcal{Q},\mathcal{Q}'}\big)\boldsymbol{\beta}_{\rm bad}^{\mathcal{Q}^\star\backslash\mathcal{Q}'}$. Now, we invoke Lemma 14 in Appendix C: as $F_{\mathcal{Q}^\star,\mathcal{Q}'}^{\rm OLS}(\boldsymbol{a}_{\rm bad}) = \big\|\boldsymbol{\beta}_{\rm bad}^{\mathcal{Q}^\star\backslash\mathcal{Q}'}\big\|_1$ plays the role of $\|\boldsymbol{\beta}\|_1$, $F_{\mathcal{Q}^\star,\mathcal{Q}}^{\rm OLS}(\boldsymbol{a}_{\rm bad})<1$ implies that $F_{\mathcal{Q}^\star,\mathcal{Q}'}^{\rm OLS}(\boldsymbol{a}_{\rm bad})\leqslant F_{\mathcal{Q}^\star,\mathcal{Q}}^{\rm OLS}(\boldsymbol{a}_{\rm bad})$. ■

We deduce from Lemmas 2 and 4 that ERC-Oxx($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$) are sufficient recovery conditions when $\mathcal{Q}\subsetneq\mathcal{Q}^\star$ has been reached (Theorem 3).

Proof of Theorem 3: Apply Lemma 4 at each iteration $j,\ldots,k-1$ until the increased subset $\mathcal{Q}'$ matches $\mathcal{Q}^\star$. The ERC-Oxx($\boldsymbol{A},\mathcal{Q}^\star,\,.\,$) assumption of Lemma 4 is always fulfilled according to Lemma 2. ■

Finally, we prove that ERC($\boldsymbol{A},\mathcal{Q}^\star$) is a necessary and sufficient condition of successful recovery for OLS (Theorem 2).

Proof of Theorem 2: The sufficient condition is a special case of Theorem 3 for $\mathcal{Q}=\emptyset$. The necessary condition identifies with that of Theorem 1 since ERC-OLS($\boldsymbol{A},\mathcal{Q}^\star,\emptyset$) simplifies to ERC($\boldsymbol{A},\mathcal{Q}^\star$). ■

B. Necessary conditions

We provide the technical analysis to prove that ERC-Oxx($\boldsymbol{A},\mathcal{Q}^\star,\mathcal{Q}$) is not only a sufficient condition of exact recovery in the worst case when $\mathcal{Q}\subsetneq\mathcal{Q}^\star$ has been reached, but also a necessary condition. We will prove Theorems 4 and 5 (see Section III) generalizing Tropp's necessary condition [1, Theorem 3.10] to any iteration of OMP and OLS.

We proceed in two stages. In the first stage, we assume that Oxx exactly recovers $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ in $j = \mathrm{Card}\,[\mathcal{Q}]$ iterations with some input vector in $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})$. This reachability assumption allows us to carry out a parallel analysis of OMP and OLS (subsection A-B1) leading to the following proposition.

Proposition 1 ***[Necessary condition for Oxx after*** j ***iterations]*** *Assume that* $\boldsymbol{A}_{\mathcal{Q}^\star}$ *is full rank and* $\mathcal{Q} \subsetneq \mathcal{Q}^\star$ *is reachable from an input in* $\mathrm{span}(\boldsymbol{A}_\mathcal{Q})$ *by Oxx. If ERC-Oxx(*$\boldsymbol{A}, \mathcal{Q}^\star, \mathcal{Q}$*) does not hold, then there exists* $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ *for which Oxx selects* $\mathcal{Q}$ *in the first* j *iterations and then a wrong atom* $\boldsymbol{a}_{\mathrm{bad}}$ *in the* $(j+1)$*-th iteration.*

This proposition coincides with Theorem 5 in the case of OMP whereas for OLS, Theorem 4 does not require the assumption that $\mathcal{Q}$ is reachable.

The second stage investigates whether the reachability assumption is automatically fulfilled or not (see subsections A-B2 and A-B3 for OLS and OMP, respectively).

1) Parallel analysis of OMP and OLS: *Proof of Proposition 1:* We proceed the proof of Lemma 4 backwards. By assumption, the right hand-side of inequality (13) is equal to

$$\|\big(\tilde{\boldsymbol{C}}^\dagger_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\boldsymbol{C}}_{\mathrm{bad}}\big)^t\|_{\infty,\infty} = \max_{\boldsymbol{a}_{\mathrm{bad}}} F^{\mathrm{Oxx}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}}) \geqslant 1.$$

By definition of induced norms, there exists a vector $\boldsymbol{v} \in \mathbb{R}^{k-j}$ satisfying $\boldsymbol{v} \neq \boldsymbol{0}$ and

$$\frac{\|\big(\tilde{\boldsymbol{C}}^\dagger_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\boldsymbol{C}}_{\mathrm{bad}}\big)^t\boldsymbol{v}\|_\infty}{\|\boldsymbol{v}\|_\infty} = \|\big(\tilde{\boldsymbol{C}}^\dagger_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\boldsymbol{C}}_{\mathrm{bad}}\big)^t\|_{\infty,\infty} \geqslant 1. \tag{20}$$

Define

$$\hat{\boldsymbol{y}} = \boldsymbol{A}_{\mathcal{Q}^\star\setminus\mathcal{Q}}(\tilde{\boldsymbol{C}}^t_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\setminus\mathcal{Q}})^{-1}\boldsymbol{v}. \tag{21}$$

The matrix inversion in (21) is well defined since $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\setminus\mathcal{Q}}$ is full rank (Corollary 3 in Appendix B) and $\tilde{\boldsymbol{C}}_{\mathcal{Q}^\star\setminus\mathcal{Q}} = \tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\setminus\mathcal{Q}}$ or $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star\setminus\mathcal{Q}}$ reads as the right product of $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\setminus\mathcal{Q}}$ with a nondegenerate diagonal matrix. By taking into account that $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\setminus\mathcal{Q}} = \boldsymbol{P}^\perp_\mathcal{Q}\boldsymbol{A}_{\mathcal{Q}^\star\setminus\mathcal{Q}}$, we obtain that

$$\boldsymbol{v} = \tilde{\boldsymbol{C}}^t_{\mathcal{Q}^\star\setminus\mathcal{Q}}\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}}. \tag{22}$$

Since the left hand-side of (20) identifies with $\phi(\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}})$ where ϕ is defined in (12), (20) yields:

$$\max_{i\notin\mathcal{Q}^\star}|\langle\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}},\tilde{\boldsymbol{c}}_i\rangle| \geqslant \max_{i\in\mathcal{Q}^\star\setminus\mathcal{Q}}|\langle\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}},\tilde{\boldsymbol{c}}_i\rangle|. \tag{23}$$

Moreover, we have $\boldsymbol{P}^\perp_\mathcal{Q}\hat{\boldsymbol{y}} \neq \boldsymbol{0}$ according to (22) and $\boldsymbol{v} \neq \boldsymbol{0}$.

Now, let $\boldsymbol{z} \in \mathrm{span}(\boldsymbol{A}_\mathcal{Q})$ denote the input for which Oxx recovers $\mathcal{Q}$. According to Lemma 15 in Appendix D, the first j iterations of Oxx with the modified input $\boldsymbol{y} = \boldsymbol{z} + \varepsilon\hat{\boldsymbol{y}}$ also select $\mathcal{Q}$ when $\varepsilon > 0$

is sufficiently small. Because $\boldsymbol{P}_{\mathcal{Q}}^{\perp}\boldsymbol{y} = \varepsilon \boldsymbol{P}_{\mathcal{Q}}^{\perp}\hat{\boldsymbol{y}}$ and (23) holds, the $(j+1)$-th iteration of Oxx necessarily selects a wrong atom. ■

At this point, we have proved Theorem 5 which is relative to OMP.

2) OLS ability to reach any subset: In order to prove Theorem 4, we establish that any subset $\mathcal{Q}$ can be reached using OLS with some input $\boldsymbol{y} \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}})$ (Lemma 3). To generate $\boldsymbol{y}$, we assign decreasing weight coefficients to the atoms $\{\boldsymbol{a}_i,\, i \in \mathcal{Q}\}$ with a rate of decrease which is high enough.

Proof of Lemma 3: Without loss of generality, we assume that the elements of $\mathcal{Q}$ correspond to the first j atoms.

Firstly, we define the vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j\}$ resulting from the orthogonalization of $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_j\}$: for all $i \leqslant j$, we have $\operatorname{span}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_i) = \operatorname{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_i)$ where $\boldsymbol{v}_1 \triangleq \boldsymbol{a}_1$ and for $i > 1$, $\boldsymbol{v}_i$ is set to the orthogonal projection of $\boldsymbol{a}_i$ onto $\operatorname{span}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{i-1})^{\perp}$.

Secondly, for arbitrary values of $\varepsilon_2, \ldots, \varepsilon_j > 0$, we define the following recursive construction:

- $\boldsymbol{y}_1 = \boldsymbol{v}_1$,
- $\boldsymbol{y}_i = \boldsymbol{y}_{i-1} + \varepsilon_i \boldsymbol{v}_i$ for $i \in \{2, \ldots, j\}$.

($\boldsymbol{y}_i$ implicitly depends on $\varepsilon_2, \ldots, \varepsilon_i$) and set $\boldsymbol{y} \triangleq \boldsymbol{y}_j$. We show by recursion that there exist $\varepsilon_2, \ldots, \varepsilon_i > 0$ such that OLS with $\boldsymbol{y}_i$ as input successively selects $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_i$ during the first i iterations (in particular, the selection rule (2) always yields a unique maximum).

The statement is obviously true for $\boldsymbol{y}_1 = \boldsymbol{a}_1$. Assume that it is true for $\boldsymbol{y}_{i-1}$ with some $\varepsilon_2, \ldots, \varepsilon_{i-1} > 0$ (these parameters will remain fixed in the following). According to Lemma 15 in Appendix D, there exists $\varepsilon_i > 0$ such that OLS with $\boldsymbol{y}_i = \boldsymbol{y}_{i-1} + \varepsilon_i \boldsymbol{v}_i$ as input selects the same atoms as with $\boldsymbol{y}_{i-1}$ during the first $i-1$ iterations, *i.e.*, $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{i-1}$ are successively chosen. At iteration i, the current active set thus reads $\mathcal{Q}' = \{1, \ldots, i-1\}$ and the OLS residual corresponding to $\boldsymbol{y}_i$ takes the form

$$\boldsymbol{r}_{\mathcal{Q}'} = \boldsymbol{P}_{\mathcal{Q}'}^{\perp}\boldsymbol{y}_{i-1} + \varepsilon_i \boldsymbol{P}_{\mathcal{Q}'}^{\perp}\boldsymbol{v}_i \;=\; \varepsilon_i \boldsymbol{v}_i$$

since $\boldsymbol{y}_{i-1} \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})$ and $\boldsymbol{v}_i \in \operatorname{span}(\boldsymbol{A}_{\mathcal{Q}'})^{\perp}$. By construction, $\boldsymbol{v}_i$ is equal to $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'} = \boldsymbol{P}_{\mathcal{Q}'}^{\perp}\boldsymbol{a}_i$, thus $\boldsymbol{r}_{\mathcal{Q}'}$ is proportional to $\tilde{\boldsymbol{a}}_i^{\mathcal{Q}'}$ and then to $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'}$. Finally, the OLS criterion (2) is maximum for the atom $\boldsymbol{a}_i$ and the maximum value is equal to $|\langle \boldsymbol{r}_{\mathcal{Q}'}, \tilde{\boldsymbol{b}}_i^{\mathcal{Q}'} \rangle| = \|\boldsymbol{r}_{\mathcal{Q}'}\|$ since $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'}$ is of unit norm.

Finally, we show that no other atom yields this maximum value. Apply Lemma 8 in Appendix B: the full rankness of $\boldsymbol{A}_{\mathcal{Q}' \cup \{i,l\}}$ (as a family of less than $\operatorname{spark}(\boldsymbol{A})$ atoms) implies that $\left[\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'}, \tilde{\boldsymbol{b}}_l^{\mathcal{Q}'}\right]$ is full rank, thus $\tilde{\boldsymbol{b}}_i^{\mathcal{Q}'}$ and $\tilde{\boldsymbol{b}}_l^{\mathcal{Q}'}$ cannot be colinear. ■

Using Lemma 3, Proposition 1 simplifies to Theorem 4 in which the assumption that $\mathcal{Q}$ is reachable by OLS is omitted.

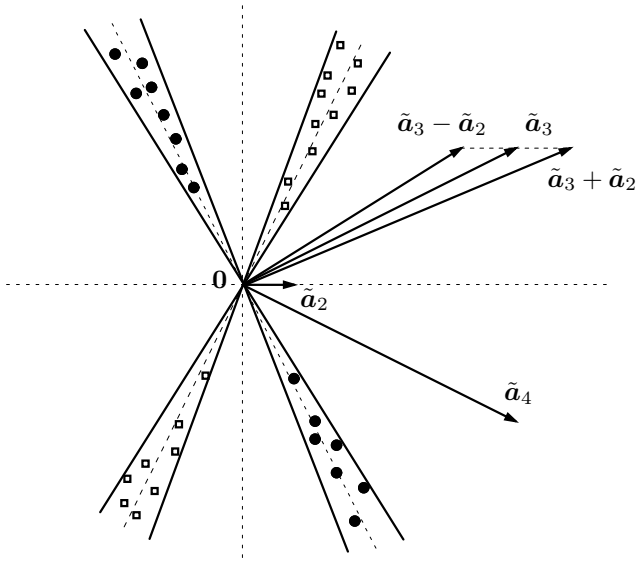


Fig. 3. Example 1: drawing of the plane $\mathrm{span}(\boldsymbol{a}_1)^\perp$. The tilde notation refers to the subset $\mathcal{Q} = \{1\}$. When θ_1 is close to 0, $\tilde{\boldsymbol{a}}_2$ is of very small norm since $\boldsymbol{a}_2$ is almost equal to $\boldsymbol{a}_1$, while $\boldsymbol{a}_3$ and $\boldsymbol{a}_4$, which are almost orthogonal to $\boldsymbol{a}_1$, yield projections $\tilde{\boldsymbol{a}}_3$ and $\tilde{\boldsymbol{a}}_4$ that are almost of unit norm. The angles $(\tilde{\boldsymbol{a}}_2, \tilde{\boldsymbol{a}}_3)$ and $(\tilde{\boldsymbol{a}}_2, \tilde{\boldsymbol{a}}_4)$ tend to θ_2 and $-\theta_2$ when $\theta_1 \to 0$. The bullet and square points correspond to positions $\boldsymbol{r}$ satisfying $|\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2\rangle| \geqslant |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_3\rangle|$ and $|\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2\rangle| \geqslant |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_4\rangle|$, respectively. These two cones only intersect at $\boldsymbol{r} = \mathbf{0}$, therefore OMP cannot successively select $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in the first two iterations.

3) OMP inability to reach some subsets: Contrary to OLS, OMP may not reach some subsets as stated in Example 1 in Section III. We now prove this result.

Proof of Example 1: Assume that OMP selects a true atom in the first iteration. Because there is a symmetry between $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, we can assume without loss of generality that $\boldsymbol{a}_1$ is selected. We show that $\boldsymbol{a}_3$ or $\boldsymbol{a}_4$ is necessarily selected in the second iteration.

As the atom dimension is $m = 3$, the residual $\boldsymbol{r}_{\{1\}}$ lies in $\mathrm{span}(\boldsymbol{a}_1)^\perp$ which is of dimension 2. The simple projection calculation $\tilde{\boldsymbol{a}}_i = \boldsymbol{a}_i - \langle \boldsymbol{a}_i, \boldsymbol{a}_1\rangle \boldsymbol{a}_1$ (the tilde notation implicitly refers to $\mathcal{Q} = \{1\}$) leads to:

$$\tilde{\boldsymbol{a}}_2 = \sin(2\theta_1) \begin{bmatrix} \sin\theta_1 \\ \cos\theta_1 \\ 0 \end{bmatrix}, \; \tilde{\boldsymbol{a}}_3 = \begin{bmatrix} \sin\theta_1 \cos\theta_1 \cos\theta_2 \\ \cos^2\theta_1 \cos\theta_2 \\ \sin\theta_2 \end{bmatrix} \text{ and } \; \tilde{\boldsymbol{a}}_4 = \begin{bmatrix} \sin\theta_1 \cos\theta_1 \cos\theta_2 \\ \cos^2\theta_1 \cos\theta_2 \\ -\sin\theta_2 \end{bmatrix}.$$

It is noticeable that when θ_1 is close to 0, $\|\tilde{\boldsymbol{a}}_2\| = |\sin(2\theta_1)|$ is small while $\tilde{\boldsymbol{a}}_3$ and $\tilde{\boldsymbol{a}}_4$ are almost of unit norm, and the angles $(\tilde{\boldsymbol{a}}_2, \tilde{\boldsymbol{a}}_3)$ and $(\tilde{\boldsymbol{a}}_2, \tilde{\boldsymbol{a}}_4)$ tend to θ_2 and $-\theta_2$ when $\theta_1 \to 0$ (see Fig. 3 for a 2D display in the plane $\mathrm{span}(\boldsymbol{a}_1)^\perp$).

It is easy to check that the set of points $\boldsymbol{r} \in \mathbb{R}^2$ satisfying $|\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2\rangle| \geqslant |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_3\rangle|$ is a 2D cone centered

around the direction that is orthogonal to $\tilde{\boldsymbol{a}}_3$ (dashed line in the south-east and north-west directions in Fig. 3). Specifically, both plain lines delimiting this cone are orthogonal to $\tilde{\boldsymbol{a}}_3 + \tilde{\boldsymbol{a}}_2$ and $\tilde{\boldsymbol{a}}_3 - \tilde{\boldsymbol{a}}_2$. Similarly, the set of points $\boldsymbol{r} \in \mathbb{R}^2$ satisfying $|\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2 \rangle| \geqslant |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_4 \rangle|$ is another 2D cone centered around the direction that is orthogonal to $\tilde{\boldsymbol{a}}_4$. When θ_1 is close to 0, both 2D cones only intersect at $\boldsymbol{r} = \boldsymbol{0}$ (since their inner angle tends towards 0), thus

$$\forall \boldsymbol{r} \in \mathbb{R}^2 \backslash \{\boldsymbol{0}\},\ |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_2 \rangle| < \max(|\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_3 \rangle|, |\langle \boldsymbol{r}, \tilde{\boldsymbol{a}}_4 \rangle|).$$

We conclude that $\boldsymbol{a}_2$ cannot be selected in the second iteration according to the OMP rule (1). ■

APPENDIX B
RE-EXPRESSION OF THE ERC-OXX FORMULAS

In this appendix, we prove Lemma 1 by successively re-expressing $\tilde{\boldsymbol{A}}^{\dagger}_{\mathcal{Q}^\star \backslash \mathcal{Q}} \tilde{\boldsymbol{a}}_{\mathrm{bad}}$ and $\tilde{\boldsymbol{B}}^{\dagger}_{\mathcal{Q}^\star \backslash \mathcal{Q}} \tilde{\boldsymbol{b}}_{\mathrm{bad}}$. Let us first show that when $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank, the matrices $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star \backslash \mathcal{Q}}$ and $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \backslash \mathcal{Q}}$ are full rank. This result is stated below as a corollary of Lemma 8.

Lemma 8 *If $\mathcal{Q} \cap \mathcal{Q}' = \emptyset$ and $\boldsymbol{A}_{\mathcal{Q} \cup \mathcal{Q}'}$ is full rank, then $\tilde{\boldsymbol{A}}^{\mathcal{Q}}_{\mathcal{Q}'}$ and $\tilde{\boldsymbol{B}}^{\mathcal{Q}}_{\mathcal{Q}'}$ are full rank.*

Proof: To prove that $\tilde{\boldsymbol{A}}^{\mathcal{Q}}_{\mathcal{Q}'}$ is full rank, we assume that $\sum_{j \in \mathcal{Q}'} \alpha_j \tilde{\boldsymbol{a}}^{\mathcal{Q}}_j = \boldsymbol{0}$ with $\alpha_j \in \mathbb{R}$. By definition of $\tilde{\boldsymbol{a}}^{\mathcal{Q}}_j = \boldsymbol{P}^{\perp}_{\mathcal{Q}} \boldsymbol{a}_j = \boldsymbol{a}_j - \boldsymbol{P}_{\mathcal{Q}} \boldsymbol{a}_j$, it follows that $\sum_{j \in \mathcal{Q}'} \alpha_j \boldsymbol{a}_j \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})$. Since $\boldsymbol{A}_{\mathcal{Q} \cup \mathcal{Q}'}$ is full rank, we conclude that all α_j's are 0.

The full rankness of $\tilde{\boldsymbol{B}}^{\mathcal{Q}}_{\mathcal{Q}'}$ directly follows from that of $\tilde{\boldsymbol{A}}^{\mathcal{Q}}_{\mathcal{Q}'}$ since for all $i \in \mathcal{Q}'$, $\tilde{\boldsymbol{b}}^{\mathcal{Q}}_i = \tilde{\boldsymbol{a}}^{\mathcal{Q}}_i / \|\tilde{\boldsymbol{a}}^{\mathcal{Q}}_i\|$ is colinear to $\tilde{\boldsymbol{a}}^{\mathcal{Q}}_i$. ■

The direct application of Lemma 8 to our context with $\mathcal{Q}' = \mathcal{Q}^\star \backslash \mathcal{Q}$ leads to the following corollary.

Corollary 3 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. For $\mathcal{Q} \subsetneq \mathcal{Q}^\star$, $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star \backslash \mathcal{Q}}$ and $\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star \backslash \mathcal{Q}}$ are full rank.*

Lemma 9 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. For $\mathcal{Q} \subsetneq \mathcal{Q}^\star$, $\tilde{\boldsymbol{A}}^{\dagger}_{\mathcal{Q}^\star \backslash \mathcal{Q}} \tilde{\boldsymbol{a}}_{\mathrm{bad}} = \left(\boldsymbol{A}^{\dagger}_{\mathcal{Q}^\star} \boldsymbol{a}_{\mathrm{bad}}\right)_{|(\mathcal{Q}^\star \backslash \mathcal{Q})}$ where $|$ denotes the restriction of a vector to a subset of its coefficients.*

Proof: The orthogonal decomposition of $\boldsymbol{a}_{\mathrm{bad}}$ on $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$ takes the form:

$$\boldsymbol{a}_{\mathrm{bad}} = \boldsymbol{A}_{\mathcal{Q}^\star} \left(\boldsymbol{A}^{\dagger}_{\mathcal{Q}^\star} \boldsymbol{a}_{\mathrm{bad}}\right) + \boldsymbol{P}^{\perp}_{\mathcal{Q}^\star} \boldsymbol{a}_{\mathrm{bad}}.$$

Projecting onto $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})^{\perp}$, we obtain

$$\tilde{\boldsymbol{a}}_{\mathrm{bad}} = \tilde{\boldsymbol{A}}_{\mathcal{Q}^\star \backslash \mathcal{Q}} \left(\boldsymbol{A}^{\dagger}_{\mathcal{Q}^\star} \boldsymbol{a}_{\mathrm{bad}}\right)_{|(\mathcal{Q}^\star \backslash \mathcal{Q})} + \boldsymbol{P}^{\perp}_{\mathcal{Q}^\star} \boldsymbol{a}_{\mathrm{bad}} \tag{24}$$

($\boldsymbol{P}_{\mathcal{Q}}^{\perp}\boldsymbol{P}_{\mathcal{Q}^\star}^{\perp} = \boldsymbol{P}_{\mathcal{Q}^\star}^{\perp}$ because $\mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})^{\perp} \subseteq \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}})^{\perp}$). For $i \in \mathcal{Q}^\star\backslash\mathcal{Q}$, $\tilde{\boldsymbol{a}}_i = \boldsymbol{a}_i - \boldsymbol{P}_{\mathcal{Q}}\boldsymbol{a}_i \in \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$. Thus, we have $\mathrm{span}(\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}) \subseteq \mathrm{span}(\boldsymbol{A}_{\mathcal{Q}^\star})$, and $\boldsymbol{P}_{\mathcal{Q}^\star}^{\perp}\boldsymbol{a}_{\mathrm{bad}}$ is orthogonal to $\mathrm{span}(\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}})$. According to Corollary 3, $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}$ is full rank. It follows from (24) that $\tilde{\boldsymbol{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}^{\dagger}\tilde{\boldsymbol{a}}_{\mathrm{bad}} = \left(\boldsymbol{A}_{\mathcal{Q}^\star}^{\dagger}\boldsymbol{a}_{\mathrm{bad}}\right)_{|(\mathcal{Q}^\star\backslash\mathcal{Q})}$. ■

Lemma 10 *Assume that $\boldsymbol{A}_{\mathcal{Q}^\star}$ is full rank. For $\mathcal{Q} \subsetneq \mathcal{Q}^\star$,*

$$\|\tilde{\boldsymbol{a}}_{\mathrm{bad}}\|\,\tilde{\boldsymbol{B}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}^{\dagger}\tilde{\boldsymbol{b}}_{\mathrm{bad}} = \boldsymbol{\Delta}_{\|\tilde{\boldsymbol{a}}_i\|}\left(\boldsymbol{A}_{\mathcal{Q}^\star}^{\dagger}\boldsymbol{a}_{\mathrm{bad}}\right)_{|(\mathcal{Q}^\star\backslash\mathcal{Q})}$$

where $\boldsymbol{\Delta}_{\|\tilde{\boldsymbol{a}}_i\|}$ stands for the diagonal matrix whose diagonal elements are $\{\|\tilde{\boldsymbol{a}}_i\|,\, i \in \mathcal{Q}^\star\backslash\mathcal{Q}\}$.

Proof: The result directly follows from $\tilde{\boldsymbol{a}}_{\mathrm{bad}} = \|\tilde{\boldsymbol{a}}_{\mathrm{bad}}\|\,\tilde{\boldsymbol{b}}_{\mathrm{bad}}$, $\tilde{\boldsymbol{b}}_i = \tilde{\boldsymbol{a}}_i/\|\tilde{\boldsymbol{a}}_i\|$ for $i \in \mathcal{Q}^\star\backslash\mathcal{Q}$, and from Lemma 9. ■

Proof of Lemma 1: The result is obvious when $\tilde{\boldsymbol{a}}_{\mathrm{bad}} = \boldsymbol{0}$. It follows from Lemmas 9 and 10 when $\tilde{\boldsymbol{a}}_{\mathrm{bad}} \neq \boldsymbol{0}$. ■

APPENDIX C
TECHNICAL RESULTS NEEDED FOR THE PROOF OF LEMMA 2

With simplified notations, the expression (19) of $F_{\mathcal{Q}^\star,\mathcal{Q}}^{\mathrm{OLS}}(\boldsymbol{a}_{\mathrm{bad}})$ reads

$$\varphi(\eta) \triangleq |\sqrt{1-\eta^2} - C\eta| + D\eta \tag{25}$$

where $\eta \in (0,1]$ and C and D take the form

$$C = \sum_{i=1}^{N} \frac{\beta_i \chi_i}{\eta_i} \tag{26}$$

$$D = \sum_{i=1}^{N} \frac{|\beta_i|}{\eta_i} \tag{27}$$

with $N \geqslant 1$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_N] \in \mathbb{R}^N$, and for all i, $\eta_i \in (0,1]$ and $\chi_i \in [-1,1]$ satisfy $\eta_i^2 + \chi_i^2 = 1$. Note that we can freely assume from (19) that $\chi_{\mathrm{bad}}^{\mathcal{Q},\mathcal{Q}'} = \pm\sqrt{1-\left(\eta_{\mathrm{bad}}^{\mathcal{Q},\mathcal{Q}'}\right)^2} \geqslant 0$. When $\chi_{\mathrm{bad}}^{\mathcal{Q},\mathcal{Q}'} < 0$, one just needs to replace $\boldsymbol{a}_{\mathrm{bad}}$ by $-\boldsymbol{a}_{\mathrm{bad}}$, leading to the replacement of $\boldsymbol{\beta}$ by $-\boldsymbol{\beta}$ in (26) and (27).

The succession of small lemmas hereafter aims at minorizing $\varphi(\eta)$ for arbitrary values of η, η_i, χ_i and $\boldsymbol{\beta}$. They lead to the main minoration result of Lemma 14.

Lemma 11 *Let $\boldsymbol{\beta} \in \mathbb{R}^N$.*

$$\textit{If } C \leqslant 0, \forall \eta \in [0,1],\ \varphi(\eta) \geqslant 1 + (\|\boldsymbol{\beta}\|_1 - 1)\eta. \tag{28}$$

$$\textit{If } C > 0,\ \min_{\eta\in[0,1]} \varphi(\eta) = \min\left(1, D/\sqrt{1+C^2}\right). \tag{29}$$

Proof: We first study the function $f(\eta) \triangleq \sqrt{1-\eta^2} - C\eta$. We have $f(0) = 1$, $f(1) = -C$, and f is concave on $[0,1]$. To minorize $\varphi(\eta) = |f(\eta)| + D\eta$, we distinguish two cases depending on the sign of C.

When $C \leqslant 0$, $f(\eta) \geqslant 0$ for all η. Since $|f| = f$ is concave, it can be minorized by the secant line joining $f(0)$ and $f(1)$, therefore, $|f(\eta)| \geqslant 1 - (C+1)\eta \geqslant 1 - \eta$. (28) follows from $\varphi(\eta) = |f(\eta)| + D\eta$ and $D \geqslant \|\boldsymbol{\beta}\|_1$ (because η_i are all in $(0,1]$).

When $C > 0$, $f(\eta) \geqslant 0$ for $\eta \in [0, z]$ and < 0 in $(z, 1]$, with $z \triangleq 1/\sqrt{1+C^2}$. $D \geqslant 0$ and $f(z) = 0$ imply that for $\eta > z$, $\varphi(\eta) \geqslant \varphi(z)$, thus the minimum of φ is reached for $\eta \in [0, z]$. On $[0, z]$, $\varphi(\eta) = f(\eta) + D\eta$ is concave, therefore the minimum value is either $\varphi(0) = 1$ or $\varphi(z) = Dz$. ■

The following two lemmas are simple inequalities linking C, D, and $\|\boldsymbol{\beta}\|_1$.

Lemma 12 $\forall \boldsymbol{\beta} \in \mathbb{R}^N$, $D^2 - C^2 \geqslant \|\boldsymbol{\beta}\|_1^2$.

Proof: By developing C^2 and D^2 from (26) and (27), we get

$$C^2 = \sum_i \frac{\beta_i^2 \chi_i^2}{\eta_i^2} + \sum_{i \neq j} \frac{\beta_i \beta_j \chi_i \chi_j}{\eta_i \eta_j}$$

$$D^2 = \sum_i \frac{\beta_i^2}{\eta_i^2} + \sum_{i \neq j} \frac{|\beta_i \beta_j|}{\eta_i \eta_j}$$

Since $\forall i$, $\eta_i^2 + \chi_i^2 = 1$, we have:

$$\begin{aligned} D^2 - C^2 &= \sum_i \beta_i^2 + \sum_{i \neq j} \frac{|\beta_i \beta_j|}{\eta_i \eta_j} \left(1 - \sigma_i \sigma_j \chi_i \chi_j\right) \\ &= \left[\sum_i |\beta_i|\right]^2 + \sum_{i \neq j} |\beta_i \beta_j| \left[\frac{1 - \sigma_i \sigma_j \chi_i \chi_j}{\eta_i \eta_j} - 1\right] \end{aligned} \tag{30}$$

with $\sigma_i = \operatorname{sgn}(\beta_i) = \pm 1$ if $\beta_i \neq 0$, and $\sigma_i = 1$ otherwise. Because η_i and χ_i satisfy $\eta_i^2 + \chi_i^2 = 1$, they reread $\eta_i = \cos\theta_i$ and $\chi_i = \sin\theta_i$, so $\eta_i \eta_j + \sigma_i \sigma_j \chi_i \chi_j = \cos(\theta_i \pm \theta_j) \leqslant 1$ which proves that the last bracketed expression in (30) is non-negative. Finally, (30) yields $D^2 - C^2 \geqslant \|\boldsymbol{\beta}\|_1^2$. ■

Lemma 13 $\forall \boldsymbol{\beta} \in \mathbb{R}^N$, $\|\boldsymbol{\beta}\|_1 \leqslant 1$ *implies that* $\|\boldsymbol{\beta}\|_1 \leqslant D/\sqrt{1+C^2}$.

Proof: $(1+C^2)\|\boldsymbol{\beta}\|_1^2 \leqslant \|\boldsymbol{\beta}\|_1^2 + C^2 \leqslant D^2$ according to Lemma 12. ■

We can now establish the main lemma that will enable us to conclude that if $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}}(\boldsymbol{a}_{\mathrm{bad}}) < 1$, $F^{\mathrm{OLS}}_{\mathcal{Q}^\star,\mathcal{Q}'}(\boldsymbol{a}_{\mathrm{bad}})$ is monotonically nonincreasing when $\mathcal{Q}' \supsetneq \mathcal{Q}$ is growing.

Lemma 14 $\forall \boldsymbol{\beta} \in \mathbb{R}^N$, $\forall \eta \in [0,1]$, $\varphi(\eta) < 1$ *implies that* $\|\boldsymbol{\beta}\|_1 \leqslant \varphi(\eta)$.

Proof: Apply Lemma 11.

When $C \leqslant 0$, (28) and $\varphi(\eta) < 1$ imply that $(\|\boldsymbol{\beta}\|_1 - 1) < 0$. Since $\eta \leqslant 1$, the lower bound of (28) is larger than $1 + (\|\boldsymbol{\beta}\|_1 - 1) = \|\boldsymbol{\beta}\|_1$.

When $C > 0$, (29) and $\varphi(\eta) < 1$ imply that the minimum value of φ on $[0, 1]$ is $D/\sqrt{1 + C^2} < 1$, then $D^2 - C^2 < 1$. Lemmas 12 and 13 imply that $\|\boldsymbol{\beta}\|_1 \leqslant 1$ and then $\|\boldsymbol{\beta}\|_1 \leqslant D/\sqrt{1 + C^2} \leqslant \varphi(\eta)$. ■

APPENDIX D
BEHAVIOR OF OXX WHEN THE INPUT VECTOR IS SLIGHTLY MODIFIED

Lemma 15 *Let $\boldsymbol{y}_1$ and $\boldsymbol{y}_2 \in \mathbb{R}^m$. Assume that the selection rule* (1)-(2) *of Oxx with $\boldsymbol{y}_1$ as input is strict in the first $j > 0$ iterations (the maximizer is unique). Then, when $\varepsilon > 0$ is sufficiently small, Oxx selects the same atoms with $\boldsymbol{y}(\varepsilon) = \boldsymbol{y}_1 + \varepsilon \boldsymbol{y}_2$ as with $\boldsymbol{y}_1$ in the first j iterations.*

Proof: We show by recursion that there exists $\varepsilon_l > 0$ such that the first l iterations of Oxx ($l = 1, \ldots, j$) with $\boldsymbol{y}(\varepsilon)$ and $\boldsymbol{y}_1$ as inputs yield the same atoms whenever $\varepsilon < \varepsilon_l$.

Let $l \geqslant 1$. We denote by $\mathcal{Q}$ the subset of cardinality $l - 1$ delivered by Oxx with $\boldsymbol{y}_1$ as input after $l - 1$ iterations. By assumption, $\mathcal{Q}$ is also yielded with $\boldsymbol{y}(\varepsilon)$ when $\varepsilon < \varepsilon_{l-1}$. Since $\boldsymbol{y}(\varepsilon) = \boldsymbol{y}_1 + \varepsilon \boldsymbol{y}_2$, the Oxx residual takes the form $\boldsymbol{r}_{\mathcal{Q}} = \boldsymbol{r}_1 + \varepsilon \boldsymbol{r}_2$ where $\boldsymbol{r}_{\mathcal{Q}}$, $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are obtained by projecting $\boldsymbol{y}(\varepsilon)$, $\boldsymbol{y}_1$, and $\boldsymbol{y}_2$, respectively onto $\operatorname{span}(\boldsymbol{A}_{\mathcal{Q}})^{\perp}$. Hence, for $i \notin \mathcal{Q}$,

$$\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle = \langle \boldsymbol{r}_1, \tilde{\boldsymbol{c}}_i \rangle + \varepsilon \langle \boldsymbol{r}_2, \tilde{\boldsymbol{c}}_i \rangle. \tag{31}$$

Let $\boldsymbol{a}_{\text{new}}$ denote the new atom selected by Oxx in the l-th iteration with $\boldsymbol{y}_1$ as input and let i_{new} refer to the corresponding index in the dictionary. By assumption, the atom selection is strict, *i.e.*,

$$|\langle \boldsymbol{r}_1, \tilde{\boldsymbol{c}}_{\text{new}} \rangle| > \max_{i \neq i_{\text{new}}} |\langle \boldsymbol{r}_1, \tilde{\boldsymbol{c}}_i \rangle|. \tag{32}$$

Taking the limit of (31) when $\varepsilon \to 0$, we obtain that for any i, $|\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle|$ tends toward $|\langle \boldsymbol{r}_1, \tilde{\boldsymbol{c}}_i \rangle|$. (32) implies that when $\varepsilon < \varepsilon_{l-1}$ is sufficiently small,

$$|\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_{\text{new}} \rangle| > \max_{i \neq i_{\text{new}}} |\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle|$$

by continuity of $|\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_i \rangle|$ ($i \neq i_{\text{new}}$) and $|\langle \boldsymbol{r}_{\mathcal{Q}}, \tilde{\boldsymbol{c}}_{\text{new}} \rangle|$ with respect to ε. Thus, Oxx with $\boldsymbol{y}(\varepsilon)$ as input selects $\boldsymbol{a}_{\text{new}}$ in the l-th iteration. ■

REFERENCES

[1] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation", *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[2] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries", *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[3] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition", in *Proc. 27th Asilomar Conf. on Signals, Systems and Computers*, Nov. 1993, vol. 1, pp. 40–44.

[4] 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.

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

[6] T. Blumensath and M. E. Davies, "On the difference between Orthogonal Matching Pursuit and Orthogonal Least Squares", Tech. Rep., University of Edinburgh, Mar. 2007.

[7] A. J. Miller, *Subset Selection in Regression*, Chapman and Hall, London, UK, 2nd edition, Apr. 2002.

[8] 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.

[9] L. Rebollo-Neira and D. Lowe, "Optimized orthogonal matching pursuit approach", *IEEE Signal Process. Lett.*, vol. 9, no. 4, pp. 137–140, Apr. 2002.

[10] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems", *Proc. IEEE, invited paper (Special Issue "Applications of sparse representation and compressive sensing")*, vol. 98, no. 5, pp. 948–958, June 2010.

[11] M. E. Davies and Y. C. Eldar, "Rank awareness in joint sparse recovery", Tech. Rep., University of Edinburgh, Apr. 2010.

[12] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property", *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sept. 2010.

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

[14] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases", *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[15] 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.

[16] S. Bourguignon, C. Soussen, H. Carfantan, and J. Idier, "Sparse deconvolution: Comparison of statistical and deterministic approaches", in *IEEE Workshop Stat. Sig. Proc.*, Nice, France, June 2011.