Coherence-based Partial Exact Recovery Condition for OMP/OLS

Cedric Herzet, Charles Soussen, Jérôme Idier, Rémi Gribonval

Coherence-based Partial Exact Recovery Condition for OMP/OLS

C. Herzet*, C. Soussen, J. Idier, and R. Gribonval

Abstract

We address the exact recovery of the support of a k-sparse vector with Orthogonal Matching Pursuit (OMP) and Orthogonal Least Squares (OLS) in a noiseless setting. We consider the scenario where OMP/OLS have selected good atoms during the first l iterations ($l < k$) and derive a new sufficient and worst-case necessary condition for their success in k steps. Our result is based on the coherence μ of the dictionary and relaxes Tropp's well-known condition $\mu < 1/(2k-1)$ to the case where OMP/OLS have a *partial* knowledge of the support.

Index Terms

Orthogonal Matching Pursuit; Orthogonal Least Squares; coherence; k-step analysis; exact support recovery.

I. Introduction

Sparse representations aim at describing a signal as the combination of a few elementary signals (or atoms) taken from an overcomplete dictionary $\mathbf{A}$. In particular, in a noiseless setting, one wishes to find the vector with the smallest number of non-zero elements, satisfying a set of linear constraints, that is

$$\min \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y}, \tag{1}$$

C. Herzet and R. Gribonval are with INRIA Rennes - Bretagne Atlantique, Campus de Beaulieu, F-35042 Rennes Cedex, France (e-mail: Cedric.Herzet@inria.fr; Remi.Gribonval@inria.fr).

C. Soussen is with the Centre de Recherche en Automatique de Nancy (CRAN, UMR 7039, Université de Lorraine, CNRS). Campus Sciences, B.P. 70239, F-54506 Vandœuvre-lès-Nancy, France (e-mail: Charles.Soussen@cran.uhp-nancy.fr.)

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 (e-mail: Jerome.Idier@irccyn.ec-nantes.fr).

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$. Problem (1) is usually NP-hard [1], that is accessing to the solution requires to sweep over all possible supports for $\mathbf{x}$.

In order to circumvent this bottleneck, suboptimal (but tractable) algorithms have been proposed in the literature. Among the most popular approaches, one can mention the procedures based on a relaxation of the ℓ_0 pseudo-norm (*e.g.*, Basis Pursuit [2], FOCUSS [3]) and the so-called "greedy pursuit" algorithms, *e.g.*, Matching Pursuit (MP) [4], Orthogonal Matching Pursuit (OMP) [5], Orthogonal Least Squares (OLS) [1], [6]. However, the suboptimal nature of these algorithms raises the question of their performance. In particular, if $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$, under which conditions can one ensure that a suboptimal algorithm recovers $\mathbf{x}^\star$ from $\mathbf{y}$? The goal of this paper is to provide novel elements of answer to this question for OMP and OLS.

OMP has been widely studied in the recent years, including worst case [7], [8] and probabilistic analyses [9]. The existing exact recovery analyses of OMP were also adapted to several extensions of OMP, namely regularized OMP [8], weak OMP [10], and Stagewise OMP [11]. Although OLS has been known in the literature for a few decades (often under different names [12]), exact recovery analyses of OLS remain rare for two reasons. First, OLS is significantly more time consuming than OMP, therefore discouraging the choice of OLS for "real-time" applications, like in compressive sensing. Secondly, the selection rule of OLS is more complex, as the projected atoms are normalized. This makes the OLS analysis more tricky. When the dictionary atoms are close to orthogonal, OLS and OMP have a similar behavior, as emphasized in [10]. On the contrary, for correlated dictionary (*e.g.*, in inverse problems), their behavior significantly differ and OLS may be a better choice [13]. The above arguments motivate our analysis of both OMP and OLS although in the present paper, our low mutual coherence assumptions imply that the correlation between atoms is weak, therefore we do not exhibit difference of behavior between OMP and OLS.

In [7], Tropp provided the first general analysis of OMP. More specifically, he derived a sufficient and worst-case necessary condition under which OMP is ensured to recover a k-sparse vector with a given support, in k iterations. Recently, Soussen *et al.* [13] showed that Tropp's exact recovery condition (ERC) is also sufficient and worst-case necessary for OLS.

A possible drawback of Tropp's ERC stands in its cumbersome evaluation, since it requires to solve a number of linear systems. Hence, Tropp proposed in [7] a stronger sufficient condition, easier to evaluate, guaranteeing the recovery of *any* k-sparse vector (for any support) by OMP. His condition reads:

$$\mu < \frac{1}{2k-1}, \tag{2}$$

where μ is the dictionary coherence, which only involves inner products between the dictionary atoms (see Definition 2 below). Note that (2) is also a sufficient condition for OLS since (2) implies Tropp's ERC which, in turn, is a sufficient condition for OLS. On the other hand, Cai&Wang recently emphasized that (2) is a worst-case necessary condition in some sense [14].

At this point, let us stress that the conditions mentioned above are *worst-case* necessary, that is, OMP/OLS will fail *for some* $\mathbf{y}$'s (and some particular dictionaries for (2)) as soon as they are not satisfied. However, when these conditions are not verified, one can observe in practice that OMP/OLS often succeed in recovering $\mathbf{x}^\star$ for many *other* observation vectors. In this paper, we investigate the case where (2) is not necessarily satisfied, but OMP/OLS nevertheless select l atoms belonging to the support of $\mathbf{x}^\star$ during the first l iterations. Our work is in the continuity of [13], in which the authors extended Tropp's condition to the l-th iteration of OMP and OLS. The resulting conditions are however rather complex and unpractical for numerical evaluation. In this paper, we derive a simpler (although stronger) condition based on the coherence of the dictionary. We show that

$$\mu < \frac{1}{2k - l - 1}, \tag{3}$$

is sufficient and worst-case necessary (in some sense) for the success of OMP/OLS in k steps when l atoms of the support have been selected during the first l iterations.

II. Notations

The following notations will be used in this paper. $\langle\, . \,, \,.\, \rangle$ refers to the inner product between vectors, $\|\,.\,\|$ and $\|\,.\,\|_1$ stand for the Euclidean and the ℓ_1 norms, respectively. $.^\dagger$ denotes the pseudo-inverse of a matrix. For a full rank and undercomplete matrix, we have $\mathbf{X}^\dagger = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ where $.^T$ stands for the matrix transposition. When $\mathbf{X}$ is overcomplete, $\mathrm{spark}(\mathbf{X})$ denotes the minimum number of columns from $\mathbf{X}$ that are linearly dependent [15]. $\mathbf{1}_p$ (resp $\mathbf{0}_p$) denotes the all-one (resp. all-zero) vector of dimension p. The letter $\mathcal{Q}$ denotes some subset of the column indices, and $\mathbf{X}_\mathcal{Q}$ is the submatrix of $\mathbf{X}$ gathering the columns indexed by $\mathcal{Q}$. For vectors, $\mathbf{x}_\mathcal{Q}$ denotes the subvector of $\mathbf{x}$ indexed by $\mathcal{Q}$. We will denote the cardinality of $\mathcal{Q}$ as $|\mathcal{Q}|$. We use the same notation to denote the absolute value of a scalar quantity. Finally, $\mathbf{P}_\mathcal{Q} = \mathbf{X}_\mathcal{Q}\mathbf{X}_\mathcal{Q}^\dagger$ and $\mathbf{P}_\mathcal{Q}^\perp = \mathbf{I} - \mathbf{P}_\mathcal{Q}$ denote the orthogonal projection operators onto $\mathrm{span}(\mathbf{X}_\mathcal{Q})$ and $\mathrm{span}(\mathbf{X}_\mathcal{Q})^\perp$, where $\mathrm{span}(\mathbf{X})$ stands for the column span of $\mathbf{X}$, $\mathrm{span}(\mathbf{X})^\perp$ is the orthogonal complement of $\mathrm{span}(\mathbf{X})$ and $\mathbf{I}$ is the identity matrix whose dimension is equal to the number of rows in $\mathbf{X}$.

III. OMP AND OLS

In this section, we recall the selection rules defining OMP and OLS. Throughout the paper, we will assume that the dictionary columns are normalized.

First note that any vector $\mathbf{x}$ satisfying the constraint in (1) must have a support, say $\mathcal{Q}$, such that $\mathbf{r}_{\mathcal{Q}} \triangleq \mathbf{P}_{\mathcal{Q}}^{\perp}\mathbf{y} = \mathbf{0}_m$ since $\mathbf{y}$ must belong to $\mathrm{span}(\mathbf{A}_{\mathcal{Q}})$. Hence, problem (1) can equivalently be rephrased as

$$\min |\mathcal{Q}| \quad \text{subject to } \mathbf{r}_{\mathcal{Q}} = \mathbf{0}_m. \tag{4}$$

OMP and OLS can be understood as iterative procedures searching for a solution of (4) by sequentially updating a support estimate as

$$\mathcal{Q} = \mathcal{Q} \cup \{j\}, \tag{5}$$

where

$$j \in \begin{cases} \arg\max_i |\langle \mathbf{a}_i, \mathbf{r}_{\mathcal{Q}} \rangle| & \text{for OMP} \\ \arg\min_i \|\mathbf{r}_{\mathcal{Q}\cup\{i\}}\| & \text{for OLS} \end{cases} \tag{6}$$

and $\mathbf{a}_i$ is the ith column of $\mathbf{A}$. More specifically, OMP/OLS add one new atom to the support at each iteration: OLS selects the atom minimizing the norm of the new residual $\mathbf{r}_{\mathcal{Q}\cup\{i\}}$ whereas OMP picks the atom maximizing the correlation with the current residual.

In the sequel, we will use a slightly different, equivalent, formulation of (6). Let us define

$$\tilde{\mathbf{a}}_i \triangleq \mathbf{P}_{\mathcal{Q}}^{\perp}\mathbf{a}_i, \tag{7}$$

$$\tilde{\mathbf{b}}_i \triangleq \begin{cases} \frac{\tilde{\mathbf{a}}_i}{\|\tilde{\mathbf{a}}_i\|} & \text{if } \tilde{\mathbf{a}}_i \neq \mathbf{0}_m \\ \mathbf{0}_m & \text{otherwise.} \end{cases} \tag{8}$$

Hence, $\tilde{\mathbf{a}}_i$ denotes the projection of $\mathbf{a}_i$ onto $\mathrm{span}(\mathbf{A}_{\mathcal{Q}})^{\perp}$ whereas $\tilde{\mathbf{b}}_i$ is a normalized version of $\tilde{\mathbf{a}}_i$. For simplicity, we dropped the dependence of $\tilde{\mathbf{a}}_i$ and $\tilde{\mathbf{b}}_i$ on $\mathcal{Q}$ in our notations. However, when there is a risk of confusion, we will use $\tilde{\mathbf{a}}_i^{\mathcal{Q}}$ (resp. $\tilde{\mathbf{b}}_i^{\mathcal{Q}}$) instead of $\tilde{\mathbf{a}}_i$ (resp. $\tilde{\mathbf{b}}_i$). With these notations, (6) can be re-expressed as

$$j \in \begin{cases} \arg\max_i |\langle \tilde{\mathbf{a}}_i, \mathbf{r}_{\mathcal{Q}} \rangle| & \text{for OMP} \\ \arg\max_i |\langle \tilde{\mathbf{b}}_i, \mathbf{r}_{\mathcal{Q}} \rangle| & \text{for OLS.} \end{cases} \tag{9}$$

The equivalence between (6) and (9) is straightforward for OMP by noticing that $\mathbf{r}_{\mathcal{Q}} \in \mathrm{span}(\mathbf{A}_{\mathcal{Q}})^{\perp}$. We refer the reader to [16] for a detailed calculation for OLS.

Throughout the paper, we will use the common acronym Oxx in statements that apply to both OMP and OLS. Moreover, we define the unifying notation:

$$\tilde{\mathbf{c}}_i \triangleq \begin{cases} \tilde{\mathbf{a}}_i & \text{for OMP,} \\ \tilde{\mathbf{b}}_i & \text{for OLS.} \end{cases} \tag{10}$$

Finally, we will use the notations $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{C}}$ to refer to the matrices whose columns are made up of the $\tilde{\mathbf{a}}_i$'s, $\tilde{\mathbf{b}}_i$'s and $\tilde{\mathbf{c}}_i$'s, respectively.

IV. Context and Main Result

Let us assume that $\mathbf{y}$ is a linear combination of k columns of $\mathbf{A}$, that is

$$\mathbf{y} = \mathbf{A}_{\mathcal{Q}^\star}\mathbf{x}_{\mathcal{Q}^\star} \qquad \text{with } |\mathcal{Q}^\star| = k,\ x_i \neq 0\ \forall i \in \mathcal{Q}^\star. \tag{11}$$

The atoms $\mathbf{a}_i$ ($i \in \mathcal{Q}^\star$) will be referred to as the "true" atoms. We review hereafter different conditions ensuring the success of Oxx and present our main result. The definition of "success" that will be used throughout the paper is as follows.

Definition 1 (Successful recovery) *Oxx with $\mathbf{y}$ as input succeeds if and only if it selects atoms in $\mathcal{Q}^\star$ during the first k iterations.*

The notion of successful recovery may be defined in a weaker sense: Plumbley [17, Corollary 4] first pointed out that there exist problems for which "delayed recovery" occurs after more than k steps. Specifically, Oxx can select some wrong atoms during the first k iterations but ends up with a larger support including $\mathcal{Q}^\star$ with a number of iterations slightly greater than k. In the noise-free setting (for $\mathbf{y} \in \mathrm{span}(\mathbf{A}_{\mathcal{Q}^\star})$), all atoms not belonging to $\mathcal{Q}^\star$ are then weighted by 0 in the solution vector. Recently, a delayed recovery analysis of OMP using restricted-isometry constants was proposed in [18] and then extended to the weak OMP algorithm (including OLS) in [10]. In the present paper, exactly k steps are performed, thus delayed recovery is considered as a recovery failure.

Moreover, we make clear that in special cases where the Oxx selection rule yields multiple solutions including a wrong atom, that is

$$\max_{i \in \mathcal{Q}^\star} |\langle \tilde{\mathbf{c}}_i, \mathbf{r}_{\mathcal{Q}} \rangle| = \max_{i \notin \mathcal{Q}^\star} |\langle \tilde{\mathbf{c}}_i, \mathbf{r}_{\mathcal{Q}} \rangle|, \tag{12}$$

we consider that Oxx systematically takes a wrong decision. Hence, situation (12) always leads to a recovery failure.

The first thoughtful theoretical analysis of OMP is due to Tropp, see [7, Theorems 3.1 and 3.10]. Tropp provided a sufficient and worst-case necessary condition for the exact recovery of any sparse vector with *a given* support $\mathcal{Q}^\star$. The derivation of a similar condition for OLS is more recent and is due to Soussen *et al.* in [13]. In the latter paper, the authors carried out a narrow analysis of both OMP and OLS at any iteration of the algorithm using specific recovery conditions depending not only on $\mathcal{Q}^\star$ but also on the current support $\mathcal{Q}$, whereas Tropp's ERC only involves $\mathcal{Q}^\star$ and does not depend on the iteration. The main result in [13] reads:

Theorem 1 (Soussen *et al.* 's Partial ERC [13, Theorem 3]) *Assume that $\mathbf{A}_{\mathcal{Q}^\star}$ is full rank and let $\mathcal{Q} \subset \mathcal{Q}^\star$ with $|\mathcal{Q}^\star| = k$, $|\mathcal{Q}| = l$. If Oxx with $\mathbf{y} \in \mathrm{span}(\mathbf{A}_{\mathcal{Q}^\star})$ as input selects atoms in $\mathcal{Q}$ during the first l iterations, and*

$$\max_{i \notin \mathcal{Q}^\star} \|\tilde{\mathbf{C}}^\dagger_{\mathcal{Q}^\star \backslash \mathcal{Q}} \tilde{\mathbf{c}}_i\|_1 < 1, \tag{13}$$

then Oxx only selects atoms in $\mathcal{Q}^\star \backslash \mathcal{Q}$ during the $k - l$ subsequent iterations. Conversely, if (13) *does not hold, there exists $\mathbf{y} \in \mathrm{span}(\mathbf{A}_{\mathcal{Q}^\star})$ for which OLS selects $\mathcal{Q}$ during the first l iterations and then a wrong atom $j \notin \mathcal{Q}^\star$ at the $(l+1)$th iteration.*

We note that (13), on its own, does not constitute a worst-case necessary condition for OMP if $\mathcal{Q} \neq \emptyset$. More specifically, as shown in [13], some additional "reachability" hypotheses are required for (13) to be a worst-case necessary condition for OMP.

Interestingly, when $\mathcal{Q} = \emptyset$, one recovers Tropp's ERC [7]:

$$\max_{i \notin \mathcal{Q}^\star} \|\mathbf{A}^\dagger_{\mathcal{Q}^\star} \mathbf{a}_i\|_1 < 1, \tag{14}$$

which constitutes a sufficient and worst-case necessary condition for *both* OMP and OLS at the very first iteration.

One drawback of Tropp's and Soussen *et al.* 's ERCs stands in their unpractical evaluation. Indeed, evaluating (13)-(14) requires to carry out a pseudo-inverse (and a projection for (13)) operation. Moreover, support $\mathcal{Q}^\star$ is unknown in practice. Hence, ensuring that Oxx will recover any k-sparse vector requires to test whether (14) is met for all possible supports $\mathcal{Q}^\star$ of cardinality k (resp. to evaluate (13) for all $\mathcal{Q}^\star$ and for all $\mathcal{Q} \subset \mathcal{Q}^\star$ of cardinality l).

In order to circumvent this problem, stronger conditions, but easier to evaluate, have been proposed in the literature. We can mainly distinguish between two types of "practical" guarantees: the conditions based on restricted-isometry constants (RIC) and those based on the coherence of the dictionary (see Definition 2 below).

The contributions [8], [19]–[22] provide RIC-based sufficient conditions for an exact recovery of the support in k steps by OMP. The most recent and tightest results are due to Maleh [21] and Mo&Shen

[22]. The authors proved that OMP succeeds in k steps if $\delta_{k+1} < \frac{1}{\sqrt{K+1}}$, where δ_{k+1} is the $(k+1)$-RIC of $\mathbf{A}$. In [22, Theorem 3.2], the authors showed moreover that this condition is almost tight, *i.e.*, there exists a dictionary $\mathbf{A}$ with $\delta_{k+1} = \frac{1}{\sqrt{K}}$ and a k-term representation $\mathbf{y}$ for which OMP selects a wrong atom at the first iteration. Let us mention that, by virtue of Theorem 1, these results remain valid for OLS.

On the other hand, Tropp derived in [7, Corollary 3.6] a sufficient condition for OMP, stronger than (14) but only based on the coherence of the dictionary $\mathbf{A}$.

Definition 2 *The mutual coherence μ of a dictionary $\mathbf{A}$ is defined as*

$$\mu = \max_{i \neq j} |\langle \mathbf{a}_i, \mathbf{a}_j \rangle|. \tag{15}$$

Tropp's condition reads as in (2) and ensures that (14) is satisfied. Since (14) guarantees the success of OLS (Theorem 1 for iteration $l = 0$), (2) is also a sufficient condition for OLS. Moreover, Cai&Wang recently showed in [14, Theorem 3.1] that (2) is also worst-case necessary in the following sense: there exists (at least) one k-sparse vector $\mathbf{x}^\star$ and one dictionary $\mathbf{A}$ with $\mu = \frac{1}{2k-1}$ such that Oxx[1] cannot recover $\mathbf{x}^\star$ from $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$. These results are summarized in the following theorem:

Theorem 2 (μ-based ERC for Oxx [7, Corollary 3.6], [14, Theorem 3.1]) *If* (2) *is satisfied, then Oxx succeeds in recovering any k-term representation. Conversely, there exists an instance of dictionary $\mathbf{A}$ and a k-term representation for which:* (i) $\mu = \frac{1}{2k-1}$*;* (ii) *Oxx selects a wrong atom at the first iteration.*

In this paper, we extend the work by Soussen *et al.* and provide a coherence-based sufficient and worst-case necessary condition for the success of Oxx in k iterations provided that true atoms have been selected in the first l iterations. Our main result generalizes Theorem 2 to the case where l true atoms have been selected:

Theorem 3 (μ-based Partial ERC for Oxx) *Consider a k-term representation $\mathbf{y} \in \operatorname{span}(\mathbf{A}_{\mathcal{Q}^\star})$. Assume that, at iteration $l < k$, Oxx has selected l true atoms in $\mathcal{Q}^\star$. If*

$$\mu < \frac{1}{2k - l - 1}, \tag{16}$$

then Oxx exactly recovers $\mathcal{Q}^\star$ in k iterations.

Conversely, there exists a dictionary $\mathbf{A}$ and a k-term representation $\mathbf{y}$ such that: (i) $\mu = \frac{1}{2k-l-1}$*;* (ii) *Oxx selects true atoms during the first l iterations and then a wrong atom at the $(l+1)$th iteration.*

[1]and actually, any sparse representation algorithm.

The proof of this theorem is reported to sections V, VI and VII. More specifically, we show in section V (resp. section VI) that (16) is sufficient for the success of OMP (resp. OLS) during the last $k-l$ iterations. The proof of this sufficient condition significantly differs for OMP and OLS. The result is shown for OMP by deriving an upper bound on Soussen *et al.* 's extended ERC as a function of the restricted isometry bounds of the projected dictionary. As for OLS, the proof is based on a connection between Soussen *et al.* 's ERC and the mutual coherence of the normalized projected dictionary $\tilde{\mathbf{B}}$. Finally, in section VII we prove that (16) is worst-case necessary for Oxx in the sense specified in Theorem 3. The proof is common to both OMP and OLS.

V. SUFFICIENT CONDITION FOR OMP AT ITERATION l

In this section, we prove the sufficient condition result of Theorem 3 for OMP. The result is a direct consequence of Theorem 4 stated below, which provides an upper bound on the left-hand side of (13) only depending on the coherence of the dictionary $\mathbf{A}$:

Theorem 4 *Let $\mathcal{Q} \subset \mathcal{Q}^\star$, with $|\mathcal{Q}| = l$, $|\mathcal{Q}^\star| = k$. If*

$$\mu < \frac{1}{k-1} \tag{17}$$

then

$$\max_{i \notin \mathcal{Q}^\star} \|\tilde{\mathbf{A}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\mathbf{a}}_i\|_1 \leq \frac{(k-l)\mu}{1-(k-1)\mu}. \tag{18}$$

The sufficient condition for OMP stated in Theorem 3 then derives from Theorem 4. We see that

$$\frac{(k-l)\mu}{1-(k-1)\mu} < 1 \tag{19}$$

implies (13) and is therefore sufficient for the success of OMP in k iterations. Now, (19) is equivalent to (16) which proves the result.

Before proving Theorem 4, we need to define some quantities characterizing the *projected* dictionary $\tilde{\mathbf{A}}$ appearing in the implementation of OMP (see (9)) and state some useful propositions. In the following definition, we generalize the concept of restricted isometry property (RIP) [23] to projected dictionaries, under the name projected RIP (P-RIP):

Definition 3 *Dictionary $\mathbf{A}$ satisfies the P-RIP($\underline{\delta}_{q,l}$,$\bar{\delta}_{q,l}$) if and only if $\forall \mathcal{Q}', \mathcal{Q}$ with $|\mathcal{Q}'| = q$, $|\mathcal{Q}| = l$, $\mathcal{Q} \cap \mathcal{Q}' = \emptyset$, $\forall \mathbf{x}_{\mathcal{Q}'}$ we have*

$$(1-\underline{\delta}_{q,l})\|\mathbf{x}_{\mathcal{Q}'}\|^2 \leq \|\tilde{\mathbf{A}}^{\mathcal{Q}}_{\mathcal{Q}'} \mathbf{x}_{\mathcal{Q}'}\|^2 \leq (1+\bar{\delta}_{q,l})\|\mathbf{x}_{\mathcal{Q}'}\|^2. \tag{20}$$

The definition of the standard (asymmetric) restricted isometry constants corresponds to the tightest possible bounds when $l = 0$ (see *e.g.*, [24], [25]). For $l \geq 1$, $\underline{\delta}_{q,l}$ and $\bar{\delta}_{q,l}$ can be seen as (asymmetric) *bounds* on the restricted isometry constants of *projected* dictionaries. Note that $\bar{\delta}_{q,l}$ is not necessarily non-negative since the columns of $\tilde{\mathbf{A}}$ are not normalized ($\|\tilde{\mathbf{a}}_i^{\mathcal{Q}}\| \leq 1$). Note also that many well-known properties of the standard restricted isometry constants (see [26, Proposition 3.1] for example) remain valid for $\underline{\delta}_{q,l}$ and $\bar{\delta}_{q,l}$.

The next proposition provides an upper bound on the left-hand side of (13) only depending on $\underline{\delta}_{q,l}$ and $\bar{\delta}_{q,l}$:

Proposition 1 *Let $\mathcal{Q} \subset \mathcal{Q}^\star$, with $|\mathcal{Q}| = l$, $|\mathcal{Q}^\star| = k$. If $\underline{\delta}_{k-l,l} < 1$, then*

$$\max_{i \notin \mathcal{Q}^\star} \|\tilde{\mathbf{A}}^\dagger_{\mathcal{Q}^\star \setminus \mathcal{Q}} \tilde{\mathbf{a}}_i\|_1 \leq (k-l)\,\frac{\bar{\delta}_{2,l} + \underline{\delta}_{2,l}}{2(1 - \underline{\delta}_{k-l,l})}. \tag{21}$$

The proof of Proposition 1 is reported to Appendix V. The next proposition provides some possible values for $\underline{\delta}_{q,l}$ and $\bar{\delta}_{q,l}$ as a function of the coherence of the dictionary $\mathbf{A}$:

Proposition 2 *If $\mu < 1/(l-1)$, then $\mathbf{A}$ satisfies the P-RIP($\underline{\delta}_{q,l}$,$\bar{\delta}_{q,l}$) with*

$$\bar{\delta}_{q,l} = (q-1)\mu, \tag{22}$$

$$\underline{\delta}_{q,l} = (q-1)\mu + \frac{\mu^2 q l}{1-(l-1)\mu}. \tag{23}$$

The proof of this result is reported to Appendix V. We are now ready to prove Theorem 4:

Proof: (Theorem 4) We rewrite the right-hand side of (21) as a function of μ. From Proposition 2, we have that $\mathbf{A}$ satisfies the P-RIP($\underline{\delta}_{q,l}$,$\bar{\delta}_{q,l}$) with constants defined in (22)-(23) as long as

$$\mu < \frac{1}{l-1}. \tag{24}$$

Now, we have $\mu < 1/(k-1)$ by hypothesis, which implies $\mu < 1/(l-1)$. Using (22) and (23), we calculate that:

$$\frac{\bar{\delta}_{2,l} + \underline{\delta}_{2,l}}{2} = \mu + \frac{\mu^2 l}{1-(l-1)\mu} = \frac{\mu(\mu+1)}{1-(l-1)\mu}, \tag{25}$$

$$\begin{aligned} 1 - \underline{\delta}_{k-l,l} &= 1 - (k-l-1)\mu - \frac{\mu^2 (k-l) l}{1-(l-1)\mu} && (26)\\ &= \frac{1-(k-2)\mu - (k-1)\mu^2}{1-(l-1)\mu} && (27)\\ &= \frac{(\mu+1)(1-(k-1)\mu)}{1-(l-1)\mu}. && (28) \end{aligned}$$

Therefore, the ratio in the right-hand side of (21) can be rewritten as

$$\frac{\bar{\delta}_{2,l} + \underline{\delta}_{2,l}}{2(1 - \underline{\delta}_{k-l,l})} = \frac{\mu}{1 - (k-1)\mu}. \tag{29}$$

According to (28), $\mu < 1/(k-1) \leq 1/(l-1)$ implies that $1 - \underline{\delta}_{k-l,l} > 0$. Proposition 1 combined with (29) implies that (18) is met. ■

Before concluding this section, let us remark that unlike Theorem 1, Theorem 3 does not (explicitly) require all $(m \times k)$-submatrices $\mathbf{A}_{\mathcal{Q}^\star}$ to be full rank. However, this condition is implicitly enforced by (16). Indeed, as shown in [7, Lemma 2.3],

$$\mu < \frac{1}{k-1} \tag{30}$$

implies that $\mathbf{A}_{\mathcal{Q}^\star}$ is full rank when $|\mathcal{Q}^\star| = k$. Hence, since $k - 1 < 2k - l - 1$, (16) also implies that any submatrix $\mathbf{A}_{\mathcal{Q}^\star}$ with $|\mathcal{Q}^\star| = k$ is full rank. Finally, we remark that the full rankness of $\mathbf{A}_{\mathcal{Q}^\star}$ implies that the projected submatrices $\tilde{\mathbf{A}}_{\mathcal{Q}^\star \backslash \mathcal{Q}}$ involved in Theorem 4 are also full rank [13, Corollary 3].

VI. Sufficient condition for OLS at iteration l

We now prove the sufficient condition for OLS stated in Theorem 3. The result is a consequence of Proposition 3 and Lemma 1 stated below. We first need to introduce the coherence of the *normalized* projected dictionary $\tilde{\mathbf{B}}$:

Definition 4 (Coherence of the normalized projected dictionary)

$$\mu_l^{OLS} = \max_{|\mathcal{Q}|=l} \max_{i \neq j} |\langle \tilde{\mathbf{b}}_i^{\mathcal{Q}}, \tilde{\mathbf{b}}_j^{\mathcal{Q}} \rangle|. \tag{31}$$

The following proposition gives a sufficient condition on μ_l^{OLS} under which (13) is satisfied:

Proposition 3 *Let $\mathcal{Q} \subset \mathcal{Q}^\star$, with $|\mathcal{Q}| = l$, $|\mathcal{Q}^\star| = k$. Assume that $\mathbf{A}_{\mathcal{Q}^\star}$ is full rank. If $\mu_l^{OLS} < 1/(2k - 2l - 1)$, then*

$$\max_{i \notin \mathcal{Q}^\star} \|\tilde{\mathbf{B}}^\dagger_{\mathcal{Q}^\star \backslash \mathcal{Q}} \tilde{\mathbf{b}}_i\|_1 < 1. \tag{32}$$

Proof: When $\tilde{\mathbf{b}}_i = \mathbf{0}$, the result is obvious. When $\tilde{\mathbf{b}}_i \neq \mathbf{0}$, apply [7, Corollary 3.6] (that is: if $\mathbf{A}$ has normalized columns and $\mu < 1/(2k-1)$ then Tropp's ERC is satisfied, *i.e.*, $\forall \mathcal{Q}^\star$ such that $|\mathcal{Q}^\star| = k$, $\max_{i \notin \mathcal{Q}^\star} \|\mathbf{A}^\dagger_{\mathcal{Q}^\star} \mathbf{a}_i\|_1 < 1$) to the matrix $\tilde{\mathbf{B}}$ and to $\mathcal{Q}^\star \backslash \mathcal{Q}$ of size $k - l$. The atoms of $\tilde{\mathbf{B}}_{\mathcal{Q}^\star \backslash \mathcal{Q}}$ are of unit norm (actually, $\tilde{\mathbf{B}}_{\mathcal{Q}^\star \backslash \mathcal{Q}}$ is full rank) because $\mathbf{A}_{\mathcal{Q}^\star}$ is full rank [13, Corollary 3]. ■

The next lemma provides a useful upper bound on μ_l^{OLS} as a function of the coherence μ of the dictionary $\mathbf{A}$:

Lemma 1 *If $\mu < 1/l$, then*

$$\mu_l^{OLS} \leqslant \frac{\mu}{1 - l\mu}. \tag{33}$$

The proof of this result is reported to Appendix B. The sufficient condition stated in Theorem 3 for OLS then follows from the combination of Proposition 3 and Lemma 1. Indeed, (16) implies $\mu < 1/(k-1) \leq 1/l$ since $2k - l - 1 = k - 1 + (k - l) > k - 1 \geq l$. Hence, the result follows by first applying Lemma 1:

$$\mu_l^{OLS} \leqslant \frac{\mu}{1 - l\mu} < \frac{1}{2k - 2l - 1}, \tag{34}$$

and then Proposition 3, which implies that (32) is met. $\mu < 1/(k-1)$ implies that the full rank assumption of Proposition 3 is met for any $\mathcal{Q}^\star$ of cardinality k [7, Lemma 2.3].

VII. Worst-case necessary condition for Oxx at iteration l

Cai&Wang recently showed in [14, Theorem 3.1] that there exist dictionaries $\mathbf{A}$ with $\mu = \frac{1}{2k-1}$ and linear combinations $\mathbf{y}$ of k columns of $\mathbf{A}$ such that $\mathbf{y}$ has *two* distinct k-sparse representations in $\mathbf{A}$. In other words, if $\mu < \frac{1}{2k-1}$ is not satisfied, there exist instances of dictionaries such that *no* algorithm can univocally recover some k-sparse representations. In the context of Oxx, their result can be rephrased as the following worst-case necessary condition: there exists a dictionary $\mathbf{A}$ with $\mu = \frac{1}{2k-1}$ and a support $\mathcal{Q}^\star$, with $|\mathcal{Q}^\star| = k$, such that Oxx selects a wrong atom at the first iteration.

In this section, we derive a worst-case necessary condition in the case where Oxx has selected atoms in $\mathcal{Q}^\star$ during the first l iterations. We extend Cai&Wang's analysis and exhibit a scenario in which l true atoms are selected, then the Oxx residual after l iterations has two $(k-l)$-term representations. Our result reads

Theorem 5 ((16) is a worst-case necessary condition for Oxx) *There exists a dictionary $\mathbf{A}$ with $\mu = \frac{1}{2k-l-1}$, a support $\mathcal{Q}^\star$ with $|\mathcal{Q}^\star| = k$ and $\mathbf{y} \in \operatorname{span}(\mathbf{A}_{\mathcal{Q}^\star})$, such that Oxx with $\mathbf{y}$ as input selects l atoms in $\mathcal{Q}^\star$ during the first l iterations and a wrong atom at the $(l+1)$th iteration.*

To reach the result, we adopt a dictionary construction similar to Cai&Wang's in [14]. Let $\mathbf{M} \in \mathbb{R}^{(2k-l)\times(2k-l)}$ be the matrix with ones on the diagonal and $-\frac{1}{2k-l-1}$ elsewhere. $\mathbf{M}$ will play the role of the Gram matrix $\mathbf{M} = \mathbf{A}^T\mathbf{A}$. We will exploit the eigenvalue decomposition of $\mathbf{M}$ to construct the dictionary $\mathbf{A} \in \mathbb{R}^{(2k-l-1)\times(2k-l)}$ with the desired properties. Since $\mathbf{M}$ is symmetric, it can be expressed as

$$\mathbf{M} = \mathbf{U}\Lambda\mathbf{U}^T, \tag{35}$$

where $\mathbf{U}$ (resp. Λ) is the unitary matrix whose columns are the eigenvectors (resp. the diagonal matrix of eigenvalues) of $\mathbf{M}$. It is easy to check that $\mathbf{M}$ has only two distinct eigenvalues: $\frac{2k-l}{2k-l-1}$ with multiplicity $2k-l-1$ and 0 with multiplicity one; moreover, the eigenvector associated to the null eigenvalue is equal to $\mathbf{1}_{2k-l}$. The eigenvalues are sorted in the decreasing order so that 0 appears in the lower right corner of Λ.

We define $\mathbf{A} \in \mathbb{R}^{(2k-l-1)\times(2k-l)}$ as

$$\mathbf{A} = \Upsilon\mathbf{U}^T, \tag{36}$$

where $\Upsilon \in \mathbb{R}^{(2k-l-1)\times(2k-l)}$ is such that

$$\Upsilon(i,j) = \begin{cases} \sqrt{\frac{2k-l}{2k-l-1}} & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{37}$$

Note that $\Upsilon^T\Upsilon = \Lambda$. Hence, $\mathbf{A}$ satisfies the hypotheses of Theorem 5 since

$$\mathbf{A}^T\mathbf{A} = \mathbf{U}\Upsilon^T\Upsilon\mathbf{U}^T = \mathbf{U}\Lambda\mathbf{U}^T = \mathbf{M}, \tag{38}$$

and therefore

$$\langle \mathbf{a}_i, \mathbf{a}_j \rangle = -\frac{1}{2k-l-1} \quad \forall i \neq j. \tag{39}$$

Since $\mathbf{M} = \mathbf{A}^T\mathbf{A}$, we have $\mathbf{M}\mathbf{x} = \mathbf{0}_{2k-l}$ if and only if $\mathbf{A}\mathbf{x} = \mathbf{0}_{2k-l-1}$. Moreover, since $\mathbf{M}$ has *one single* zero eigenvalue with eigenvector $\mathbf{1}_{2k-l}$, the null-space of $\mathbf{A}$ is the one-dimensional space spanned by $\mathbf{1}_{2k-l}$. Therefore, any $p < 2k-l$ columns of $\mathbf{A}$ are linearly independent, *i.e.*, $\mathrm{spark}(\mathbf{A}) = 2k-l$.

Before proceeding to the proof of Theorem 5, we need to define the concept of "reachability" of a subset $\mathcal{Q}$:

Definition 5 *A subset $\mathcal{Q}$ is said to be reachable by Oxx if there exists $\mathbf{y} \in \mathrm{span}(\mathbf{A}_{\mathcal{Q}})$ such that Oxx with $\mathbf{y}$ as input selects atoms in $\mathcal{Q}$ during the first $|\mathcal{Q}|$ iterations.*

The concept of reachability was first introduced in [13]. The authors showed that any subset $\mathcal{Q}$ with $|\mathcal{Q}| \leq \mathrm{spark}(\mathbf{A}) - 2$ is reachable by OLS, see [13, Lemma 3]. On the other hand, they emphasized that there exist dictionaries for which some subsets $\mathcal{Q}$ can never be reached by OMP, see [13, Example 1]. This scenario does however not occur for the dictionary defined in (36) as stated in the next lemma:

Lemma 2 *Let* $\mathbf{A}$ *be defined as in* (36) *with* $l < k$*. Then any subset* $\mathcal{Q}$ *with* $|\mathcal{Q}| = l$ *is reachable by Oxx.*

The proof of this result is reported to Appendix C. To prove Theorem 5, we also need the following technical lemma whose proof is reported to Appendix C:

Lemma 3 *Let* $\mathbf{A}$ *be defined as in* (36) *with* $l < k$*. Then, for any subset* $\mathcal{Q}$ *with* $|\mathcal{Q}| = l$*, there exists a vector* $\mathbf{y}$ *having two* $(k-l)$*-term representations with disjoint supports in the projected dictionary* $\tilde{\mathbf{C}}_{\backslash\mathcal{Q}} \triangleq \tilde{\mathbf{C}}_{\{1,\ldots,2k-l\}\backslash\mathcal{Q}} \in \mathbb{R}^{2k-l-1\times 2k-2l}$.

We are now ready to prove Theorem 5:

Proof: (Theorem 5) Consider the dictionary $\mathbf{A}$ defined in (36) with $l < k$. Let $\mathcal{Q}$ be a subset of cardinality l, arbitrarily chosen (say, the first l atoms of the dictionary). We will exhibit a subset $\mathcal{Q}^\star \supset \mathcal{Q}$ for which the result of Theorem 5 holds.

We first apply Lemma 2: there exists an input $\mathbf{y}_1 \in \mathrm{span}(\mathbf{A}_\mathcal{Q})$ for which Oxx selects all atoms in $\mathcal{Q}$ during the first l iterations. Then, we apply Lemma 3: there exists a vector $\mathbf{y}_2$ having two $(k-l)$-term representations in the projected dictionary $\tilde{\mathbf{C}}_{\backslash\mathcal{Q}}$. We will denote their respective supports by $\mathcal{Q}_1$ and $\mathcal{Q}_2$ with $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \emptyset$.

By virtue of [13, Lemma 15], Oxx with $\mathbf{y} = \mathbf{y}_1 + \epsilon\mathbf{y}_2$ as input selects the same atoms (*i.e.*, $\mathcal{Q}$) as with $\mathbf{y}_1$ as input during the first l iterations as long as $\epsilon > 0$ is sufficiently small. Moreover, the selection rule (9) indicates that the atom $\tilde{\mathbf{a}}_j$ selected at iteration $l+1$ satisfies:

$$j \in \arg\max_i |\langle \tilde{\mathbf{c}}_i, \mathbf{P}_\mathcal{Q}^\perp \mathbf{y}\rangle| = \arg\max_i |\langle \tilde{\mathbf{c}}_i, \mathbf{y}_2\rangle|, \tag{40}$$

since $\mathbf{P}_\mathcal{Q}^\perp \mathbf{y} = \epsilon \mathbf{P}_\mathcal{Q}^\perp \mathbf{y}_2 = \epsilon\mathbf{y}_2$. Now, we set $\mathcal{Q}^\star$ in such a way that $j \notin \mathcal{Q}^\star$:

$$\mathcal{Q}^\star = \begin{cases} \mathcal{Q} \cup \mathcal{Q}_1 & \text{if } j \in \mathcal{Q}_2, \\ \mathcal{Q} \cup \mathcal{Q}_2 & \text{if } j \in \mathcal{Q}_1. \end{cases} \tag{41}$$

To complete the proof, it is easy to check that $\mathbf{y} = \mathbf{y}_1 + \epsilon\mathbf{y}_2 \in \mathrm{span}(\mathbf{A}_{\mathcal{Q}^\star})$ because $\mathbf{y}_1 \in \mathrm{span}(\mathbf{A}_\mathcal{Q})$ and $\mathbf{y}_2 \in \mathrm{span}(\tilde{\mathbf{C}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}) = \mathrm{span}(\tilde{\mathbf{A}}_{\mathcal{Q}^\star\backslash\mathcal{Q}}) \subset \mathbf{A}_{\mathcal{Q}^\star}$. ■

VIII. Conclusions

The sufficient and worst-case necessary condition we derived for the success of Oxx after the first l iterations have been completed reads $\mu < \frac{1}{2k-l-1}$ and relaxes the coherence-based results by Tropp [7] and Cai&Wang [14] corresponding to the case $l = 0$.

Our condition is obviously pessimistic since it is a worst-case condition for *all possible* supports of cardinality l. In comparison, the conditions we elaborated in [13] are sharper (although significantly more complex) and they are dedicated to a single support of size l. The latter conditions are indeed rather unpractical since they depend on the true support which is unknown. In practice, they shall be evaluated for all possible pairs of complete/partial supports of dimension k and l, and each evaluation requires a pseudo-inverse computation. A compromise between the pessimistic coherence condition and those elaborated in [13] would be to adapt our mutual coherence results to the cumulative coherence [7], and the weak ERC condition [7], [27], [28] (also referred to as the Neumann ERC in [29]). The latter conditions are intermediate conditions at iteration 0 between the mutual coherence condition $\mu < 1/(2k-1)$ and Tropp's ERC. Their computation remains simple as only inner products between the dictionary atoms are involved. It would therefore be definitely interesting to study how this type of condition evolve when Oxx has recovered l atoms of the support. This is part of our future work.

In this paper, we did also not investigate the case where the observed vector $\mathbf{y}$ is corrupted by some additive noise. This problem has been addressed in different contributions of the recent literature, see *e.g.*, [30], [31], and is interesting on its own. The extension of the proposed partial condition to noisy settings is part of our ongoing work.

APPENDIX A
PROOF OF THE RESULTS OF SECTION V

This section contains the proofs of Propositions 1 and 2 together with some useful technical lemmas.

Lemma 4 *Assume* $\mathbf{A}$ *satisfies the P-RIP(*$\underline{\delta}_{2,l}$,$\bar{\delta}_{2,l}$*) and let*

$$\mu_l^{OMP} \triangleq \max_{|\mathcal{Q}|=l} \max_{i\neq j} |\langle \tilde{\mathbf{a}}_i^{\mathcal{Q}}, \tilde{\mathbf{a}}_j^{\mathcal{Q}} \rangle|. \tag{42}$$

Then, we have

$$\mu_l^{OMP} \leq \frac{\bar{\delta}_{2,l} + \underline{\delta}_{2,l}}{2}. \tag{43}$$

Proof: By definition of $\bar{\delta}_{2,l}$ and $\underline{\delta}_{2,l}$ we must have for all $\mathcal{Q}$, $\mathcal{Q}'$ with $|\mathcal{Q}| = l$, $|\mathcal{Q}'| = 2$ and $\mathcal{Q}' \cap \mathcal{Q} = \emptyset$:

$$1 + \bar{\delta}_{2,l} \geq \lambda_{max}(\tilde{\mathbf{A}}_{\mathcal{Q}'}^T \tilde{\mathbf{A}}_{\mathcal{Q}'}), \tag{44}$$

$$1 - \underline{\delta}_{2,l} \leq \lambda_{min}(\tilde{\mathbf{A}}_{\mathcal{Q}'}^T \tilde{\mathbf{A}}_{\mathcal{Q}'}), \tag{45}$$

where $\lambda_{max}(\mathbf{M})$ (resp. $\lambda_{min}(\mathbf{M})$) denotes the largest (resp. smallest) eigenvalue of $\mathbf{M}$. Moreover, if $\mathcal{Q}' = \{i, j\}$, it is easy to check that the eigenvalues of $\tilde{\mathbf{A}}_{\mathcal{Q}'}^T \tilde{\mathbf{A}}_{\mathcal{Q}'}$ can be expressed as

$$\lambda(\tilde{\mathbf{A}}_{\mathcal{Q}'}^T \tilde{\mathbf{A}}_{\mathcal{Q}'}) = \frac{\|\tilde{\mathbf{a}}_i\|^2 + \|\tilde{\mathbf{a}}_j\|^2 \pm \Delta}{2},$$

where

$$\Delta = \sqrt{(\|\tilde{\mathbf{a}}_i\|^2 + \|\tilde{\mathbf{a}}_j\|^2)^2 + 4(\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j \rangle^2 - \|\tilde{\mathbf{a}}_i\|^2 \, \|\tilde{\mathbf{a}}_j\|^2)} \tag{46}$$

$$= \sqrt{(\|\tilde{\mathbf{a}}_i\|^2 - \|\tilde{\mathbf{a}}_j\|^2)^2 + 4\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j \rangle^2}. \tag{47}$$

Hence

$$\lambda_{max}(\tilde{\mathbf{A}}_{\mathcal{Q}'}^T \tilde{\mathbf{A}}_{\mathcal{Q}'}) - \lambda_{min}(\tilde{\mathbf{A}}_{\mathcal{Q}'}^T \tilde{\mathbf{A}}_{\mathcal{Q}'}) = \Delta \geq 2|\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j \rangle|.$$

Using (44)-(45), we thus obtain $\forall i, j \notin \mathcal{Q}$:

$$\bar{\delta}_{2,l} + \underline{\delta}_{2,l} \geq 2|\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j \rangle|. \tag{48}$$

Now, this inequality also holds if $i \in \mathcal{Q}$ or $j \in \mathcal{Q}$ since the right hand-side of (48) is then equal to zero. The result then follows from the definition of μ_l^{OMP}. ■

Lemma 5 *Let* $|\mathcal{Q}|{=}l$ *and* $\mathcal{Q}' \cap \mathcal{Q}'' = \emptyset$*, then* $\forall \mathbf{u} \in \mathbb{R}^{|\mathcal{Q}''|}$,

$$\|\tilde{\mathbf{A}}_{\mathcal{Q}'}^T \tilde{\mathbf{A}}_{\mathcal{Q}''} \mathbf{u}\| \leq \mu_l^{OMP} \sqrt{|\mathcal{Q}'||\mathcal{Q}''|}\, \|\mathbf{u}\|. \tag{49}$$

Proof: We have

$$\|\tilde{\mathbf{A}}_{\mathcal{Q}'}^T \tilde{\mathbf{A}}_{\mathcal{Q}''} \mathbf{u}\| = \sqrt{\sum_{i \in \mathcal{Q}'} \langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{A}}_{\mathcal{Q}''} \mathbf{u} \rangle^2} \tag{50}$$

$$= \sqrt{\sum_{i \in \mathcal{Q}'} \big(\sum_{j \in \mathcal{Q}''} u_j \, \langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j \rangle \big)^2} \tag{51}$$

$$\leq \sqrt{\sum_{i \in \mathcal{Q}'} \big(\sum_{j \in \mathcal{Q}''} |u_j| \, |\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j \rangle| \big)^2} \tag{52}$$

$$\leq \mu_l^{OMP} \sqrt{|\mathcal{Q}'|}\, \|\mathbf{u}\|_1 \tag{53}$$

$$\leq \mu_l^{OMP} \sqrt{|\mathcal{Q}'||\mathcal{Q}''|}\, \|\mathbf{u}\|. \tag{54}$$

■

Using Lemmas 4 and 5, we can now prove Propositions 1 and 2:

Proof: (Proposition 1) $\forall\, i \notin \mathcal{Q}^\star$, the following inequalities hold:

$$\|\tilde{\mathbf{A}}^\dagger_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\mathbf{a}}_i\|_1 \leq \sqrt{k-l}\,\|\tilde{\mathbf{A}}^\dagger_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\mathbf{a}}_i\|_2, \tag{55}$$

$$\leq \frac{\sqrt{k-l}}{1-\underline{\delta}_{k-l,l}}\|\tilde{\mathbf{A}}^T_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\mathbf{a}}_i\|_2, \tag{56}$$

$$\leq \frac{k-l}{1-\underline{\delta}_{k-l,l}}\mu_l^{OMP}, \tag{57}$$

$$\leq \frac{k-l}{1-\underline{\delta}_{k-l,l}}\frac{\bar{\delta}_{2,l}+\underline{\delta}_{2,l}}{2}, \tag{58}$$

where the first inequality follows from the equivalence of norms; the second from RIC properties (see [26, Proposition 3.1]); the third from Lemma 5 and the fourth from Lemma 4. ■

Proof: (Proposition 2) First, notice that $\mathbf{A}$ satisfies the P-RIP($\underline{\delta}_{q,0}$,$\bar{\delta}_{q,0}$) $\forall\, q$ with

$$\bar{\delta}_{q,0} = \underline{\delta}_{q,0} = (q-1)\mu, \tag{59}$$

see *e.g.*, [7, Lemma 2.3]. Hence, (22) is a consequence of the following inequalities:

$$\|\mathbf{P}^\perp_{\mathcal{Q}}\mathbf{A}_{\mathcal{Q}'}\mathbf{x}_{\mathcal{Q}'}\|^2 \leq \|\mathbf{A}_{\mathcal{Q}'}\mathbf{x}_{\mathcal{Q}'}\|^2 \leq (1+\bar{\delta}_{q,0})\|\mathbf{x}_{\mathcal{Q}'}\|^2. \tag{60}$$

Lower bound (23) may derived by noticing that

$$\|\mathbf{P}^\perp_{\mathcal{Q}}\mathbf{A}_{\mathcal{Q}'}\mathbf{x}_{\mathcal{Q}'}\|^2 = \|\mathbf{A}_{\mathcal{Q}'}\mathbf{x}_{\mathcal{Q}'}\|^2 - \|\mathbf{P}_{\mathcal{Q}}\mathbf{A}_{\mathcal{Q}'}\mathbf{x}_{\mathcal{Q}'}\|^2, \tag{61}$$

and

$$\|\mathbf{A}_{\mathcal{Q}'}\mathbf{x}_{\mathcal{Q}'}\|^2 \geq (1-\underline{\delta}_{q,0})\|\mathbf{x}_{\mathcal{Q}'}\|^2, \tag{62}$$

$$\|\mathbf{P}_{\mathcal{Q}}\mathbf{A}_{\mathcal{Q}'}\mathbf{x}_{\mathcal{Q}'}\|^2 = \|(\mathbf{A}^\dagger_{\mathcal{Q}})^T\mathbf{A}^T_{\mathcal{Q}}\mathbf{A}_{\mathcal{Q}'}\mathbf{x}_{\mathcal{Q}'}\|^2 \tag{63}$$

$$\leq \frac{\|\mathbf{A}^T_{\mathcal{Q}}\mathbf{A}_{\mathcal{Q}'}\mathbf{x}_{\mathcal{Q}'}\|^2}{1-\underline{\delta}_{l,0}}, \tag{64}$$

$$\leq \frac{\mu^2 lq\,\|\mathbf{x}_{\mathcal{Q}'}\|^2}{1-\underline{\delta}_{l,0}}, \tag{65}$$

where inequality (64) follows from standard relationships between the RIC properties of $\mathbf{A}$ and transforms of $\mathbf{A}$, and $1-\underline{\delta}_{l,0} \geq 0$ is a consequence of hypothesis $\mu < 1/(l-1)$ [7, Lemma 2.3]; (65) is a consequence of Lemma 5. ■

Appendix B

Proof of the results of section VI

Proof: (Lemma 1) The proof is recursive. Obviously, the result holds for $l = 0$ since $\mu_0^{OLS} = \mu$.

Let $\mathcal{Q}$ with $|\mathcal{Q}| = l \geq 1$ and consider $\mathcal{R}$ such that $\mathcal{Q} = \mathcal{R} \cup \{i\}$ with $|\mathcal{R}| = l - 1$. According to [13, Lemma 5], if $j \notin \mathcal{Q}$, we have the orthogonal decomposition

$$\tilde{\mathbf{b}}_j^{\mathcal{R}} = \eta_j \tilde{\mathbf{b}}_j^{\mathcal{Q}} + \langle \tilde{\mathbf{b}}_j^{\mathcal{R}}\,,\, \tilde{\mathbf{b}}_i^{\mathcal{R}} \rangle\, \tilde{\mathbf{b}}_i^{\mathcal{R}}. \tag{66}$$

Moreover, assumption $\mu < 1/l$ implies that $\mathbf{A}_{\mathcal{Q}\cup\{j\}}$, $\mathbf{A}_{\mathcal{R}\cup\{j\}}$ and $\mathbf{A}_{\mathcal{R}\cup\{i\}}$ are full column rank as families of at most $l+1$ atoms [7, Lemma 2.3] which in turn implies that $\tilde{\mathbf{a}}_j^{\mathcal{Q}}$, $\tilde{\mathbf{a}}_j^{\mathcal{R}}$ and $\tilde{\mathbf{a}}_i^{\mathcal{R}}$ are nonzero [13, Corollary 3]. Therefore, $\|\tilde{\mathbf{b}}_j^{\mathcal{Q}}\|$, $\|\tilde{\mathbf{b}}_j^{\mathcal{R}}\|$ and $\|\tilde{\mathbf{b}}_i^{\mathcal{R}}\|$ are all of unit norm, and then (66) yields $\eta_j = \pm\sqrt{1 - \langle \tilde{\mathbf{b}}_j^{\mathcal{R}}\,,\, \tilde{\mathbf{b}}_i^{\mathcal{R}} \rangle^2}$. If j and $j' \notin \mathcal{Q}$, it follows that

$$\langle \tilde{\mathbf{b}}_j^{\mathcal{Q}}\,,\, \tilde{\mathbf{b}}_{j'}^{\mathcal{Q}} \rangle = \frac{\langle \tilde{\mathbf{b}}_j^{\mathcal{R}}\,,\, \tilde{\mathbf{b}}_{j'}^{\mathcal{R}} \rangle - \langle \tilde{\mathbf{b}}_j^{\mathcal{R}}\,,\, \tilde{\mathbf{b}}_i^{\mathcal{R}} \rangle \langle \tilde{\mathbf{b}}_{j'}^{\mathcal{R}}\,,\, \tilde{\mathbf{b}}_i^{\mathcal{R}} \rangle}{\eta_j \eta_{j'}}. \tag{67}$$

Majorizing the inner products $|\langle \tilde{\mathbf{b}}_j^{\mathcal{R}}\,,\, \tilde{\mathbf{b}}_i^{\mathcal{R}} \rangle|$ by μ_{l-1}^{OLS} and using (33), we get:

$$\begin{aligned}
|\langle \tilde{\mathbf{b}}_j^{\mathcal{Q}}\,,\, \tilde{\mathbf{b}}_{j'}^{\mathcal{Q}} \rangle| &\leqslant \frac{\mu_{l-1}^{OLS} + (\mu_{l-1}^{OLS})^2}{1 - (\mu_{l-1}^{OLS})^2} && (68)\\
&= \frac{\mu_{l-1}^{OLS}}{1 - \mu_{l-1}^{OLS}} && (69)\\
&\leqslant \frac{\mu}{1 - (l-1)\mu - \mu} = \frac{\mu}{1 - l\mu} && (70)
\end{aligned}$$

leading to (33). ∎

Appendix C
Proof of the results of section VII

In this appendix, we provide a proof of Lemma 2. We use the notation $\mathcal{R}$ instead of $\mathcal{Q}$ to denote the current support. This change of notation is done to avoid confusion: in the rest of the paper, we have $|\mathcal{Q}| = l$ whereas in this appendix, the support cardinality may differ from l.

We first need to prove the following technical lemma:

Lemma 6 *Let $\mathbf{A}$ be defined as in* (36)*. Then, we have for all $\mathcal{R}$ with $|\mathcal{R}| < 2k - l$ and $i, j \notin \mathcal{R}$, $i \neq j$:*

$$\langle \tilde{\mathbf{a}}_i^{\mathcal{R}}, \tilde{\mathbf{a}}_j^{\mathcal{R}} \rangle = -\mu - \mu^2 \mathbf{1}_{|\mathcal{R}|}^T (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{1}_{|\mathcal{R}|}, \tag{71}$$

$$\|\tilde{\mathbf{a}}_i^{\mathcal{R}}\|^2 = 1 - \mu^2 \mathbf{1}_{|\mathcal{R}|}^T (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{1}_{|\mathcal{R}|}. \tag{72}$$

Proof: First recall that $\mathrm{spark}(\mathbf{A}) = 2k - l$ (see section VII). Therefore, $\mathbf{A}_{\mathcal{R}}$ is full rank when $|\mathcal{R}| < 2k - l$ and $\tilde{\mathbf{a}}_i^{\mathcal{R}}$ reads

$$\tilde{\mathbf{a}}_i^{\mathcal{R}} = \mathbf{P}_{\mathcal{R}}^{\perp} \mathbf{a}_i = \mathbf{a}_i - \mathbf{P}_{\mathcal{R}} \mathbf{a}_i = \mathbf{a}_i - \mathbf{A}_{\mathcal{R}} (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{A}_{\mathcal{R}}^T \mathbf{a}_i. \tag{73}$$

Using this expression, we have

$$\langle \tilde{\mathbf{a}}_i^{\mathcal{R}}, \tilde{\mathbf{a}}_j^{\mathcal{R}} \rangle = \langle \mathbf{a}_i, \mathbf{a}_j \rangle - \mathbf{a}_i^T \mathbf{A}_{\mathcal{R}} (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{A}_{\mathcal{R}}^T \mathbf{a}_j, \tag{74}$$

$$\|\tilde{\mathbf{a}}_i^{\mathcal{R}}\|^2 = 1 - \mathbf{a}_i^T \mathbf{A}_{\mathcal{R}} (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{A}_{\mathcal{R}}^T \mathbf{a}_i. \tag{75}$$

Taking into account that the inner product between any pair of atoms is equal to $-\mu$ by definition of $\mathbf{M} = \mathbf{A}^T \mathbf{A}$, we obtain the result.

■

Proof: (*Lemma 2*) We prove a result slightly more general than the statement of Lemma 2: for the dictionary defined as in (36), any subset $\mathcal{R}$ with $p \triangleq |\mathcal{R}| \leq 2k - l - 2$ can be reached by Oxx. Lemma 2 corresponds to the case $p = l$ ($p \leq 2k - l - 2$ is always satisfied as long as $l < k$).

The result is true for OLS by virtue of [13, Lemma 3] which states that any subset $\mathcal{R}$ of an *arbitrary* dictionary $\mathbf{A}$ is reachable as long as $|\mathcal{R}| \leq \mathrm{spark}(\mathbf{A}) - 2$. In particular, the latter condition is verified by the dictionary $\mathbf{A}$ and the subset $\mathcal{R}$ considered here since $\mathrm{spark}(\mathbf{A}) = 2k - l$ and $|\mathcal{R}| \leq 2k - l - 2$ by hypothesis.

We prove hereafter that the result is also true for OMP. Without loss of generality, we assume that the elements of $\mathcal{R}$ correspond to the first p atoms of $\mathbf{A}$ (the analysis performed hereafter remains valid for any other support $\mathcal{R}$ of cardinality p since the content of the Gram matrix $\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}}$ is constant whatever the support $\mathcal{R}$: see (39)). For arbitrary values of $\epsilon_2, \ldots, \epsilon_p > 0$, we define the following recursive construction:

- $\mathbf{y}_1 = \mathbf{a}_1$,
- $\mathbf{y}_{p+1} = \mathbf{y}_p + \epsilon_{p+1} \mathbf{a}_{p+1}$

($\mathbf{y}_{p+1}$ implicitly depends on $\epsilon_2, \ldots, \epsilon_{p+1}$). We show by recursion that for all $p \in \{1, \ldots, 2k-l-2\}$, there exist $\epsilon_2, \ldots, \epsilon_p > 0$ such that OMP with the dictionary defined as in (36) and $\mathbf{y}_p$ as input successively selects $\mathbf{a}_1, \ldots, \mathbf{a}_p$ during the first p iterations (in particular, the selection rule (9) always yields a unique maximum).

The statement is obviously true for $\mathbf{y}_1 = \mathbf{a}_1$. Assume that it is true for $\mathbf{y}_p$ ($p < 2k - l - 2$) with some $\epsilon_2, \ldots, \epsilon_p > 0$ (these parameters will remain fixed in the following). According to [13, Lemma 15], there exists $\epsilon_{p+1} > 0$ such that OMP with $\mathbf{y}_{p+1} = \mathbf{y}_p + \epsilon_{p+1} \mathbf{a}_{p+1}$ as input selects the same atoms as with $\mathbf{y}_p$ during the first p iterations, *i.e.*, $\mathbf{a}_1, \ldots, \mathbf{a}_p$ are successively chosen. At iteration p, the current active set reads $\mathcal{R} = \{1, \ldots, p\}$ and the corresponding residual takes the form

$$\mathbf{r}_{\mathcal{R}} = \epsilon_{p+1} \tilde{\mathbf{a}}_{p+1}^{\mathcal{R}}. \tag{76}$$

Thus, $\mathbf{a}_{p+1}$ is chosen at iteration $p+1$ if and only if

$$|\langle \tilde{\mathbf{a}}_i^{\mathcal{R}}, \tilde{\mathbf{a}}_{p+1}^{\mathcal{R}} \rangle| < \|\tilde{\mathbf{a}}_{p+1}^{\mathcal{R}}\|^2 \qquad \forall\, i \neq p+1. \tag{77}$$

Now, $|\mathcal{R}| = p < 2k-l$ by hypothesis, then Lemma 6 applies. Using (71)-(72), it is easy to see that (77) is equivalent to

$$\mu + 2\mu^2 \mathbf{1}_p^T (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{1}_p < 1. \tag{78}$$

Since $\mu = \frac{1}{2k-l-1} < \frac{1}{p+1} < \frac{1}{p-1}$, we have $(1-(p-1)\mu) > 0$. Then, [7, Lemma 2.3] and $\|\mathbf{1}_p\|^2 = p$ yield:

$$\mathbf{1}_p^T (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{1}_p \leq \frac{p}{1-(p-1)\mu}. \tag{79}$$

Using the majoration $\mu < 1/(p+1)$, it follows that:

$$\begin{aligned}
\mu + 2\mu^2 \mathbf{1}_p^T (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{1}_p &\leqslant \mu \left(1 + \frac{2\mu p}{1-(p-1)\mu}\right) && (80)\\
&= \mu \left(\frac{1+(p+1)\mu}{1-(p-1)\mu}\right) && (81)\\
&< \frac{1}{p+1} \left(\frac{2}{1-\frac{p-1}{p+1}}\right) = 1 && (82)
\end{aligned}$$

which proves that the condition (78), and then (77) is met. OMP therefore recovers the subset $\mathcal{R} \cup \{p+1\} = \{1, \dots, p+1\}$.

■

Proof: (Lemma 3) Using Lemma 6, we notice that $\tilde{\mathbf{C}}_{\backslash \mathcal{Q}} = \beta \tilde{\mathbf{A}}_{\backslash \mathcal{Q}}$ for some $\beta > 0$ since $\|\tilde{\mathbf{a}}_i\|$ does not depend on i and $\tilde{\mathbf{c}}_i \neq \mathbf{0}$. Defining $\mathbf{v} \triangleq \mathbf{1}_{2k-2l}$, we obtain

$$\begin{aligned}
\tilde{\mathbf{C}}_{\backslash \mathcal{Q}} \mathbf{v} &= \beta \tilde{\mathbf{A}}_{\backslash \mathcal{Q}} \mathbf{v} && (83)\\
&= \beta \tilde{\mathbf{A}} \mathbf{1}_{2k-l} = \beta \mathbf{P}_{\mathcal{Q}}^{\perp} \mathbf{A} \mathbf{1}_{2k-l} = \mathbf{0}_{2k-l-1}, && (84)
\end{aligned}$$

since $\mathbf{1}_{2k-l}$ belongs to the null-space of $\mathbf{A}$.

Let us partition the elements of $\mathbf{v} = \mathbf{1}_{2k-l}$ into two subsets $\mathcal{Q}_1 \cup \mathcal{Q}_2$ with $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \emptyset$ and $|\mathcal{Q}_1| = |\mathcal{Q}_2| = k-l$, and define $\mathbf{y} \triangleq \tilde{\mathbf{C}}_{\mathcal{Q}_1 \backslash \mathcal{Q}} \mathbf{1}_{k-l}$. According to (84), $\mathbf{y}$ rereads $-\tilde{\mathbf{C}}_{\mathcal{Q}_2 \backslash \mathcal{Q}} \mathbf{1}_{k-l}$, therefore $\mathbf{y}$ has two $(k-l)$-sparse representations with disjoint supports in $\tilde{\mathbf{C}}_{\backslash \mathcal{Q}}$.

■

REFERENCES

[1] B. K. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM J. Comput.*, vol. 24, pp. 227–234, Apr. 1995.

[2] S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic decomposition by Basis Pursuit,” *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1999.

[3] I. Gorodnitsky and D. R. Bhaskar, “Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm,” *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[4] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[5] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition,” in *Proc. 27th Ann. Asilomar Conf. Signals, Systems, and Computers*, 1993.

[6] S. Chen, S. A. Billings, and W. Luo, “Orthogonal least squares methods and their application to non-linear system identification,” *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.

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

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

[9] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[10] S. Foucart, “Stability and robustness of weak orthogonal matching pursuits,” in *AMS Spring, Southern Conference*. 2011, Springer Proceedings in Mathematics.

[11] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, “Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit,” *IEEE Trans. Information Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[12] T. Blumensath and M. E. Davies, “On the difference between orthogonal matching pursuit and orthogonal least squares,” Tech. Rep., University of Edinburgh, Mar. 2007.

[13] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, “Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares,” Tech. Rep., CRAN, IRCCYN, INRIA-Rennes Bretagne Atlantique, Nov. 2012.

[14] T. T. Cai, L. Wang, and G. Xu, “Stable recovery of sparse signals and an oracle inequality,” *IEEE Trans. Information Theory*, vol. 56, no. 7, pp. 3516–3522, July 2010.

[15] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.

[16] L. Rebollo-Neira and D. Lowe, “Optimized orthogonal matching pursuit approach,” *IEEE Signal Processing Letters*, vol. 9, no. 4, pp. 137–140, Apr. 2002.

[17] M. D. Plumbley, “On polar polytopes and the recovery of sparse representations,” *IEEE Trans. Information Theory*, vol. 53, no. 9, pp. 3188–3195, Sept. 2007.

[18] T. Zhang, “Sparse recovery with orthogonal matching pursuit under RIP,” *IEEE Trans. Information Theory*, vol. 57, no. 9, pp. 6215–6221, Sept. 2011.

[19] S. Huang and J. Zhu, “Recovery of sparse signals using OMP and its variants: convergence analysis based on RIP,” *Inverse Problems*, vol. 27, no. 3, pp. 035003+, Mar. 2011.

[20] E. Liu and V. N. Temlyakov, “The orthogonal super greedy algorithm and applications in compressed sensing,” *IEEE Trans. Information Theory*, vol. 58, no. 4, pp. 2040–2047, Apr. 2012.

[21] R. Maleh, "Improved RIP analysis of orthogonal matching pursuit," *CoRR*, vol. abs/1102.4311, 2011.

[22] Q. Mo and Y. Shen, "A remark on the restricted isometry property in orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 58, no. 6, pp. 3654–3656, June 2012.

[23] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Information Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[24] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395–407, May 2009.

[25] M. E. Davies and R. Gribonval, "Restricted isometry constants where ℓ^p sparse recovery can fail for $0 \ll p \leq 1$," *IEEE Trans. Information Theory*, vol. 55, no. 5, pp. 2203–2214, May 2009.

[26] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301–321, 2009.

[27] C. Dossal and S. Mallat, "Sparse spike deconvolution with minimum scale," in *Signal Processing with Adaptive Sparse Structured Representations (SPARS workshop)*, Rennes, France, Apr. 2005, pp. 1–4.

[28] R. Gribonval and M. Nielsen, "Beyond sparsity: Recovering structured representations by ℓ^1 minimization and greedy algorithms," *Adv. Comput. Math.*, vol. 28, no. 1, pp. 23–41, 2008.

[29] D. A. Lorenz, S. Schiffler, and D. Trede, "Beyond convergence rates: exact recovery with the tikhonov regularization with sparsity constraints," *Inverse Problems*, vol. 27, no. 8, pp. 1–17, Nov. 2011.

[30] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Information Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[31] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Trans. Information Theory*, vol. 57, no. 7, pp. 4680–4688, July 2011.