Exact recovery conditions for sparse representations with partial support information

Cédric Herzet, Charles Soussen, Jérôme Idier, Rémi Gribonval

Exact Recovery Conditions for Sparse Representations with Partial Support Information

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

Abstract—**We address the exact recovery of a k-sparse vector in the noiseless setting when some partial information on the support is available. This partial information takes the form of either a subset of the true support or an approximate subset including wrong atoms as well. We derive a new sufficient and worst-case necessary (in some sense) condition for the success of some procedures based on ℓ_p-relaxation, Orthogonal Matching Pursuit (OMP) and Orthogonal Least Squares (OLS). Our result is based on the coherence μ of the dictionary and relaxes the well-known condition $\mu < 1/(2k-1)$ ensuring the recovery of any k-sparse vector in the non-informed setup. It reads $\mu < 1/(2k-g+b-1)$ when the informed support is composed of g good atoms and b wrong atoms. We emphasize that our condition is complementary to some restricted-isometry based conditions by showing that none of them implies the other.**

Because this mutual coherence condition is common to all procedures, we carry out a finer analysis based on the Null Space Property (NSP) and the Exact Recovery Condition (ERC). Connections are established regarding the characterization of ℓ_p-relaxation procedures and OMP in the informed setup. First, we emphasize that the truncated NSP enjoys an ordering property when p is decreased. Second, the partial ERC for OMP (ERC-OMP) implies in turn the truncated NSP for the informed ℓ_1 problem, and the truncated NSP for $p < 1$.

Index Terms—**Partial support information; ℓ_p relaxation; Orthogonal Matching Pursuit; Orthogonal Least Squares; mutual 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

$$\min \|\mathbf{x}\|_0 \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y}, \qquad (P_0)$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$. Unfortunately, problem (P_0) is of combinatorial nature and, therefore, its resolution reveals to be intractable in most practical settings [1].

In order to address this issue, suboptimal (but tractable) algorithms have been proposed in the literature. Among the most popular procedures, let us mention: *i)* the algorithms based on the ℓ_p-relaxation of the ℓ_0 pseudo-norm; *ii)* the greedy algorithms, seen as suboptimal discrete search algorithms to address (P_0). On the one hand, the ℓ_p relaxation of (P_0) can be expressed as

$$\min \|\mathbf{x}\|_p \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y}, \qquad (P_p)$$

with $p \in (0, 1]$. Practical implementations of (P_1), also named Basis Pursuit [2] can be done optimally using linear programming algorithms, see *e.g.,* [3]; suboptimal procedures looking for a solution of (P_p) with $p \in (0, 1)$ are for example derived in [4], [5]. On the other hand, (forward) greedy procedures build a sparse vector by gradually increasing the active subset starting from the empty set. At each iteration, a new atom is appended to the active subset. Standard greedy procedures include, by increasing order of complexity, Matching Pursuit (MP) [6], Orthogonal Matching Pursuit (OMP) [7], Orthogonal Least Squares (OLS) [1], [8] and variants thereof, namely regularized OMP [9], weak OMP [10], stagewise OMP [11], etc.

In this paper, we focus on a variation of the sparse representation problem in which the decoder has some information (possibly erroneous) about the support of the sparse vector. This new paradigm has recently been introduced independently in several contributions and finds practical and analytical interests in many setups.

In [12]–[17], the authors focussed on the problem of recovering a sequence of sparse vectors with a strong dependence on their supports. This type of settings occurs for example in video compression or dynamic magnetic resonance imaging where the supports of the sought vectors commonly evolve slowly with time. More specifically, this set of papers focusses on an ℓ_1-relaxation of the following problem (or some slightly different variants thereof):

$$\min_{\mathbf{x}} \|\mathbf{x}_{\bar{\mathcal{Q}}}\|_0 \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y}, \qquad (P_{0,\mathcal{Q}})$$

where $\mathcal{Q}$ is an estimate of the sought support and $\mathbf{x}_{\bar{\mathcal{Q}}}$ represents the vector made up of the elements of $\mathbf{x}$ whose index is not in $\mathcal{Q}$.

More generally, the paradigm of sparse representation with side support information is of interest when some of the coefficients of the sparse decomposition can be easily identified a priori. For example, as mentioned in [15], in wavelet image processing, the coefficients weighting the scaling functions are likely to be non-zero and this information should be (ideally) taken into account in any processing. It also happens in many practical situations that some coefficients of the sparse decomposition (typically those with high amplitudes) can be identified by simple thresholding. This observation is the essence of the algorithm proposed in [18] where the authors look for a solution of (P_0) by successively applying thresholding operations on the solution of ℓ_1-relaxations of $(P_{0,\mathcal{Q}})$ to obtain a sequence of refined support estimates.

A slightly different, but related, perspective was considered in [19] for OMP and in [20] for both OMP and OLS. In these papers, the authors derived guarantees of success for OMP and OLS by assuming that atoms belonging to some subset $\mathcal{Q}$ have been selected during the first iterations. The goal of such approaches is to provide a finer analysis of OMP/OLS at intermediate iterations by noting that the standard uniform recovery conditions ensuring the success of OMP/OLS from the first iteration are rather pessimistic. It is quite obvious that the conditions derived in these papers also apply to situations where OMP/OLS are initialized with support $\mathcal{Q}$ (rather than with the empty support). In the sequel, we will refer to this variant of OMP (resp. OLS) as $\text{OMP}_{\mathcal{Q}}$ (resp. $\text{OLS}_{\mathcal{Q}}$). Clearly, $\text{OMP}_{\mathcal{Q}}$/$\text{OLS}_{\mathcal{Q}}$

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@univ-lorraine.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).

can be understood as greedy procedures looking for a solution of $(P_{0,\mathcal{Q}})$.

In this paper, we derive *uniform* recovery conditions for $\text{OMP}_{\mathcal{Q}}$/$\text{OLS}_{\mathcal{Q}}$ and ℓ_p-relaxed versions of $(P_{0,\mathcal{Q}})$ in the paradigm of partially-informed decoders. Our conditions are valid for $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$ where $\mathbf{x}^\star$ is any k-sparse vector. Let us briefly summarize the related literature.

First, generalizing the well-known "Null-Space Property" (NSP) derived in [21], the authors of [17], [18], [22] proposed a "truncated" NSP, which is a sufficient and worst-case necessary condition for the success of

$$\min_{\mathbf{x}} \|\mathbf{x}_{\bar{\mathcal{Q}}}\|_p \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{y}, \qquad (P_{p,\mathcal{Q}})$$

with $p \in [0, 1]$. Secondly, in [14]–[16], a series of sufficient conditions based on restricted isometry constants (RICs) were proposed to guarantee the success of $(P_{1,\mathcal{Q}})$ (or some variants thereof).

Concerning $\text{OMP}_{\mathcal{Q}}$/$\text{OLS}_{\mathcal{Q}}$, the authors in [20] derived a partial "Exact Recovery Condition" (ERC) extending Tropp's ERC to the partially-informed paradigm considered in this paper. The extended condition was shown to be sufficient but also worst-case necessary for the success of OMP/OLS when some support $\mathcal{Q}$ has been selected at an intermediate iteration. In [19], the authors proposed a sufficient condition based on RICs and depending on the number of "good" and "bad" atoms selected in $\mathcal{Q}$, that is the number of elements of $\mathcal{Q}$ which are (resp. are not) in the support of $\mathbf{x}^\star$.

In this paper, we derive a new simple recovery guarantee for $\text{OMP}_{\mathcal{Q}}$, $\text{OLS}_{\mathcal{Q}}$ and $(P_{p,\mathcal{Q}})$ for $p \in [0, 1]$. Our condition only depends on the mutual coherence of the dictionary μ and the number of good and bad atoms selected in the estimated support $\mathcal{Q}$:

$$\mu < \frac{1}{2k - g + b - 1}, \qquad (1)$$

where g (resp. b) denotes the number of "good" (resp. "bad") atoms in $\mathcal{Q}$. We show that (1) is sufficient for the success of $(P_{p,\mathcal{Q}})$ with $p \in [0, 1]$, $\text{OMP}_{\mathcal{Q}}$ and $\text{OLS}_{\mathcal{Q}}$. We emphasize moreover that (1) is worst-case necessary in the following sense: there exists a dictionary $\mathbf{A}$ with $\mu = \frac{1}{2k-g+b-1}$, a combination $\mathbf{y}$ of k columns of $\mathbf{A}$ and a support $\mathcal{Q}$ containing g good and b bad atoms such that neither $(P_{p,\mathcal{Q}})$ nor $\text{OMP}_{\mathcal{Q}}$/$\text{OLS}_{\mathcal{Q}}$ can recover $\mathbf{x}^\star$. Our condition generalizes, within the informed paradigm, the well-known condition $\mu < \frac{1}{2k-1}$ ensuring the success of Basis-Pursuit and OMP/OLS in the standard setup, see *e.g.*, [21], [23], [24]. In particular, we see that if the informed support $\mathcal{Q}$ contains more than 50% of good atoms, (1) leads to a weaker condition than its standard counterpart.

Although ensuring the success of $(P_{p,\mathcal{Q}})$ and $\text{OMP}_{\mathcal{Q}}$/$\text{OLS}_{\mathcal{Q}}$, condition (1) does not allow for a discrimination of the performance achievable by these algorithms. In order to address this question, we analyze some connections existing between the conditions previously proposed in the literature. First, we show that the truncated NSP derived in [17], [18], [22] enjoys a nesting property, namely: if the truncated NSP is satisfied for some $p \in [0, 1]$, then it is also verified for any other $q \in [0, p]$. From a worst-case point of view, this result tends to show that the resolution of $(P_{p,\mathcal{Q}})$ with $p \in [0, 1)$ is more favorable than ℓ_1-based approaches[1]. In particular, as a corrolary of this result, we have that all uniform conditions previously proposed for $(P_{1,\mathcal{Q}})$ also guarantee the success of $(P_{p,\mathcal{Q}})$ with $p \in [0, 1)$. Second, we establish that the partial ERC derived in [20] for $\text{OMP}_{\mathcal{Q}}$ is also a sufficient condition of success for $(P_{1,\mathcal{Q}})$. This generalizes the result derived by Tropp in the standard (non-informed) setup [23] to the partially-informed context considered in this paper. On the other hand, we emphasize that, unlike in the standard setup, such a connection does not hold between $(P_{1,\mathcal{Q}})$ and $\text{OLS}_{\mathcal{Q}}$.

Finally, we also study the connection between the proposed coherence-based condition (1) and some RIC-based conditions previously proposed in the context of orthogonal greedy algorithms. First, we illustrate the complementarity of (1) with the RIC guarantees proposed in [19] for OMP. We emphasize that no condition implies the other one. Secondly, we show that the RIC condition proposed in [19] for the success of $\text{OMP}_{\mathcal{Q}}$ also enjoys a form of quasi-tightness for both $\text{OMP}_{\mathcal{Q}}$ and $\text{OLS}_{\mathcal{Q}}$.

The rest of this paper is organized as follows. In section II, we set the notations that will be used throughout the paper. In section III, we review the main expressions defining the recursions of OMP/OLS and briefly discuss their application to the informed problem $(P_{0,\mathcal{Q}})$. Our contributions and their positioning within the current state of the art are discussed in section IV. Finally, the remaining sections and appendices are dedicated to the proofs of our results.

[1]We note however that, unlike the convex ℓ_1 problem, reaching the global minimum of ℓ_p problems is not guaranteed in practice.

II. NOTATIONS

The following notations will be used in this paper. $\langle \,.\, , \,.\, \rangle$ refers to the inner product between vectors and $\|\,.\,\|$ stands for the Euclidean norm. $\|\,.\,\|_p$ with $0 \leq p \leq 1$ will denote the ℓ_p (pseudo) norm. Of particular interest, the ℓ_0 pseudo norm, $\|\,.\,\|_0$, counts the number of non-zero elements in its argument. With a slight abuse of notation and for the sake of conciseness in some of our statements, we will assume that $\|\,.\,\|_0^0 \triangleq \|\,.\,\|_0$. We will use the notation $\mathbf{X}^\dagger$ to denote the pseudo-inverse of a matrix $\mathbf{X}$. For a full-rank and undercomplete matrix $\mathbf{X}$, 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, $\text{spark}(\mathbf{X})$ denotes the minimum number of columns from $\mathbf{X}$ that are linearly dependent [25]. $\mathbf{1}_m$ (resp $\mathbf{0}_m$) denotes the all-one (resp. all-zero) vector of dimension $m \times 1$. $\mathbf{I}_m$ is the $m \times m$ identity matrix. Caligraphic letters (as $\mathcal{Q}$, $\mathcal{R}$, $\mathcal{S}$, *etc*) will be used to denote some subsets of indices of the columns of the dictionary; the complementary of these sets in $\{1, \ldots, n\}$ will be denoted as $\bar{\mathcal{Q}}$, $\bar{\mathcal{R}}$, $\bar{\mathcal{S}}$, *etc*. In the main body of the paper, we will usually reserve the specific notations $\mathcal{Q}$ and $\mathcal{Q}^\star$ for, respectively, the informed support and the support of the sought sparse vector. $\mathbf{X}_{\mathcal{Q}}$ 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. Given a subset of the columns of the dictionary $\mathbf{A}_{\mathcal{Q}} \in \mathbb{R}^{m \times |\mathcal{Q}|}$, $\mathbf{P}_{\mathcal{Q}} = \mathbf{A}_{\mathcal{Q}}\mathbf{A}_{\mathcal{Q}}^\dagger$ and $\mathbf{P}_{\mathcal{Q}}^\perp = \mathbf{I}_m - \mathbf{P}_{\mathcal{Q}}$ denote the orthogonal projection operators onto $\text{span}(\mathbf{A}_{\mathcal{Q}})$ and $\text{span}(\mathbf{A}_{\mathcal{Q}})^\perp$, where $\text{span}(\mathbf{X})$ stands for the column span of $\mathbf{X}$, $\text{span}(\mathbf{X})^\perp$ is the orthogonal complement of $\text{span}(\mathbf{X})$. $\mathbf{r}^{\mathcal{Q}} = \mathbf{P}_{\mathcal{Q}}^\perp \mathbf{y} = \mathbf{y} - \mathbf{P}_{\mathcal{Q}}\mathbf{y}$ denotes the data residual induced by the orthogonal projection of $\mathbf{y}$ onto $\text{span}(\mathbf{A}_{\mathcal{Q}})$. Finally, we will use the notation $\ker(\mathbf{X}) \triangleq \text{span}(\mathbf{X}^T)^\perp$ to denote the null space of $\mathbf{X}$; $\ker_0(\mathbf{X})$ is the null-space of $\mathbf{X}$ minus the all-zero vector.

III. OMP AND OLS

In this section, we recall the selection rules defining OMP and OLS, and discuss their application to the support-informed problem $(P_{0,\mathcal{Q}})$. Throughout the paper, we will use the common acronym Oxx in statements that apply to both OMP and OLS.

First note that any vector $\mathbf{x}$ satisfying the constraint in (P_0) must have a support, say $\tilde{\mathcal{Q}} = \{i \,|\, x_i \neq 0\}$, such that $\mathbf{P}_{\tilde{\mathcal{Q}}}^\perp \mathbf{y} = \mathbf{0}_m$ since $\mathbf{y}$ must belong to $\text{span}(\mathbf{A}_{\tilde{\mathcal{Q}}})$. Hence, problem (P_0) can equivalently be rephrased as

$$\min_{\tilde{\mathcal{Q}}} |\tilde{\mathcal{Q}}| \quad \text{subject to } \mathbf{P}_{\tilde{\mathcal{Q}}}^\perp \mathbf{y} = \mathbf{0}_m. \qquad (2)$$

Oxx can be understood as an iterative procedure searching for a solution of (2) by generating a sequence of support estimates $\{\mathcal{Q}^{(\ell)}\}$ as

$$\mathcal{Q}^{(\ell+1)} = \mathcal{Q}^{(\ell)} \cup \{j\},$$

where

$$j \in \begin{cases} \arg\max_i |\langle \mathbf{a}_i, \mathbf{r}^{\mathcal{Q}^{(\ell)}} \rangle| & \text{for OMP} \\ \arg\min_i \|\mathbf{r}^{\mathcal{Q}^{(\ell)}\cup\{i\}}\| & \text{for OLS,} \end{cases} \tag{3}$$

$\mathbf{r}^{\mathcal{Q}^{(\ell)}} \triangleq \mathbf{P}^{\perp}_{\mathcal{Q}^{\ell}}\mathbf{y}$ is the current data residual and $\mathbf{a}_i$ is the ith column of $\mathbf{A}$. More specifically, Oxx adds one new atom to the estimated support at each iteration: OLS selects the atom minimizing the norm of the new residual $\mathbf{r}^{\mathcal{Q}^{(\ell)}\cup\{i\}}$ whereas OMP picks the atom maximizing the correlation with the current residual.

Oxx is commonly initialized with the empty set, *i.e.*, $\mathcal{Q}^{(0)} = \emptyset$. However, when some initial estimate of the support, say $\mathcal{Q}$, is available, nothing prevents us from initializing Oxx with $\mathcal{Q}^{(0)} = \mathcal{Q}$. We will refer to this variant of Oxx as $\text{Oxx}_{\mathcal{Q}}$[2]. On the one hand, $\text{Oxx}_{\mathcal{Q}}$ can readily be seen as a greedy procedure looking for a solution of $(P_{0,\mathcal{Q}})$. On the other hand, the behavior of $\text{Oxx}_{\mathcal{Q}}$ can be understood from a different perspective, namely the analysis of the standard Oxx algorithm at an intermediate iteration. Indeed, let us assume that Oxx has selected atoms in $\mathcal{Q}$ during the first $|\mathcal{Q}|$ iterations. Then, the next step of Oxx will be identical to the first iteration of $\text{Oxx}_{\mathcal{Q}}$. Although we will mainly stick to the former vision hereafter, the results that will be derived in the paper can be interpreted from these two perspectives.

In the sequel, we will often use a slightly different, equivalent, formulation of (3) based on orthogonal projections. For some subset of the column indices $\mathcal{R}$, let us define

$$\tilde{\mathbf{a}}_i^{\mathcal{R}} \triangleq \mathbf{P}^{\perp}_{\mathcal{R}}\mathbf{a}_i,$$
$$\tilde{\mathbf{b}}_i^{\mathcal{R}} \triangleq \begin{cases} \tilde{\mathbf{a}}_i^{\mathcal{R}}/\|\tilde{\mathbf{a}}_i^{\mathcal{R}}\| & \text{if } \tilde{\mathbf{a}}_i^{\mathcal{R}} \neq \mathbf{0}_m \\ \mathbf{0}_m & \text{otherwise.} \end{cases}$$

$\tilde{\mathbf{a}}_i^{\mathcal{R}}$ denotes the projection of $\mathbf{a}_i$ onto $\text{span}(\mathbf{A}_{\mathcal{R}})^{\perp}$ whereas $\tilde{\mathbf{b}}_i^{\mathcal{R}}$ is a normalized version of $\tilde{\mathbf{a}}_i^{\mathcal{R}}$. With these notations, (3) can be re-expressed as

$$j \in \arg\max_i |\langle \tilde{\mathbf{c}}_i^{\mathcal{Q}^{(\ell)}}, \mathbf{r}^{\mathcal{Q}^{(\ell)}} \rangle|, \tag{4}$$

where

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

The equivalence between (3) and (4) is straightforward for OMP by noticing that $\mathbf{r}^{\mathcal{Q}^{(\ell)}} \in \text{span}(\mathbf{A}_{\mathcal{Q}^{(\ell)}})^{\perp}$. We refer the reader to [26] for a detailed derivation of the equivalence for OLS.

In the sequel, we will use the notations $\tilde{\mathbf{A}}^{\mathcal{R}} \triangleq (\tilde{\mathbf{a}}_1^{\mathcal{R}}\,\tilde{\mathbf{a}}_2^{\mathcal{R}}\ldots\tilde{\mathbf{a}}_n^{\mathcal{R}}) \in \mathbb{R}^{m\times n}$, $\tilde{\mathbf{B}}^{\mathcal{R}} \triangleq (\tilde{\mathbf{b}}_1^{\mathcal{R}}\,\tilde{\mathbf{b}}_2^{\mathcal{R}}\ldots\tilde{\mathbf{b}}_n^{\mathcal{R}}) \in \mathbb{R}^{m\times n}$ and $\tilde{\mathbf{C}}^{\mathcal{R}} \triangleq (\tilde{\mathbf{c}}_1^{\mathcal{R}}\,\tilde{\mathbf{c}}_2^{\mathcal{R}}\ldots\tilde{\mathbf{c}}_n^{\mathcal{R}}) \in \mathbb{R}^{m\times n}$ to refer to the matrices whose columns are made up of the $\tilde{\mathbf{a}}_i^{\mathcal{R}}$'s, $\tilde{\mathbf{b}}_i^{\mathcal{R}}$'s and $\tilde{\mathbf{c}}_i^{\mathcal{R}}$'s, respectively. When the set of indices $\mathcal{R}$ corresponds to the informed support $\mathcal{Q}$, we will usually drop the dependence on $\mathcal{R}$ and use the simplified notations $\tilde{\mathbf{a}}_i$, $\tilde{\mathbf{b}}_i$, $\tilde{\mathbf{c}}_i$, $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{C}}$.

IV. Context and Main Results

Let us assume that $\mathbf{y}$ is a linear combination of k columns of $\mathbf{A}$ indexed by $\mathcal{Q}^{\star}$, that is

$$\mathbf{y} = \mathbf{A}\mathbf{x}^{\star} \qquad \text{with } x_i^{\star} \neq 0 \Leftrightarrow i \in \mathcal{Q}^{\star},\ |\mathcal{Q}^{\star}| = k. \tag{5}$$

[2]Let us note that, at the first iteration of $\text{Oxx}_{\mathcal{Q}}$, the residual is initialized by $\mathbf{r}^{\mathcal{Q}} \triangleq \mathbf{P}^{\perp}_{\mathcal{Q}}\mathbf{y}$, *i.e.*, the data $\mathbf{y}$ are being projected onto $\text{span}(\mathbf{A}_{\mathcal{Q}})^{\perp}$. In other words, $\text{Oxx}_{\mathcal{Q}}$ behaves similarly with $\mathbf{y}$ or $\mathbf{P}^{\perp}_{\mathcal{Q}}\mathbf{y}$ as input vector.

In this section, we review some standard conditions ensuring the correct reconstruction of $\mathbf{x}^{\star}$ (with and without partial information on the support) and recast our contributions within these existing results. We will use the following conventions: the atoms whose indices are in $\mathcal{Q}^{\star}$ will be referred to as "good" atoms whereas atoms whose indices are not in $\mathcal{Q}^{\star}$ will be dubbed "bad" atoms. If an initial estimate of the support $\mathcal{Q}^{\star}$ is available, say $\mathcal{Q}$, we will denote by $g \triangleq |\mathcal{Q}^{\star} \cap \mathcal{Q}|$ the number of good atoms in $\mathcal{Q}$ and by $b \triangleq |\bar{\mathcal{Q}}^{\star} \cap \mathcal{Q}|$ the number of bad atoms in $\mathcal{Q}$. We will always implicitly assume that $g < k$ since otherwise the informed problem $(P_{0,\mathcal{Q}})$ becomes trivial. Finally, we will suppose that the columns of $\mathbf{A}$ are normalized throughout the paper.

Our contributions will be both at the level of $\text{Oxx}_{\mathcal{Q}}$ and $(P_{p,\mathcal{Q}})$. In the next subsection we will focus on the conditions pertaining to Oxx and $\text{Oxx}_{\mathcal{Q}}$ whereas in subsection IV-B, we will describe the guarantees associated to the success of (P_p) and $(P_{p,\mathcal{Q}})$. Let us mention that our contributions are uniform conditions derived within the context of worst-case analyses. Hence, hereafter, we will essentially limit our discussion to the contributions in this line of thought.

Before proceeding, we recall the standard definitions of the restricted isometry constant (RIC) and mutual coherence that will be used in our discussion:

Definition 1 *The k-th order restricted isometry constant of $\mathbf{A}$ is the smallest non-negative value δ_k such that the following inequalities*

$$(1-\delta_k)\|\mathbf{x}\|^2 \leq \|\mathbf{A}\mathbf{x}\|^2 \leq (1+\delta_k)\|\mathbf{x}\|^2$$

are verified for any k-sparse vector $\mathbf{x}$.

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

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

A. Results and state-of-the-art conditions for Oxx and $\text{Oxx}_{\mathcal{Q}}$

OMP has been widely studied in the recent years, including worst-case [23], [27] and probabilistic analyses [28]. The existing exact recovery analyses of OMP were also adapted to several extensions of OMP, namely regularized OMP [9], weak OMP [10], and stagewise OMP [11]. Although OLS has been known in the literature for a few decades (often under different names [29]), 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 analysis of OLS more tricky. When the dictionary atoms are close to orthogonal, OLS and OMP have similar behaviors, as emphasized in [10]. On the contrary, for correlated dictionaries (*e.g.*, in ill-conditioned inverse problems), their behaviors significantly differ and OLS may be a better choice [20]. The above arguments motivate our analysis of both OMP and OLS.

Let us first rigorously define the notion of "success" that will be used for $\text{Oxx}_{\mathcal{Q}}$ throughout the paper:

Definition 3 (Successful recovery) *$\text{Oxx}_{\mathcal{Q}}$ with $\mathbf{y}$ defined in* (5) *as input succeeds if and only if it selects atoms in $\mathcal{Q}^{\star}\backslash\mathcal{Q}$ during the first $k-g$ iterations.*

In particular, this definition implies that $\text{Oxx}_{\mathcal{Q}}$ exactly reconstructs $\mathbf{x}^{\star}$ after $k-g$ iterations, as long as $\mathbf{A}_{\mathcal{Q}^{\star}\cup\mathcal{Q}}$ is full rank. When $\mathcal{Q} = \emptyset$, $\text{Oxx}_{\mathcal{Q}}$ reduces to the standard implementation of Oxx. In this case, Definition 3 matches the classical "k-step" analysis encountered in many contributions of the literature.

We will assume that, in special cases where the $\text{Oxx}_{\mathcal{Q}}$ selection rule yields multiple solutions including a wrong atom, that is

$$\max_{i\in\mathcal{Q}^\star}|\langle\tilde{\mathbf{c}}_i^{\mathcal{Q}^{(\ell)}},\mathbf{r}^{\mathcal{Q}^{(\ell)}}\rangle|=\max_{i\notin\mathcal{Q}^\star}|\langle\tilde{\mathbf{c}}_i^{\mathcal{Q}^{(\ell)}},\mathbf{r}^{\mathcal{Q}^{(\ell)}}\rangle|, \tag{6}$$

$\text{Oxx}_{\mathcal{Q}}$ systematically makes a bad decision. Hence, situation (6) always leads to a recovery failure.

Let us mention that the notion of successful recovery may be defined in a weaker sense than in Definition 3: Plumbley [30, 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 $\mathbf{y}\in\text{span}(\mathbf{A}_{\mathcal{Q}^\star})$), all atoms not belonging to $\mathcal{Q}^\star$ are then weighted by 0 in the solution vector (under some full-rank assumptions). Recently, a delayed recovery analysis of OMP using restricted-isometry constants was proposed in [31] and then extended to the weak OMP algorithm (including OLS) in [10].

To some extent, the definition of success considered in this paper also partially covers the setup of delayed recovery. Indeed, keeping in mind that $\text{Oxx}_{\mathcal{Q}}$ can be understood as a particular instance of Oxx in which atoms in $\mathcal{Q}$ have been selected during the first $g+b$ iterations, any condition ensuring the success of $\text{Oxx}_{\mathcal{Q}}$ in the sense of Definition 3 also guarantees the success of Oxx in $k+b$ iterations as long as atoms in $\mathcal{Q}$ are selected during the first $g+b$ iterations. Conditions under which g good and b bad atoms are selected during the first iterations are however not discussed in the rest of the paper.

Regarding k-step analyses, the first thoughtful theoretical study of OMP is due to Tropp, see [23, Th. 3.1 and Th. 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 [20]. In the latter paper, the authors carried out a narrow analysis of both OMP and OLS at any intermediate iteration of the algorithms. Their recovery conditions depend not only on $\mathcal{Q}^\star$ but also on the support $\mathcal{Q}^{(\ell)}$ estimated by Oxx at a given iteration ℓ. Recasting this analysis within the framework of sparse recovery with partial support information, $\mathcal{Q}^{(\ell)}$ plays the role of the estimated support $\mathcal{Q}$, and the main result in [20] can be rewritten as:

Theorem 1 (Soussen *et al.* 's partial ERC [20, Th. 3]) *Assume that $\mathbf{A}_{\mathcal{Q}^\star\cup\mathcal{Q}}$ is full rank with $|\mathcal{Q}^\star|=k$, $|\mathcal{Q}^\star\cap\mathcal{Q}|=g<k$, and $|\bar{\mathcal{Q}}^\star\cap\mathcal{Q}|=b$. If*

$$\max_{i\notin\mathcal{Q}^\star}\|\tilde{\mathbf{C}}^\dagger_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\mathbf{c}}_i\|_1<1, \tag{7}$$

then for any $\mathbf{y}\in\text{span}(\mathbf{A}_{\mathcal{Q}^\star})$, $\text{Oxx}_{\mathcal{Q}}$ only selects atoms in $\mathcal{Q}^\star\setminus\mathcal{Q}$ during the first $k-g$ iterations. Conversely, if (7) *does not hold, there exists $\mathbf{y}\in\text{span}(\mathbf{A}_{\mathcal{Q}^\star})$ for which $\text{Oxx}_{\mathcal{Q}}$ selects a bad atom $j\notin\mathcal{Q}^\star$ at the first iteration.*

The proof of Theorem 1 is a straightforward adaptation of [20, Th. 3]. For conciseness reasons, we therefore skip it. Let us just mention that the original formulation of [20, Th. 3] involves a vector $\mathbf{y}\in\text{span}(\mathbf{A}_{\mathcal{Q}^\star\cup\mathcal{Q}})$. Because any vector $\mathbf{y}\in\text{span}(\mathbf{A}_{\mathcal{Q}^\star\cup\mathcal{Q}})$ can be uniquely decomposed as $\mathbf{y}=\mathbf{y}_1+\mathbf{y}_2$ with $\mathbf{y}_1\in\text{span}(\mathbf{A}_{\mathcal{Q}^\star})$, $\mathbf{y}_2\in\text{span}(\mathbf{A}_{\bar{\mathcal{Q}}^\star\cap\mathcal{Q}})$ under full-rank conditions, and because $\text{Oxx}_{\mathcal{Q}}$ has the same behavior with $\mathbf{y}$ and $\mathbf{y}_1$ as inputs (the component $\mathbf{y}_2$ indexed by $\mathcal{Q}$ is not taken into account), both sufficient and necessary parts in Theorem 1 involve data vectors $\mathbf{y}\in\text{span}(\mathbf{A}_{\mathcal{Q}^\star})$.

Interestingly, when $\mathcal{Q}=\emptyset$, Theorem 1 reduces to Tropp's ERC [23]:

$$\max_{i\notin\mathcal{Q}^\star}\|\mathbf{A}^\dagger_{\mathcal{Q}^\star}\mathbf{a}_i\|_1<1, \tag{8}$$

which constitutes a sufficient and worst-case necessary condition for Oxx when no support information is available (or, equivalently, at the very first iteration of the algorithm).

A tight condition for the recovery of any k-sparse vector from any support estimate $\mathcal{Q}$ such that $|\mathcal{Q}^\star\cap\mathcal{Q}|=g$, $|\bar{\mathcal{Q}}^\star\cap\mathcal{Q}|=b$ can therefore be expressed as

$$\theta_{\text{Oxx}}(k,g,b)<1,$$

where

$$\theta_{\text{Oxx}}(k,g,b)\triangleq\max_{|\mathcal{Q}^\star|=k}\max_{\substack{|\bar{\mathcal{Q}}^\star\cap\mathcal{Q}|=b\\|\mathcal{Q}^\star\cap\mathcal{Q}|=g}}\Big\{\max_{i\notin\mathcal{Q}^\star}\|\tilde{\mathbf{C}}^\dagger_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\mathbf{c}}_i\|_1\Big\}. \tag{9}$$

Unfortunately, the main drawback of (9) stands in its cumbersome (combinatorial) evaluation. In order to circumvent this issue, 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 and those based on the mutual coherence of the dictionary.

The contributions [27], [32]–[36] provide RIC-based sufficient conditions for the exact recovery of the support $\mathcal{Q}^\star$ in k steps by OMP. The most recent and tightest results are due to Maleh [34], Mo&Shen [35] and Wang&Shim [36]. The authors proved that OMP succeeds in k steps if

$$\delta_{k+1}<\frac{1}{\sqrt{k}+1}. \tag{10}$$

In [35, Th. 3.2] and [36, Example 1], 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 (this result was actually first conjectured by Dai&Milenkovic in [37]). Let us mention that, by virtue of Theorem 1, these results remain valid for OLS. Indeed, when $\mathcal{Q}=\emptyset$, (8) is a worst-case necessary condition of exact recovery for both OMP and OLS. Moreover, since (10) is a uniform sufficient condition for OMP, (10) implies (8). Very recently, Karahanoglu and Erdogan [19] showed that the condition

$$\delta_{k+b+1}<\frac{1}{\sqrt{k-g}+1} \tag{11}$$

is sufficient for the success of $\text{OMP}_{\mathcal{Q}}$ when some support information is available at the decoder. Similar conditions are still not available for $\text{OLS}_{\mathcal{Q}}$ and remain an open problem in the literature.

In this paper, we emphasize that the RIC-based condition (11) also enjoys a type of worst-case necessity. In particular, the following result shows that (11) is almost tight for the success of $\text{OMP}_{\mathcal{Q}}$ in the following sense:

Lemma 1 (Quasi worst-case necessity of (11) for $\text{Oxx}_{\mathcal{Q}}$) *There exists a dictionary $\mathbf{A}$, a k-term representation $\mathbf{y}$ and a set $\mathcal{Q}$ with $|\mathcal{Q}^\star\cap\mathcal{Q}|=g$ and $|\bar{\mathcal{Q}}^\star\cap\mathcal{Q}|=b$, such that:* (i) *$\delta_{k+b+1}=\frac{1}{\sqrt{k-g}}$;* (ii) *$\text{Oxx}_{\mathcal{Q}}$ with $\mathbf{y}$ as input selects a bad atom at the first iteration.*

The proof of this lemma is reported to section IX. Let us mention that the result stated in Lemma 1 is valid for both OMP and OLS. Hence, although (11) has not been proved to be a sufficient condition for the success of $\text{OLS}_{\mathcal{Q}}$, this result shows that one cannot expect to achieve much better guarantees in terms of RICs for this algorithm.

Regarding uniform conditions based on the mutual coherence of the dictionary, Tropp showed in [23, Cor. 3.6] that

$$\mu < \frac{1}{2k-1} \tag{12}$$

is sufficient for the success of OMP in k steps. As a matter of fact, (12) therefore ensures that (8) is satisfied and thus also guarantees the success of OLS (Theorem 1 with $\mathcal{Q} = \emptyset$). Moreover, Cai&Wang recently showed in [38, Th. 3.1] that (12) 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[3] 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 uniform condition for Oxx [23, Cor. 3.6], [38, Th. 3.1]] *If* (12) *is satisfied, then Oxx succeeds in recovering any k-term representation. Conversely, there exist 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 provide a coherence-based sufficient and worst-case necessary condition for the success of $\text{Oxx}_\mathcal{Q}$. Our result generalizes Theorem 2 as follows:

Theorem 3 (μ-based uniform condition for $\text{Oxx}_\mathcal{Q}$) *Consider a k-term representation $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$ and a subset $\mathcal{Q}$ such that $|\mathcal{Q}^\star \cap \mathcal{Q}| = g$ and $|\bar{\mathcal{Q}}^\star \cap \mathcal{Q}| = b$. If $\mu < \frac{1}{2k-g+b-1}$ holds, then $\text{Oxx}_\mathcal{Q}$ recovers $\mathbf{x}^\star$ in $k-g$ iterations. Conversely, there exist a dictionary $\mathbf{A}$ and a k-term representation $\mathbf{y}$ such that:* (i) $\mu = \frac{1}{2k-g+b-1}$*;* (ii) *$\text{Oxx}_\mathcal{Q}$ with $\mathbf{y}$ as input selects a bad atom at the first iteration.*

The proof of this theorem is reported to sections V, VI and VIII. More specifically, we show in section V (resp. section VI) that (1) is sufficient for the success of $\text{OMP}_\mathcal{Q}$ (resp. $\text{OLS}_\mathcal{Q}$) in $k-g$ iterations. The proof of this sufficient condition significantly differs for $\text{OMP}_\mathcal{Q}$ and $\text{OLS}_\mathcal{Q}$. The result is shown for $\text{OMP}_\mathcal{Q}$ by deriving an upper bound on Soussen *et al.* 's ERC-OMP condition (7) as a function of the restricted isometry bounds of the projected dictionary $\tilde{\mathbf{A}}$. As for $\text{OLS}_\mathcal{Q}$, the proof is based on a connection between Soussen *et al.* 's ERC-OLS condition (7) and the mutual coherence of the normalized projected dictionary $\tilde{\mathbf{B}}$. Finally, in section VIII we prove that (1) is worst-case necessary for $\text{Oxx}_\mathcal{Q}$ in the sense specified in Theorem 3. This proof is common to both $\text{OMP}_\mathcal{Q}$ and $\text{OLS}_\mathcal{Q}$. If $b = 0$, we also prove a slightly stronger result by showing that the subset $\mathcal{Q}$ appearing in the converse part of Theorem 3 can indeed be "reached" by Oxx, initialized with the empty support. More formally, the following result holds:

Lemma 2 (μ-based partial uniform condition for Oxx) *There exist a dictionary $\mathbf{A}$, a k-term representation $\mathbf{y}$ and a set $\mathcal{Q} \subset \mathcal{Q}^\star$ with $|\mathcal{Q}| = g$, such that:* (i) $\mu = \frac{1}{2k-g-1}$*;* (ii) *Oxx with $\mathbf{y}$ as input selects atoms in $\mathcal{Q}$ during the first g iterations and an atom $\mathbf{a}_i$, $i \notin \mathcal{Q}^\star \backslash \mathcal{Q}$ at the $(g+1)$th iteration.*

This result is of in interest in the analysis of Oxx at intermediate iterations since it shows that if $\mu < \frac{1}{2k-g-1}$ is not satisfied, there exist scenarios where Oxx selects good atoms during the first g iterations and then fails at the subsequent step.

[3]and actually, any sparse representation algorithm.

B. Results and state-of-the-art conditions for (P_p) and $(P_{p,\mathcal{Q}})$

The performance associated to the resolution of (P_p) has been widely studied during the last decade. Among the noticeable works dealing with uniform and (worst-case) necessary conditions, one can first mention the seminal paper by Fuchs [39] in which the author showed that the success of (P_1) only depends on the sign of the nonzero components in $\mathbf{x}^\star$. More recently, Wang *et al.* provided in [40] sufficient and worst-case necessary conditions for the success of (P_p), with $p \in (0,1)$, depending on the sign-pattern of $\mathbf{x}^\star$. On the other hand, Gribonval&Nielsen derived in [21] the "Null-Space Property", a tight condition for the recovery of any k-sparse vector via (P_p).

Other conditions have also been proposed in terms of RIC and mutual coherence. On the one hand, the use of RIC-based conditions was ignited by Candes, Romberg and Tao in their seminal work [41]. Candes refined this result in [42] and some improvements were proposed by other authors in [5], [43].

On the other hand, guarantees for (P_0) and (P_1) based on the mutual coherence were early proposed in [44] for the particular case of sparse representations in a union of two orthogonal bases. Several authors later proved independently that condition (12) ensures the success of (P_0) and (P_1) for any k-sparse vector in arbitrary redundant dictionaries, see *e.g.*, [21], [39]. This condition was then shown to be valid for the success of (P_p) with $p \in [0,1]$ in [24]. Finally, Cai&Wang emphasized in [38, Th. 3.1] that (12) is also worst-case necessary (in some sense) for the success of (P_p).

Recently, several authors took a look at conditions ensuring the success of $(P_{p,\mathcal{Q}})$ when some partial information $\mathcal{Q}$ is available about the support $\mathcal{Q}^\star$. First, a "truncated" NSP generalizing the standard NSP has been derived in [17, Th. 2.1], [18, Th. 3.1] and [22, Th. 3.1]:

Theorem 4 (Truncated NSP) *Assume that* $\text{spark}(\mathbf{A}) > k+b$ *and let*

$$\theta_p(k,g,b) \triangleq \max_{|\mathcal{Q}^\star|=k} \max_{\substack{|\bar{\mathcal{Q}}^\star \cap \mathcal{Q}|=b \\ |\mathcal{Q}^\star \cap \mathcal{Q}|=g}} \max_{\mathbf{v}\in\ker_0(\mathbf{A})} \left\{ \frac{\|\mathbf{v}_{\mathcal{Q}^\star \backslash \mathcal{Q}}\|_p^p}{\|\mathbf{v}_{\overline{\mathcal{Q} \cup \mathcal{Q}^\star}}\|_p^p} \right\}. \tag{13}$$

For any $p \in [0,1]$, if

$$\theta_p(k,g,b) \le 1, \tag{14}$$

then any k-sparse vector $\mathbf{x}^\star$ is a minimizer of $(P_{p,\mathcal{Q}})$ for any partial support estimate $\mathcal{Q}$ such that $|\mathcal{Q}^\star \cap \mathcal{Q}| = g$ and $|\bar{\mathcal{Q}}^\star \cap \mathcal{Q}| = b$. Moreover, if the inequality in (14) *holds strictly, $\mathbf{x}^\star$ is the unique solution of $(P_{p,\mathcal{Q}})$. Conversely, if* (14) *is not satisfied, there exist a k-sparse vector $\mathbf{x}^\star$ and a support estimate $\mathcal{Q}$ satisfying $|\mathcal{Q}^\star \cap \mathcal{Q}| = g$ and $|\bar{\mathcal{Q}}^\star \cap \mathcal{Q}| = b$, such that $\mathbf{x}^\star$ is not a minimizer of $(P_{p,\mathcal{Q}})$ with $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$ as input.*

We note that the denominator in the right-hand side of (13) is always non-zero because of the hypothesis $\text{spark}(\mathbf{A}) > k+b = |\mathcal{Q}^\star \cup \mathcal{Q}|$ (see also Appendix C). The direct part of Theorem 4 is proved in [18] and [22] for $(P_{1,\mathcal{Q}})$ and $(P_{p,\mathcal{Q}})$, respectively. In [17], the authors demonstrated both the direct and converse parts of Theorem 4 for $(P_{1,\mathcal{Q}})$. We verified that the converse part of Theorem 4 also holds for $(P_{p,\mathcal{Q}})$, $p \in [0,1]$. The proof is very similar to the exposition in [17] and [21], and is therefore not reported here. We note that Theorem 4 reduces to the standard NSP as soon as $g = b = 0$.

Several authors also proposed recovery guarantees in terms of RICs, see [14]–[16]. In [14], the authors identified a sufficient condition for the success of $(P_{1,\mathcal{Q}})$ and show that the latter condition is weaker than a condition derived in [41] for the non-informed setting as long as $\mathcal{Q}$ contains a "sufficiently" large number of good atoms. This result was later extended by Jacques [15] to the cases of compressible signals and noisy observations. Finally, in [16], Friedlander *et al.* generalized the RIC condition derived in [41], [42] to the partially-informed paradigm considered in this paper. In particular, the authors showed that the following condition[4]

$$\delta_{2k} < \left(1+\sqrt{2\left(1+\frac{b-g}{k}\right)}\right)^{-1}, \tag{15}$$

is sufficient for the success of $(P_{1,\mathcal{Q}})$. Interestingly, if $g=b=0$, one recovers the standard condition $\delta_{2k} < (1+\sqrt{2})^{-1}$ by Candes for the success of (P_1), [42]. Finally, we also mention the work by Khajehnejad *et al.* [13] where a Grassman angle approach was used to characterize a class of signal which can be recovered by (a variant of) $(P_{1,\mathcal{Q}})$.

In this paper, we will show that the quantities $\theta_p(k,g,b)$ involved in the truncated NSP obey an ordering property and can be related to the partial ERC stated in Theorem 1 (see Theorems 6 and 7 below). As a consequence of these results, together with Theorem 3, we obtain that a coherence-based condition, similar to the one obtained for Oxx$_\mathcal{Q}$, holds for the success of $(P_{p,\mathcal{Q}})$:

Theorem 5 (μ-based uniform condition for $(P_{p,\mathcal{Q}})$) *Consider a k-term representation $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$ and a support $\mathcal{Q}$ such that $|\mathcal{Q}^\star \cap \mathcal{Q}| = g$ and $|\bar{\mathcal{Q}}^\star \cap \mathcal{Q}| = b$. If $\mu < \frac{1}{2k-g+b-1}$ holds, then $\mathbf{x}^\star$ is the unique minimizer of $(P_{p,\mathcal{Q}})$. Conversely, there exist a dictionary $\mathbf{A}$ and a k-term representation $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$ such that:* (i) *$\mu = \frac{1}{2k-g+b-1}$;* (ii) *$\mathbf{x}^\star$ is not the unique minimizer of $(P_{p,\mathcal{Q}})$.*

The direct part of Theorem 5 is proved in section VII. The converse part will be shown in section VIII.

Interestingly, similar to the result by Friedlander *et al.* in (15), we notice that our coherence-based condition becomes weaker than its standard counterpart (12) as soon as $g > b$, that is, when at least 50% of the atoms of $\mathcal{Q}$ belongs to $\mathcal{Q}^\star$. In other words, the success of $(P_{p,\mathcal{Q}})$ is ensured under conditions less restrictive than for (P_p) as soon as $\mathcal{Q}$ provides a "sufficiently reliable" information about $\mathcal{Q}^\star$.

C. Relationships between conditions for $(P_{p,\mathcal{Q}})$ and Oxx$_\mathcal{Q}$

In this section, we discuss the implications (or non-implications) existing between some of the conditions mentioned above. First, we emphasize that an ordering property, similar to the one derived by Gribonval&Nielsen in [24, Lemma 7] for (P_p), still holds for the truncated NSPs defined in Theorem 4:

Theorem 6 (Ordering property of truncated NSPs) *If $0 \le q \le p \le 1$ and $\operatorname{spark}(\mathbf{A}) > k+b$, the following ordering property holds:*

$$\theta_0(k,g,b) \le \theta_q(k,g,b) \le \theta_p(k,g,b) \le \theta_1(k,g,b). \tag{16}$$

The proof of this result is reported to section VII. Clearly, one recovers Gribonval&Nielsen's ordering property as a particular case of (16) as soon as $g=b=0$. This ordering property implies that any uniform condition for $(P_{p,\mathcal{Q}})$ is also a sufficient condition of success for $(P_{q,\mathcal{Q}})$ with $q \in [0,p]$. In particular, the guarantees derived in [14]–[16] for $(P_{1,\mathcal{Q}})$ also ensure the success of $(P_{p,\mathcal{Q}})$ for $p \in [0,1]$.

Secondly, we show that the truncated NSPs share some connections with the partial ERC for OMP defined in (9). Specifically, we have

Theorem 7 *If $\operatorname{spark}(\mathbf{A}) > k+b$, then*

$$\theta_1(k,g,b) \le \theta_{\mathrm{OMP}}(k,g,b). \tag{17}$$

The proof of this result is reported to section VII. This inclusion generalizes Tropp's result [23, Th. 3.3] to the paradigm of sparse representation with partial support information, namely ERC-OMP is a sufficient condition of success for $(P_{1,\mathcal{Q}})$ (and thus for any $(P_{p,\mathcal{Q}})$ with $p \in [0,1]$ by virtue of Theorems 4 and 6). As an important by-product of this observation, it turns out that any uniform guarantee of success for OMP$_\mathcal{Q}$ is also a sufficient condition of success for $(P_{p,\mathcal{Q}})$.

It is noticeable that an ordering similar to (17) does not generally hold between $\theta_1(k,g,b)$ and $\theta_{\mathrm{OLS}}(k,g,b)$ for all k,g,b. Indeed, on the one hand, $\theta_{\mathrm{OLS}}(k,0,0) \ge \theta_1(k,0,0)$ since $\theta_{\mathrm{OLS}}(k,0,0) = \theta_{\mathrm{OMP}}(k,0,0)$. On the other hand, we exhibit hereafter an example in which $\theta_{\mathrm{OLS}}(k,g,b) < \theta_1(k,g,b)$:

Example 1 *In this example, we construct a dictionary such that*

$$\theta_1(k,g,b) > 1,$$
$$\theta_{\mathrm{OLS}}(k,g,b) < 1,$$

for some k,g,b. Let $n \ge 3$ and define the matrix

$$\mathbf{G} = \begin{pmatrix} \mathbf{I}_{n-2} & \beta\mathbf{1}_{n-2} & \beta\mathbf{1}_{n-2} \\ \beta\mathbf{1}_{n-2}^T & 1 & \alpha \\ \beta\mathbf{1}_{n-2}^T & \alpha & 1 \end{pmatrix}, \qquad \text{for } \alpha,\ \beta \in \mathbb{R},$$

which will play the role of the Gram matrix of the dictionary, that is $\mathbf{G} = \mathbf{A}^T\mathbf{A}$. Since $\mathbf{G}$ is symmetric it allows for the following eigenvalue decomposition:

$$\mathbf{G} = \mathbf{U}\Lambda\mathbf{U}^T,$$

where $\mathbf{U}$ (resp. Λ) is the unitary matrix whose columns are the eigenvectors (resp. the diagonal matrix of eigenvalues) of $\mathbf{G}$. Letting

$$\alpha = \frac{1}{2}\gamma^2(n-2) - 1,$$
$$\beta = -\frac{\gamma}{2},$$

with

$$|\gamma| < (n-2)^{-1},$$

it is easy to see that $\mathbf{G}$ is a semi-definite positive matrix with one single zero eigenvalue. The zero eigenvalue is located in the lower-right corner of Λ and the corresponding eigenvector writes, up to a normalization factor, as

$$\mathbf{v} = [\gamma\mathbf{1}_{n-2}^T \quad 1 \quad 1]^T. \tag{18}$$

We define $\mathbf{A} \in \mathbb{R}^{n-1\times n}$ as

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

where $\Upsilon \in \mathbb{R}^{n-1\times n}$ is such that

$$\Upsilon(i,j) = \begin{cases} \sqrt{\Lambda(i,i)} & \text{if } i=j, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, $\Upsilon^T\Upsilon = \Lambda$ and $\mathbf{A}^T\mathbf{A} = \mathbf{U}\Upsilon^T\Upsilon\mathbf{U}^T = \mathbf{U}\Lambda\mathbf{U}^T = \mathbf{G}$.

Now, $\mathbf{A}$ is such that $\theta_{\mathrm{OLS}}(k,g,b) < 1 < \theta_1(k,g,b)$ for $k=2$, $g=1$, $b=0$. Indeed, on the one hand it can easily be seen that $\ker(\mathbf{A}) = \ker(\mathbf{G})$ and $\ker(\mathbf{A})$ corresponds therefore to the one-dimensional subspace defined by $\mathbf{v}$ in (18)*. This implies that* $\operatorname{spark}(\mathbf{A}) = n$*. Moreover, considering $\mathcal{Q}^\star = \{n-1,n\}$ and $\mathcal{Q} = \{n-1\}$, we have*

$$\theta_1(2,1,0) \ge \frac{\|\mathbf{v}_{\mathcal{Q}^\star\setminus\mathcal{Q}}\|_1}{\|\mathbf{v}_{\overline{\mathcal{Q}^\star\cup\mathcal{Q}}}\|_1} = \frac{1}{(n-2)\gamma} > 1$$

[4]We have adapted the formulation of the condition derived in [16] to the particular setup and notations considered in this paper.

where the first inequality follows from the definition of $\theta_1(2,1,0)$ *and the last one from the fact that* $\gamma < (n-2)^{-1}$.

On the other hand, since $\text{spark}(\mathbf{A}) = n \geq k+1$ *and there is only one atom in* $\mathcal{Q}^\star \backslash \mathcal{Q}$*, we have from [20, Th. 6] that necessarily* $\theta_{\text{OLS}}(2,1,0) < 1$.

In the previous example, we provided a simple scenario where $\text{OLS}_\mathcal{Q}$ succeeds in recovering $\mathbf{x}^\star$ when $g = k-1$. It can be observed that $(P_{0,\mathcal{Q}})$ also succeeds in this particular example. Indeed, using the definition of $\mathbf{v}$ in (18), we have

$$\theta_0(2,1,0) = \frac{1}{(n-2)} < 1.$$

More generally, it can be shown that there exists an equivalence between the success of $(P_{0,\mathcal{Q}})$ and $\text{OLS}_\mathcal{Q}$ when $g = k-1$. This follows from the fact that the problem resolved by $\text{OLS}_\mathcal{Q}$ when $g = k-1$, that is (3), is exactly equivalent to $(P_{0,\mathcal{Q}})$. From a more technical point of view, it can easily be seen that condition $\theta_0(k, k-1, b) < 1$ can be rephrased as

$$k + b + 1 < \text{spark}(\mathbf{A}).$$

Now, by slightly extending the arguments developed in [20, Th. 6], the latter condition is also sufficient and worst-case necessary for the success of $\text{OLS}_\mathcal{Q}$ when $g = k-1$. This observation thus demonstrates the optimality of $\text{OLS}_\mathcal{Q}$ when the informed support contains all the correct atoms but one.

D. Non-implication between the mutual and RIP conditions for $\text{Oxx}_\mathcal{Q}$

In Theorem 3, we derived a novel guarantee of success for $\text{Oxx}_\mathcal{Q}$ in terms of mutual coherence of the dictionary. On the other hand, other conditions were previously proposed in terms of RICs for $\text{OMP}_\mathcal{Q}$, see (11). Hence, one legitimate question arises: is there any implication from (1) to (11) or vice versa? We show hereafter that the answer to this question is negative. In particular, we exhibit two particular instances of dictionary such that (1) is satisfied but (11) is not, and vice versa. We construct our examples in the case where $b = 0$ for the sake of conciseness. Similar constructions can however be applied to derive examples in the general case $b \neq 0$.

Example 2 (A satisfies (1) but not (11)) *Let us consider* $\mathbf{A} \in \mathbb{R}^{(k+1)\times(k+1)}$ *such that*

$$\mathbf{G} \triangleq \mathbf{A}^T\mathbf{A}$$

and

$$G_{i,j} = \left\{ \begin{array}{ll} -\mu & i \neq j \\ 1 & i = j \end{array} \right.$$

with $\mu \leq 1/k$*. We have therefore*

$$\begin{aligned} &\lambda_{max}(\mathbf{G}) = 1 + \mu \qquad \textit{(with multiplicity k)}, \\ &\lambda_{min}(\mathbf{G}) = 1 - k\mu, \end{aligned}$$

and

$$\begin{aligned} \delta_{k+1} &= \max\{1 - \lambda_{min}(\mathbf{G}), \lambda_{max}(\mathbf{G}) - 1\} \\ &= k\mu. \end{aligned}$$

We can freely set $\mu = \alpha/(2k-g-1)$ *with* $0 \leq g < k$ *and* $\alpha \in (0,1)$ *since this yields* $\mu < 1/k$*. Then,* μ *trivially satisfies* (1)*. On the other hand,* δ_{k+1} *can be written as*

$$\delta_{k+1} = \frac{\alpha k}{2k-g-1} \geq \alpha/2. \tag{19}$$

For any $g < k-1$*, there exist* $\alpha \in (0,1)$ *and* k *such that* (11) *is not verified. For example, for* k *sufficiently large and fixed* $g < k-1$*,* δ_{k+1} *in* (19) *does not satisfy* (11) *since the right-hand side of* (11) *tends towards 0 when* k *tends to infinity.*

Example 3 (A satisfies (11) but not (1)) *Let*

$$\mathbf{A} \triangleq \left(\begin{array}{ccc} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{H} \end{array} \right) \in \mathbb{R}^{(k+1)\times(k+1)} \tag{20}$$

be such that

$$\begin{aligned} &\mathbf{a}_1^T\mathbf{a}_2 = \mu, \\ &\mathbf{H}^T\mathbf{a}_1 = \mathbf{H}^T\mathbf{a}_2 = \mathbf{0}_{k-1}, \\ &\mathbf{H}^T\mathbf{H} = \mathbf{I}_{k-1}. \end{aligned}$$

Then, we easily have

$$\begin{aligned} &\lambda_{max}(\mathbf{G}) = 1 + \mu, \\ &\lambda_{min}(\mathbf{G}) = 1 - \mu, \end{aligned}$$

and

$$\begin{aligned} \delta_{k+1} &= \max\{1 - \lambda_{min}(\mathbf{G}), \lambda_{max}(\mathbf{G}) - 1\} \\ &= \mu. \end{aligned}$$

Let us set $\delta_{k+1} = \mu = \alpha/(\sqrt{k-g}+1)$ *with* $\alpha \in (0,1)$*. Then,* δ_{k+1} *trivially satisfies* (11)*. On the other hand,* $\mu > 1/(2k-g-1)$ *holds for sufficiently large* k *and a fixed value of* $g < k$*.*

Finally, we mention that, following the same procedures as above, one can derive examples for which (15) is satisfied but (1) is not for some value of k, g, b, and vice versa. The details are however not reported here for the sake of conciseness.

V. Sufficiency of (1) for $\text{OMP}_\mathcal{Q}$

In this section, we prove the direct part of Theorem 3 for $\text{OMP}_\mathcal{Q}$. The result is a direct consequence of Proposition 1 stated below, which provides an upper bound on the left-hand side of (7) only depending on the mutual coherence of $\mathbf{A}$:

Proposition 1 *Let* $\mathcal{Q}^\star$ *and* $\mathcal{Q}$ *be such that* $|\mathcal{Q}^\star| = k$*,* $|\mathcal{Q}^\star \cap \mathcal{Q}| = g$ *and* $|\bar{\mathcal{Q}}^\star \cap \mathcal{Q}| = b$*. If*

$$\mu < \frac{1}{k+b-1},$$

then

$$\max_{i \notin \mathcal{Q}^\star} \|\tilde{\mathbf{A}}^\dagger_{\mathcal{Q}^\star \backslash \mathcal{Q}} \tilde{\mathbf{a}}_i\|_1 \leq \frac{(k-g)\mu}{1-(k+b-1)\mu}. \tag{21}$$

The sufficient condition for $\text{OMP}_\mathcal{Q}$ stated in Theorem 3 then derives from Proposition 1 and Theorem 1. Indeed, we see from Proposition 1 that

$$\frac{(k-g)\mu}{1-(k+b-1)\mu} < 1 \tag{22}$$

implies (7). Moreover, by reorganizing the latter expression, it is easy to see that (22) is equivalent to (1). To prove Theorem 3 it thus remains to apply Theorem 1. Now, the full-rankness of $\mathbf{A}_{\mathcal{Q}^\star \cup \mathcal{Q}}$ in the hypotheses of Theorem 1 is implicitly enforced by (1). Indeed, as shown in [23, Lemma 2.3], $\mu < \frac{1}{k+b-1}$ implies that $\mathbf{A}_{\mathcal{Q}^\star \cup \mathcal{Q}}$ is full rank whenever $|\mathcal{Q}^\star \cup \mathcal{Q}| = k+b$. Hence, since $k+b-1 < 2k-g+b-1$, (1) in turn implies that any submatrix $\mathbf{A}_{\mathcal{Q}^\star \cup \mathcal{Q}}$ with $|\mathcal{Q}^\star \cup \mathcal{Q}| = k+b$ is full rank. Then, applying Theorem 1, we have that (1) is sufficient for the success of $\text{OMP}_\mathcal{Q}$ in $k-g$ iterations.

Before proving Proposition 1, we need to define some quantities characterizing the *projected* dictionary $\tilde{\mathbf{A}}$ appearing in the implementation of OMP (see (4)) and state some useful lemmas. In the following definition, we generalize the concept of restricted isometry property (RIP) [41] to projected dictionaries, under the name projected RIP (P-RIP):

Definition 4 *Dictionary* $\mathbf{A}$ *satisfies the P-RIP(*$\underline{\delta}_{q,l}$*,*$\bar{\delta}_{q,l}$*) if and only if* $\forall \mathcal{R}, \mathcal{S}$ *with* $|\mathcal{R}| = l$, $|\mathcal{S}| = q$, $\mathcal{R} \cap \mathcal{S} = \emptyset$, $\forall \mathbf{x}_{\mathcal{S}}$ *we have*

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

The definition of the standard (asymmetric) restricted isometry constants corresponds to the tightest possible bounds when $l = 0$ (see *e.g.*, [5], [45]). 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 the *projected* dictionary $\tilde{\mathbf{A}}^{\mathcal{R}}$. Note that $\bar{\delta}_{q,l}$ might be negative since the columns of $\tilde{\mathbf{A}}^{\mathcal{R}}$ are not normalized ($\|\tilde{\mathbf{a}}_i^{\mathcal{R}}\| \leq 1$). Note also that many well-known properties of the standard restricted isometry constants (see [46, Proposition 3.1] for example) remain valid for $\underline{\delta}_{q,l}$ and $\bar{\delta}_{q,l}$.

The next lemma provides an upper bound on the left-hand side of (7) only depending on the P-RIP constants:

Lemma 3 *Let* $\mathcal{Q}^\star$ *and* $\mathcal{Q}$ *be such that* $|\mathcal{Q}^\star| = k$, $|\mathcal{Q}^\star \cap \mathcal{Q}| = g$ *and* $|\bar{\mathcal{Q}}^\star \cap \mathcal{Q}| = b$. *If* $\underline{\delta}_{k-g,g+b} < 1$, *then*

$$\max_{i \notin \mathcal{Q}^\star} \|\tilde{\mathbf{A}}_{\mathcal{Q}^\star \backslash \mathcal{Q}}^{\dagger}\tilde{\mathbf{a}}_i\|_1 \leq (k-g)\,\frac{\bar{\delta}_{2,g+b} + \underline{\delta}_{2,g+b}}{2(1 - \underline{\delta}_{k-g,g+b})}. \tag{23}$$

The proof of Lemma 3 is reported to Appendix A. The next lemma provides some possible values for $\underline{\delta}_{q,l}$ and $\bar{\delta}_{q,l}$ as a function of the mutual coherence of $\mathbf{A}$:

Lemma 4 *If* $\mu < 1/(l-1)$, *then* $\mathbf{A}$ *satisfies the P-RIP(*$\underline{\delta}_{q,l}$*,*$\bar{\delta}_{q,l}$*) for any* $q \geq 0$ *with*

$$\bar{\delta}_{q,l} = (q-1)\mu, \tag{24}$$

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

The proof of this result is reported to Appendix A. We are now ready to prove Proposition 1:

Proof: (Proposition 1) We rewrite the right-hand side of (23) as a function of μ. From Lemma 4, we have that $\mathbf{A}$ satisfies the P-RIP($\underline{\delta}_{q,l}$,$\bar{\delta}_{q,l}$) with constants defined in (24)-(25) as long as

$$\mu < \frac{1}{l-1}. \tag{26}$$

Now, we have $\mu < 1/(k+b-1)$ by hypothesis, which implies $\mu < 1/(g+b-1)$. Thus, Lemma 4 can be applied with $l = g+b$. Using (24) and (25), we calculate that:

$$\begin{aligned}
\frac{\bar{\delta}_{2,g+b} + \underline{\delta}_{2,g+b}}{2} &= \mu + \frac{\mu^2(g+b)}{1-(g+b-1)\mu} \\
&= \frac{\mu(\mu+1)}{1-(g+b-1)\mu}, \\
1 - \underline{\delta}_{k-g,g+b} &= 1 - (k-g-1)\mu - \frac{\mu^2(k-g)(g+b)}{1-(g+b-1)\mu} \\
&= \frac{1-(k+b-2)\mu - (k+b-1)\mu^2}{1-(g+b-1)\mu} \\
&= \frac{(\mu+1)(1-(k+b-1)\mu)}{1-(g+b-1)\mu}.
\end{aligned} \tag{27}$$

Therefore, the ratio in the right-hand side of (23) can be rewritten as

$$\frac{\bar{\delta}_{2,g+b} + \underline{\delta}_{2,g+b}}{2(1 - \underline{\delta}_{k-g,g+b})} = \frac{\mu}{1-(k+b-1)\mu}. \tag{28}$$

According to (27), $\mu < 1/(k+b-1) \leq 1/(g+b-1)$ implies that $1 - \underline{\delta}_{k-g,g+b} > 0$. Lemma 3 combined with (28) implies that (21) is met. ■

VI. Sufficiency of (1) for $\mathrm{OLS}_{\mathcal{Q}}$

We now prove the sufficient condition for $\mathrm{OLS}_{\mathcal{Q}}$ stated in Theorem 3. The result is a consequence of Proposition 2 and Lemma 5 stated below. We first need to introduce the coherence of the *normalized* projected dictionary $\tilde{\mathbf{B}}^{\mathcal{R}}$:

Definition 5 (Coherence of the normalized projected dictionary)

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

The following proposition gives a sufficient condition on μ_{g+b}^{OLS} under which (7) is satisfied:

Proposition 2 *Let* $\mathcal{Q}^\star$ *and* $\mathcal{Q}$ *be such that* $|\mathcal{Q}^\star| = k$, $|\mathcal{Q}^\star \cap \mathcal{Q}| = g$ *and* $|\bar{\mathcal{Q}}^\star \cap \mathcal{Q}| = b$. *Assume that* $\mathbf{A}_{\mathcal{Q}^\star \cup \mathcal{Q}}$ *is full rank. If* $\mu_{g+b}^{OLS} < 1/(2k-2g-1)$, *then*

$$\max_{i \notin \mathcal{Q}^\star} \|\tilde{\mathbf{B}}_{\mathcal{Q}^\star \backslash \mathcal{Q}}^{\dagger}\tilde{\mathbf{b}}_i\|_1 < 1. \tag{29}$$

Proof: When $\tilde{\mathbf{b}}_i = \mathbf{0}$, the result is obvious. When $\tilde{\mathbf{b}}_i \neq \mathbf{0}$, apply [23, Cor. 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 $|\mathcal{Q}^\star| = k$, $\max_{i \notin \mathcal{Q}^\star} \|\mathbf{A}_{\mathcal{Q}^\star}^{\dagger}\mathbf{a}_i\|_1 < 1$) to the matrix $\tilde{\mathbf{B}}$ and to $\mathcal{Q}^\star \backslash \mathcal{Q}$ of size $k-g$. 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 $\mathbf{A}_{\mathcal{Q}^\star \cup \mathcal{Q}}$ is full rank [20, Cor. 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 5 *If* $\mu < 1/l$, *then*

$$\mu_l^{OLS} \leq \frac{\mu}{1-l\mu}. \tag{30}$$

The proof of this result is reported to Appendix B. The sufficient condition stated in Theorem 3 for $\mathrm{OLS}_{\mathcal{Q}}$ then follows from the combination of Proposition 2 and Lemma 5. Indeed, (1) implies $\mu < 1/(k+b-1) \leq 1/(g+b)$ since $2k-g+b-1 = k+b-1+(k-g) > k+b-1 \geq g+b$. Hence, the result follows by first applying Lemma 5 and (1):

$$\mu_{g+b}^{OLS} \leq \frac{\mu}{1-(g+b)\mu} < \frac{1}{2k-2g-1},$$

and then Proposition 2, which implies (29). $\mu < 1/(k+b-1)$ implies that the full rank assumption of Proposition 2 is met for any $\mathcal{Q}^\star \cup \mathcal{Q}$ of cardinality $k+b$ [23, Lemma 2.3].

VII. Ordering properties and sufficiency of (1) for the success of $(P_{p,\mathcal{Q}})$

In this section, we elaborate on the proofs of Theorems 5 (direct part), 6 and 7. These results have been gathered in this section since they are all related to some guarantees of success for $(P_{p,\mathcal{Q}})$: Theorem 5 shows that (1) is a sufficient and worst-case necessary condition for the success of $(P_{p,\mathcal{Q}})$; Theorem 6 establishes an ordering property between the truncated NSPs for different values of $p \in [0,1]$; Theorem 7 emphasizes that the ERC-OMP (9) is also a sufficient condition for the success of $(P_{1,\mathcal{Q}})$ and in turn, of $(P_{p,\mathcal{Q}})$ for $p < 1$.

Theorems 6 and 7 follow from some technical lemmas which are stated below and proved in Appendix C. The proof of the direct part of Theorem 5 is a consequence of Theorems 6, 7 and is discussed at the end of this section. The proof of the converse part of Theorem 5 is reported to the next section.

We first turn our attention to the proof of the NSP ordering stated in Theorem 6. The result follows from the following lemma:

Lemma 6 *Assume* $\text{spark}(\mathbf{A}) > k+b$ *and let* $\forall \mathbf{v} \in \ker_0(\mathbf{A})$*:*

$$\theta_p(k,g,b,\mathbf{v}) \triangleq \max_{|\mathcal{Q}^\star|=k} \max_{\substack{|\bar{\mathcal{Q}}^\star\cap\mathcal{Q}|=b \\ |\mathcal{Q}^\star\cap\mathcal{Q}|=g}} \left\{ \frac{\|\mathbf{v}_{\mathcal{Q}^\star\setminus\mathcal{Q}}\|_p^p}{\|\mathbf{v}_{\overline{\mathcal{Q}\cup\mathcal{Q}^\star}}\|_p^p} \right\}. \tag{31}$$

Then, the following inequality holds for $0 \leq q < p \leq 1$*:*

$$\theta_q(k,g,b,\mathbf{v}) \leq \theta_p(k,g,b,\mathbf{v}). \tag{32}$$

Obviously, taking the supremum with respect to $\mathbf{v} \in \ker_0(\mathbf{A})$ of both sides in (32) leads to the result stated in Theorem 6.

Secondly, the inequality relating $\theta_1(k,g,b)$ to $\theta_{\text{OMP}}(k,g,b)$ in Theorem 7 is a consequence of the next result:

Lemma 7 *If* $\text{spark}(\mathbf{A}) > k+b$*, then*

$$\frac{\|\mathbf{v}_{\mathcal{Q}^\star\setminus\mathcal{Q}}\|_1}{\|\mathbf{v}_{\overline{\mathcal{Q}^\star\cup\mathcal{Q}}}\|_1} \leq \max_{i\notin\mathcal{Q}^\star} \|\tilde{\mathbf{A}}^\dagger_{\mathcal{Q}^\star\setminus\mathcal{Q}}\tilde{\mathbf{a}}_i\|_1 \tag{33}$$

for any $\mathbf{v} \in \ker_0(\mathbf{A})$ *and* $\mathcal{Q}^\star$*,* $\mathcal{Q}$ *with* $|\mathcal{Q}^\star| = k$*,* $|\mathcal{Q}^\star\cap\mathcal{Q}| = g < k$*,* $|\bar{\mathcal{Q}}^\star\cap\mathcal{Q}| = b$*.*

Theorem 7 then follows by taking the supremum of both sides of (33) with respect to $\mathbf{v} \in \ker_0(\mathbf{A})$ and $\mathcal{Q}^\star$, $\mathcal{Q}$ with $|\mathcal{Q}^\star| = k$, $|\mathcal{Q}^\star\cap\mathcal{Q}| = g$ and $|\bar{\mathcal{Q}}^\star\cap\mathcal{Q}| = b$.

We are now ready to prove the sufficiency of (1) for $(P_{p,\mathcal{Q}})$:

Proof: (Direct part of Theorem 5) On the one hand, let us first note that (1) is sufficient for the success of $\text{OMP}_\mathcal{Q}$ for any k-sparse representation $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$ (Theorem 3). Hence, (1) implies that $\theta_{\text{OMP}}(k,g,b) < 1$ since the latter condition is worst-necessary for the success of $\text{OMP}_\mathcal{Q}$ (Theorem 1). On the other hand, from [23, Th. 2.4], we have that (1) is sufficient for $\text{spark}(\mathbf{A}) > 2k-g+b > k+b$. Applying successively Theorems 7 and 6, we have

$$\theta_p(k,g,b) < 1 \qquad \forall p \in [0,1].$$

The result then follows from Theorem 4. ∎

VIII. Worst-case necessity of (1)

A. General case $b \geq 0$

Cai&Wang recently showed in [38, Th. 3.1] that there exist dictionaries $\mathbf{A}$ with $\mu = \frac{1}{2k-1}$ and a vector $\mathbf{y} \in \text{span}(\mathbf{A})$ having *two* disjoint 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 show that (1) is worst-case necessary for $(P_{p,\mathcal{Q}})$ and $\text{Oxx}_\mathcal{Q}$ in the sense defined in Theorems 3 and 5, respectively. To prove the result for $(P_{p,\mathcal{Q}})$, we will construct a dictionary $\mathbf{A}$ satisfying $\mu = \frac{1}{2k-g+b-1}$ and such that

$$\theta_p(k,g,b) = 1 \qquad \forall p \in [0,1]. \tag{34}$$

The result then immediately follows from Theorem 4. Invoking Theorem 7 and the converse part of Theorem 1, (34) also leads to the result for $\text{OMP}_\mathcal{Q}$: in particular, $\theta_{\text{OMP}}(k,g,b) \geq 1$. On the other hand, since $\theta_{\text{OLS}}(k,g,b)$ and $\theta_1(k,g,b)$ do not enjoy a nesting property similar to (17), specific arguments need to be derived to prove the worst-case necessity of (1) for $\text{OLS}_\mathcal{Q}$. Regarding $\text{OLS}_\mathcal{Q}$ (and actually, also $\text{OMP}_\mathcal{Q}$), we will show using the same dictionary as for $(P_{p,\mathcal{Q}})$, that there exists a k-term representation $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$ satisfying the hypotheses of Theorem 3 and such that $\text{Oxx}_\mathcal{Q}$ selects a wrong atom at the first iteration. The proofs for $\text{Oxx}_\mathcal{Q}$ and $(P_{p,\mathcal{Q}})$ use a dictionary construction similar to Cai&Wang's in [38].

Let $\mathbf{G} \in \mathbb{R}^{(2k-g+b)\times(2k-g+b)}$ be the matrix with ones on the diagonal and $-\mu \triangleq -\frac{1}{2k-g+b-1}$ elsewhere. $\mathbf{G}$ will play the role of the Gram matrix $\mathbf{G} = \mathbf{A}^T\mathbf{A}$. We will exploit the eigenvalue decomposition of $\mathbf{G}$ to construct the dictionary $\mathbf{A} \in \mathbb{R}^{(2k-g+b-1)\times(2k-g+b)}$ with the desired properties. Since $\mathbf{G}$ is symmetric, it can be expressed as

$$\mathbf{G} = \mathbf{U}\Lambda\mathbf{U}^T,$$

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

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

$$\mathbf{A} = \Upsilon\mathbf{U}^T, \tag{35}$$

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

$$\Upsilon(i,j) = \begin{cases} \sqrt{1+\mu} & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\Upsilon^T\Upsilon = \Lambda$. Hence, $\mathbf{A}$ satisfies the statement (i) in the converse part of Theorems 3 and 5 since

$$\mathbf{A}^T\mathbf{A} = \mathbf{U}\Upsilon^T\Upsilon\mathbf{U}^T = \mathbf{U}\Lambda\mathbf{U}^T = \mathbf{G},$$

and therefore

$$\langle \mathbf{a}_i, \mathbf{a}_j \rangle = -\mu \quad \forall i \neq j. \tag{36}$$

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

Taking these observations into account, it easily follows that (34) holds since

$$\|\mathbf{v}_{\mathcal{Q}^\star\setminus\mathcal{Q}}\|_p^p = \|\mathbf{v}_{\overline{\mathcal{Q}^\star\cup\mathcal{Q}}}\|_p^p = k-g,$$

for $\mathbf{v} = \mathbf{1}_{2k-g+b} \in \ker_0(\mathbf{A})$, and $\forall\,\mathcal{Q}^\star$,$\mathcal{Q}$ with $|\mathcal{Q}^\star| = k$, $|\mathcal{Q}^\star\cap\mathcal{Q}| = g$ and $|\bar{\mathcal{Q}}^\star\cap\mathcal{Q}| = b$. This proves the necessary part of Theorem 5.

We now address the case of OLS. Although the OMP necessity result is already obtained from the $(P_{p,\mathcal{Q}})$ necessity result, the construction related to OLS is also valid for OMP. For the sake of generality, we develop our arguments for both OMP and OLS hereafter. We first need the following technical lemma whose proof is reported to Appendix D:

Lemma 8 *Let* $\mathbf{A}$ *be defined as in* (35)*. Then, for any subset* $\mathcal{Q}$ *with* $|\mathcal{Q}| = g+b$*, there exists a vector* $\tilde{\mathbf{y}}$ *having two* $(k-g)$*-term representations with disjoint supports in the projected dictionary* $\tilde{\mathbf{C}}_{\setminus\mathcal{Q}} \triangleq \tilde{\mathbf{C}}_{\{1,\ldots,2k-g+b\}\setminus\mathcal{Q}} \in \mathbb{R}^{(2k-g+b-1)\times(2k-2g)}$*.*

We are now ready to prove the worst-case necessity of (1) for $\text{Oxx}_{\mathcal{Q}}$:

Proof: (Converse part of Theorem 3) We show that there exists a k-sparse representation $\mathbf{y}$ such that $\text{Oxx}_{\mathcal{Q}}$ selects a wrong at the first iteration with the dictionary $\mathbf{A}$ defined in (35). Our construction of such $\mathbf{y}$ is as follows. Let $\mathcal{Q}$ be a subset of cardinality $g+b$, arbitrarily chosen (say, the first $g+b$ atoms of the dictionary). We consider the following decomposition $\mathcal{Q} = \mathcal{Q}_g \cup \mathcal{Q}_b$ where $\mathcal{Q}_g$ and $\mathcal{Q}_b$ are the subsets collecting respectively the good and the bad atoms in $\mathcal{Q}$, with $\mathcal{Q}_g \cap \mathcal{Q}_b = \emptyset$. Let $\tilde{\mathbf{y}}_2$ be a vector having two $(k-g)$-term representations in the projected dictionary $\tilde{\mathbf{C}}_{\setminus\mathcal{Q}}$. We note that such a vector $\tilde{\mathbf{y}}_2$ exists by virtue of Lemma 8. We will denote the respective supports of the two representations of $\tilde{\mathbf{y}}_2$ by $\mathcal{Q}_1$ and $\mathcal{Q}_2$ with $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \emptyset$. Hence,

$$\tilde{\mathbf{y}}_2 = \tilde{\mathbf{C}}_{\mathcal{Q}_1}\mathbf{x}_{\mathcal{Q}_1} = \tilde{\mathbf{C}}_{\mathcal{Q}_2}\mathbf{x}_{\mathcal{Q}_2},$$

for some vectors $\mathbf{x}_{\mathcal{Q}_1}$ and $\mathbf{x}_{\mathcal{Q}_2}$. We then define the k-sparse representation

$$\mathbf{y} \triangleq \mathbf{y}_1 + \mathbf{y}_2,$$

where $\mathbf{y}_1 = \mathbf{A}_{\mathcal{Q}_g}\mathbf{1}_{|\mathcal{Q}_g|}$ and $\mathbf{y}_2 = \mathbf{A}_{\mathcal{Q}_i}\mathbf{x}_{\mathcal{Q}_i} \in \text{span}(\mathbf{A}_{\mathcal{Q}_i})$ with $i=1$ or $i=2$. The specific value of i will be determined hereafter so that a failure situation occurs.
The selection rule (4) indicates that the atom $\tilde{\mathbf{a}}_j$ selected by $\text{Oxx}_{\mathcal{Q}}$ at the first iteration satisfies:

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

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

$$\mathcal{Q}^\star = \begin{cases} \mathcal{Q}_g \cup \mathcal{Q}_1 & \text{if } j \in \mathcal{Q}_2, \\ \mathcal{Q}_g \cup \mathcal{Q}_2 & \text{if } j \in \mathcal{Q}_1. \end{cases} \tag{37}$$

To complete the proof, it is easy to check that $\mathbf{y} = \mathbf{y}_1 + \mathbf{y}_2 \in \text{span}(\mathbf{A}_{\mathcal{Q}^\star})$: indeed, $\mathbf{y}_1 \in \text{span}(\mathbf{A}_{\mathcal{Q}_g}) \subset \text{span}(\mathbf{A}_{\mathcal{Q}^\star})$ and $\mathbf{y}_2 \in \text{span}(\mathbf{A}_{\mathcal{Q}^\star\setminus\mathcal{Q}}) \subset \text{span}(\mathbf{A}_{\mathcal{Q}^\star})$. ■

B. Special case $b = 0$

We now turn our attention to the proof of Lemma 2, which is related to the standard version of Oxx, initialized with the empty support. We first need to define the concept of "reachability" of a subset $\mathcal{Q}$:

Definition 6 *A subset $\mathcal{Q}$ is said to be reachable by Oxx if there exists $\mathbf{y} \in \text{span}(\mathbf{A}_{\mathcal{Q}})$ 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 [20]. The authors showed that any subset $\mathcal{Q}$ with $|\mathcal{Q}| \leq \text{spark}(\mathbf{A}) - 2$ is reachable by OLS, see [20, 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 [20, Example 1]. This scenario does however not occur for the dictionary defined in (35) as stated in the next lemma:

Lemma 9 *Let $\mathbf{A}$ be defined as in* (35) *with $b = 0$. Then any subset $\mathcal{Q}$ with $|\mathcal{Q}| = g < k$ is reachable by Oxx.*

The proof of this result is reported to Appendix D. We are now ready to prove Lemma 2:

Proof: (Lemma 2) Consider the dictionary $\mathbf{A}$ defined in (35) with $b = 0$. Let $\mathcal{Q}$ be a subset of cardinality g, arbitrarily chosen (say, the first g atoms of the dictionary). We will exhibit a subset $\mathcal{Q}^\star \supset \mathcal{Q}$ for which the result of Lemma 2 holds.

We first apply Lemma 9: there exists an input $\mathbf{y}_1 \in \text{span}(\mathbf{A}_{\mathcal{Q}})$ for which Oxx selects all atoms in $\mathcal{Q}$ during the first g iterations. Then, we apply Lemma 8: there exists a vector $\tilde{\mathbf{y}}_2$ having two $(k-g)$-term representations in the projected dictionary $\tilde{\mathbf{C}}_{\setminus\mathcal{Q}}$. We will denote their respective supports by $\mathcal{Q}_1$ and $\mathcal{Q}_2$ with $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \emptyset$. We then define $\mathbf{y}_2$ as in the proof of the converse of Theorem 3.

By virtue of [20, Lemma 15], Oxx with $\mathbf{y} = \mathbf{y}_1 + \varepsilon\mathbf{y}_2$ as input selects the same atoms (*i.e.*, $\mathcal{Q}$) as with $\mathbf{y}_1$ as input during the first g iterations as long as $\varepsilon > 0$ is sufficiently small. Moreover, defining $\mathcal{Q}^\star$ as in (37) and applying the same reasoning as in the proof of the converse part of Theorem 3, we have that $\mathbf{y} \in \text{span}(\mathbf{A}_{\mathcal{Q}^\star})$ and is such that Oxx selects a bad atom at iteration $g+1$. ■

IX. Quasi-tightness of (11) for $\text{OMP}_{\mathcal{Q}}$

In this section, we provide an instance of dictionary such that $\delta_{k+b+1} = 1/\sqrt{k-g}$ and $\text{Oxx}_{\mathcal{Q}}$ fails at the first iteration. Our dictionary construction is along the same lines as [35, Th. 3.2].

Proof: (Lemma 1) We first consider the case $g \leq k-2$. Let us define $\mathbf{A} \in \mathbb{R}^{(k+b+1)\times(k+b+1)}$ as

$$\mathbf{A} = \begin{pmatrix} \mathbf{I}_{g+b} & \mathbf{0}_{(g+b)\times(k-g+1)} \\ \mathbf{0}_{(k-g+1)\times(g+b)} & \mathbf{M} \end{pmatrix} \tag{38}$$

where

$$\mathbf{M} \triangleq \begin{pmatrix} & & & \frac{1}{k-g} \\ & \mathbf{I}_{k-g} & & \vdots \\ & & & \frac{1}{k-g} \\ 0 & \dots & 0 & \sqrt{\frac{k-g-1}{k-g}} \end{pmatrix}$$

On the one hand, it can be seen that the eigenvalues of the Gram matrix $\mathbf{G} = \mathbf{A}^T\mathbf{A}$ are $\lambda = 1$ with multiplicity $k+b-1$ and $\lambda = 1 \pm \frac{1}{\sqrt{k-g}}$ with multiplicity 1. Hence, $\delta_{k+b+1} = \frac{1}{\sqrt{k-g}}$.

On the other hand, there exist $\mathcal{Q}^\star$ and $\mathcal{Q}$ satisfying the hypotheses of Lemma 1 and such that $\text{Oxx}_{\mathcal{Q}}$ fails at the first iteration for some representation $\mathbf{y} = \mathbf{A}\mathbf{x}^\star$ indexed by $\mathcal{Q}^\star$. Let us set $\mathcal{Q} = \{1,\dots,g+b\}$, $\mathcal{Q}^\star = \{b+1,\dots,k+b\}$ in such a way that there is only one wrong atom outside of $\mathcal{Q} \cup \mathcal{Q}^\star$, namely the last atom. We set

$$x_i^\star = \begin{cases} 1 & \text{if } i \in \mathcal{Q}^\star \\ 0 & \text{otherwise.} \end{cases}$$

With this particular choice, we have $\mathbf{y} = \mathbf{A}_{\mathcal{Q}^\star}\mathbf{1}_k$ and:

$$\tilde{\mathbf{C}}_{\{1,\dots,k+b+1\}\setminus\mathcal{Q}} = \mathbf{A}_{\{1,\dots,k+b+1\}\setminus\mathcal{Q}} = \begin{pmatrix} \mathbf{0}_{g+b\times k-g+1} \\ \mathbf{M} \end{pmatrix},$$

$$\mathbf{r}^{\mathcal{Q}} = \tilde{\mathbf{C}}_{\mathcal{Q}^\star\setminus\mathcal{Q}}\mathbf{1}_{k-g} = \begin{pmatrix} \mathbf{0}_{g+b} \\ \mathbf{1}_{k-g} \\ 0 \end{pmatrix},$$

and therefore,

$$\langle \tilde{\mathbf{c}}_i, \mathbf{r}^{\mathcal{Q}}\rangle = 1 \quad \forall i \geq g+b+1.$$

Since $k+b+1 \notin \mathcal{Q}^\star$, a failure situation as in (6) occurs.

The special case $g = k-1$ leads to the degenerate situation $\delta_{k+b+1} = 1$ in Lemma 1. This case is handled by proposing a dictionary having two identical columns. We define $\mathbf{A}$ as in (38) with

$$\mathbf{M} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{2\times 2}.$$

We have obviously $\delta_{k+b+1} = 1$ since that the dictionary has two identical columns. $\mathrm{Oxx}_{\mathcal{Q}}$ then trivially fails with $\mathbf{y}$, $\mathcal{Q}^\star$ and $\mathcal{Q}$ defined as above.

■

X. Conclusions

We derived a new sufficient and worst-case necessary condition, $\mu < \frac{1}{2k-g+b-1}$, for the success of OMP, OLS and some procedures based on ℓ_p relaxation. Our result both applies to the context of sparse representations with support side information, and to the analysis of greedy algorithms at intermediate iterations. Our condition relaxes the well-known coherence-based result $\mu < \frac{1}{2k-1}$ derived in the non-informed setup by several authors, see *e.g.*, [21], [23], [39]. Moreover, it is shown to be complementary with some similar conditions based on restricted isometry constants [16], [19].

We also carried out a fine analysis of some relations existing between conditions of success for OMP/OLS and ℓ_p-relaxed procedures in the informed setup. We showed that the truncated NSP, characterizing the success of ℓ_p-relaxed procedures in the informed setup, enjoys some ordering property. Moreover, we established a direct implication between the ERC-OMP derived in [20] and the truncated NSP for the informed ℓ_1-relaxed problem.

Appendix A
Proof of the results of section V

This section contains the proofs of Lemmas 3 and 4 together with some useful technical lemmas.

Lemma 10 *Assume* $\mathbf{A}$ *satisfies the P-RIP(*$\underline{\delta}_{2,l}$*,*$\bar{\delta}_{2,l}$*) and let*

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

Then, we have

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

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

$$1 + \bar{\delta}_{2,l} \geq \lambda_{max}(\tilde{\mathbf{A}}_{\mathcal{S}}^T \tilde{\mathbf{A}}_{\mathcal{S}}), \quad (39)$$

$$1 - \underline{\delta}_{2,l} \leq \lambda_{min}(\tilde{\mathbf{A}}_{\mathcal{S}}^T \tilde{\mathbf{A}}_{\mathcal{S}}), \quad (40)$$

where $\lambda_{max}(\mathbf{M})$ (resp. $\lambda_{min}(\mathbf{M})$) denotes the largest (resp. smallest) eigenvalue of $\mathbf{M}$ and we used the short-hand notation $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}^{\mathcal{R}}$. Moreover, if $\mathcal{S} = \{i,j\}$, it is easy to check that the eigenvalues of $\tilde{\mathbf{A}}_{\mathcal{S}}^T \tilde{\mathbf{A}}_{\mathcal{S}}$ can be expressed as

$$\lambda(\tilde{\mathbf{A}}_{\mathcal{S}}^T \tilde{\mathbf{A}}_{\mathcal{S}}) = \frac{\|\tilde{\mathbf{a}}_i\|^2 + \|\tilde{\mathbf{a}}_j\|^2 \pm \Delta}{2},$$

where

$$\begin{aligned}\Delta &= \sqrt{(\|\tilde{\mathbf{a}}_i\|^2 + \|\tilde{\mathbf{a}}_j\|^2)^2 + 4(\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j\rangle^2 - \|\tilde{\mathbf{a}}_i\|^2\,\|\tilde{\mathbf{a}}_j\|^2)} \\ &= \sqrt{(\|\tilde{\mathbf{a}}_i\|^2 - \|\tilde{\mathbf{a}}_j\|^2)^2 + 4\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j\rangle^2}.\end{aligned}$$

Hence

$$\lambda_{max}(\tilde{\mathbf{A}}_{\mathcal{S}}^T \tilde{\mathbf{A}}_{\mathcal{S}}) - \lambda_{min}(\tilde{\mathbf{A}}_{\mathcal{S}}^T \tilde{\mathbf{A}}_{\mathcal{S}}) = \Delta \geq 2|\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j\rangle|.$$

Using (39)-(40), we thus obtain $\forall i,j \notin \mathcal{R}$:

$$\bar{\delta}_{2,l} + \underline{\delta}_{2,l} \geq 2|\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j\rangle|. \quad (41)$$

Now, this inequality also holds if $i \in \mathcal{R}$ or $j \in \mathcal{R}$ since the right hand-side of (41) is then equal to zero. The result then follows from the definition of μ_l^{OMP}. ■

Lemma 11 *Let* $|\mathcal{R}|=l$ *and* $\mathcal{S}\cap\mathcal{S}' = \emptyset$*, then* $\forall \mathbf{u} \in \mathbb{R}^{|\mathcal{S}'|}$*,*

$$\|(\tilde{\mathbf{A}}_{\mathcal{S}}^{\mathcal{R}})^T \tilde{\mathbf{A}}_{\mathcal{S}'}^{\mathcal{R}} \mathbf{u}\| \leq \mu_l^{OMP} \sqrt{|\mathcal{S}||\mathcal{S}'|}\,\|\mathbf{u}\|.$$

Proof: Let us use the short-hand notation $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}^{\mathcal{R}}$. We have:

$$\begin{aligned}\|\tilde{\mathbf{A}}_{\mathcal{S}}^T \tilde{\mathbf{A}}_{\mathcal{S}'}\mathbf{u}\| &= \sqrt{\sum_{i\in\mathcal{S}} \langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{A}}_{\mathcal{S}'}\mathbf{u}\rangle^2} \\ &= \sqrt{\sum_{i\in\mathcal{S}} \big(\sum_{j\in\mathcal{S}'} u_j\,\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j\rangle\big)^2} \\ &\leq \sqrt{\sum_{i\in\mathcal{S}} \big(\sum_{j\in\mathcal{S}'} |u_j|\,|\langle \tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_j\rangle|\big)^2} \\ &\leq \mu_l^{OMP}\sqrt{|\mathcal{S}|}\,\|\mathbf{u}\|_1 \\ &\leq \mu_l^{OMP}\sqrt{|\mathcal{S}||\mathcal{S}'|}\,\|\mathbf{u}\|.\end{aligned}$$

■

Using Lemmas 10 and 11, we can now prove Lemmas 3 and 4:

Proof: (Lemma 3) $\forall\, i \notin \mathcal{Q}^\star$, the following inequalities hold:

$$\begin{aligned}\|\tilde{\mathbf{A}}_{\mathcal{Q}^\star\setminus\mathcal{Q}}^\dagger \tilde{\mathbf{a}}_i\|_1 &\leq \sqrt{k-g}\,\|\tilde{\mathbf{A}}_{\mathcal{Q}^\star\setminus\mathcal{Q}}^\dagger \tilde{\mathbf{a}}_i\|, \\ &\leq \frac{\sqrt{k-g}}{1-\underline{\delta}_{k-g,g+b}}\|\tilde{\mathbf{A}}_{\mathcal{Q}^\star\setminus\mathcal{Q}}^T \tilde{\mathbf{a}}_i\|, \\ &\leq \frac{k-g}{1-\underline{\delta}_{k-g,g+b}}\mu_{g+b}^{OMP}, \\ &\leq \frac{k-g}{1-\underline{\delta}_{k-g,g+b}}\frac{\bar{\delta}_{2,g+b}+\underline{\delta}_{2,g+b}}{2},\end{aligned}$$

where the first inequality follows from the equivalence of norms; the second from RIC properties (see [46, Prop. 3.1]); the third from Lemma 11 and the fourth from Lemma 10. ■

Proof: (Lemma 4) First, notice that $\mathbf{A}$ satisfies the P-RIP($\underline{\delta}_{q,0}$,$\bar{\delta}_{q,0}$) $\forall\, q$ with

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

see *e.g.*, [23, Lemma 2.3]. Let $\mathcal{R},\mathcal{S}$ with $|\mathcal{R}| = l$, $|\mathcal{S}| = q$, $\mathcal{R}\cap\mathcal{S} = \emptyset$. Then, (24) is a consequence of the following inequalities:

$$\|\mathbf{P}_{\mathcal{R}}^\perp \mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}}\|^2 \leq \|\mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}}\|^2 \leq (1+\bar{\delta}_{q,0})\|\mathbf{x}_{\mathcal{S}}\|^2.$$

Lower bound (25) is derived by noticing that

$$\|\mathbf{P}_{\mathcal{R}}^\perp \mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}}\|^2 = \|\mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}}\|^2 - \|\mathbf{P}_{\mathcal{R}}\mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}}\|^2,$$

and

$$\begin{aligned}\|\mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}}\|^2 &\geq (1-\underline{\delta}_{q,0})\|\mathbf{x}_{\mathcal{S}}\|^2, \\ \|\mathbf{P}_{\mathcal{R}}\mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}}\|^2 &= \|(\mathbf{A}_{\mathcal{R}}^\dagger)^T\mathbf{A}_{\mathcal{R}}^T\mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}}\|^2 \\ &\leq \frac{\|\mathbf{A}_{\mathcal{R}}^T\mathbf{A}_{\mathcal{S}}\mathbf{x}_{\mathcal{S}}\|^2}{1-\underline{\delta}_{l,0}}, \quad (42) \\ &\leq \frac{\mu^2 lq\,\|\mathbf{x}_{\mathcal{S}}\|^2}{1-\underline{\delta}_{l,0}}, \quad (43)\end{aligned}$$

where inequality (42) 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 $\mu < 1/(l-1)$ [23, Lemma 2.3]; (43) is a consequence of Lemma 11. ■

Appendix B
Proof of the results of section VI

Proof: (Lemma 5) The proof is recursive. Obviously, the result holds for $l = 0$ since $\mu_0^{OLS} = \mu$.

Let $\mathcal{R}$ with $|\mathcal{R}| = l \geq 1$ and assume that the result holds for $l-1$. If $j \in \mathcal{R}$, the bound $|\langle \tilde{\mathbf{b}}_j^{\mathcal{R}}, \tilde{\mathbf{b}}_{j'}^{\mathcal{R}} \rangle| \leq \frac{\mu}{1-l\mu}$ trivially holds since $\tilde{\mathbf{b}}_j^{\mathcal{R}} = \mathbf{0}_m$.

Let us then consider the case where $j \notin \mathcal{R}$ and $j' \notin \mathcal{R}$. First, from the assumption $\mu < 1/l$, we have that $\mathbf{A}_{\mathcal{R}\cup\{j\}}$ is full column rank as a family of $l+1$ atoms [23, Lemma 2.3]. Let us then consider $\mathcal{S}$ such that $\mathcal{R} = \mathcal{S} \cup \{i\}$ with $|\mathcal{S}| = l-1$ and let us apply [20, Lemma 5]. We obtain the following orthogonal decomposition:

$$\tilde{\mathbf{b}}_j^{\mathcal{S}} = \eta_j \tilde{\mathbf{b}}_j^{\mathcal{R}} + \langle \tilde{\mathbf{b}}_j^{\mathcal{S}}, \tilde{\mathbf{b}}_i^{\mathcal{S}} \rangle \tilde{\mathbf{b}}_i^{\mathcal{S}},$$

with

$$\eta_j \triangleq \pm\sqrt{1 - \langle \tilde{\mathbf{b}}_j^{\mathcal{S}}, \tilde{\mathbf{b}}_i^{\mathcal{S}} \rangle^2} \neq 0.$$

Exploiting this decomposition, we can thus write:

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

Taking the absolute value of both sides and majorizing the inner products on the right-hand side by μ_{l-1}^{OLS}, we obtain:

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

where the last inequality follows from the fact that (30) is assumed to hold for $l-1$. This proves the result for $|\mathcal{R}| = l$ and completes the recursion. ■

Appendix C
Proof of the results of section VII

Before proceeding to the proofs of Lemmas 6 and 7, we emphasize that $\text{spark}(\mathbf{A}) > k+b$ and $\mathbf{v} \in \ker_0(\mathbf{A})$ are sufficient conditions for $\mathbf{v}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}}$ not to be equal to zero because $\mathcal{Q}^\star \cup \mathcal{Q}$ is composed of $k+b$ elements. This implies that (13), (31) and (33) are always well-defined, as their denominators are nonzero.

Proof: (Lemma 6) As an initial remark, let us mention that, for any $\mathbf{v} \in \ker_0(\mathbf{A})$, a couple $(\mathcal{Q}^\star, \mathcal{Q})$ maximizing the right-hand side of (31) should be such that $\mathbf{v}_{\mathcal{Q}^\star \setminus \mathcal{Q}}$ (resp. $\mathbf{v}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}}$) collects the elements of $\mathbf{v}$ with the largest (resp. smallest) amplitudes, because $t \mapsto t^p$ is an increasing function on $\mathbb{R}^+$. In the rest of the proof, we will therefore assume that $\mathcal{Q}^\star$ and $\mathcal{Q}$ satisfy this requirement.

Let $\mathbf{w}^T \triangleq [\mathbf{v}_{\mathcal{Q}^\star \setminus \mathcal{Q}}^T, \mathbf{v}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}}^T]$. Taking our initial remark into account, $\theta_p(k,g,b,\mathbf{v})$ can be expressed as

$$\theta_p(k,g,b,\mathbf{v}) = \frac{\|\mathbf{v}_{\mathcal{Q}^\star \setminus \mathcal{Q}}\|_p^p}{\|\mathbf{v}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}}\|_p^p} = \frac{\sum_{i=1}^{k-g} |w_i|^p}{\|\mathbf{w}\|_p^p - \sum_{i=1}^{k-g} |w_i|^p}.$$

Showing (32) is therefore equivalent to proving that

$$\frac{\sum_{i=1}^{k-g} |w_i|^q}{\|\mathbf{w}\|_q^q - \sum_{i=1}^{k-g} |w_i|^q} \leq \frac{\sum_{i=1}^{k-g} |w_i|^p}{\|\mathbf{w}\|_p^p - \sum_{i=1}^{k-g} |w_i|^p}.$$

which can also be rewritten as

$$\frac{\sum_{i=1}^{k-g} |w_i|^q}{\|\mathbf{w}\|_q^q} \leq \frac{\sum_{i=1}^{k-g} |w_i|^p}{\|\mathbf{w}\|_p^p} \qquad \text{for } q < p. \tag{44}$$

Now, in [24, Th. 5], it is proved that (44) holds for any vector $\mathbf{w}$ whose $k-g$ first elements have the largest magnitudes. Observing that $\mathbf{w}$ satisfies the latter condition, we obtain the result. ■

Proof: (Lemma 7) For any $\mathbf{v} \in \ker_0(\mathbf{A})$, we have

$$\mathbf{A}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \mathbf{v}_{\mathcal{Q}^\star \setminus \mathcal{Q}} = -\mathbf{A}_{\mathcal{Q}} \mathbf{v}_{\mathcal{Q}} - \mathbf{A}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}} \mathbf{v}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}}.$$

Applying the orthogonal projector onto $\text{span}(\mathbf{A}_{\mathcal{Q}})^\perp$ to both sides, we obtain

$$\tilde{\mathbf{A}}_{\mathcal{Q}^\star \setminus \mathcal{Q}} \mathbf{v}_{\mathcal{Q}^\star \setminus \mathcal{Q}} = -\tilde{\mathbf{A}}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}} \mathbf{v}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}}.$$

Let us note that $\mathbf{A}_{\mathcal{Q}^\star \cup \mathcal{Q}}$ is full-rank by hypothesis and, by virtue of [20, Cor. 3], $\tilde{\mathbf{A}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}$ is therefore also a full-rank matrix. This leads to

$$\mathbf{v}_{\mathcal{Q}^\star \setminus \mathcal{Q}} = -\tilde{\mathbf{A}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}^\dagger \tilde{\mathbf{A}}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}} \mathbf{v}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}}.$$

Taking the ℓ_1 norm of both sides and using the definition of the ℓ_1 matrix induced norm, we have

$$\frac{\|\mathbf{v}_{\mathcal{Q}^\star \setminus \mathcal{Q}}\|_1}{\|\mathbf{v}_{\overline{\mathcal{Q}^\star \cup \mathcal{Q}}}\|_1} \leq \max_{i \notin \mathcal{Q}^\star \cup \mathcal{Q}} \|\tilde{\mathbf{A}}_{\mathcal{Q}^\star \setminus \mathcal{Q}}^\dagger \tilde{\mathbf{a}}_i\|_1. \tag{45}$$

The result then follows from the fact that $\tilde{\mathbf{a}}_i = \mathbf{0}_m \ \forall i \in \mathcal{Q}$. ■

Appendix D
Proof of the results of section VIII

In this appendix, we provide a proof of Lemmas 8 and 9. We first need to prove the following technical lemma:

Lemma 12 *Let* $\mathbf{A}$ *be defined as in* (35). *Then, we have for all* $\mathcal{R}$ *with* $|\mathcal{R}| < 2k-g+b$ *and* $i, j \notin \mathcal{R}$, $i \neq j$:

$$\langle \tilde{\mathbf{a}}_i^{\mathcal{R}}, \tilde{\mathbf{a}}_j^{\mathcal{R}} \rangle = -\mu - \mu^2 \mathbf{1}_{|\mathcal{R}|}^T (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{1}_{|\mathcal{R}|}, \tag{46}$$

$$\|\tilde{\mathbf{a}}_i^{\mathcal{R}}\|^2 = 1 - \mu^2 \mathbf{1}_{|\mathcal{R}|}^T (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{1}_{|\mathcal{R}|}. \tag{47}$$

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

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

Using this expression, we have

$$\begin{aligned} \langle \tilde{\mathbf{a}}_i^{\mathcal{R}}, \tilde{\mathbf{a}}_j^{\mathcal{R}} \rangle &= \langle \mathbf{a}_i, \mathbf{a}_j \rangle - \mathbf{a}_i^T \mathbf{A}_{\mathcal{R}} (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{A}_{\mathcal{R}}^T \mathbf{a}_j, \\ \|\tilde{\mathbf{a}}_i^{\mathcal{R}}\|^2 &= 1 - \mathbf{a}_i^T \mathbf{A}_{\mathcal{R}} (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{A}_{\mathcal{R}}^T \mathbf{a}_i. \end{aligned}$$

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

Proof: (Lemma 8) Using Lemma 12 for $|\mathcal{R}| = g+b$, we notice that $\tilde{\mathbf{C}}_{\setminus \mathcal{Q}}^{\mathcal{Q}} = \beta \tilde{\mathbf{A}}_{\setminus \mathcal{Q}}^{\mathcal{Q}}$ for some $\beta > 0$ since $\|\tilde{\mathbf{a}}_i^{\mathcal{Q}}\|$ does not depend on i. Moreover, $\tilde{\mathbf{a}}_i^{\mathcal{Q}} \neq \mathbf{0}_m$ (and therefore $\tilde{\mathbf{c}}_i^{\mathcal{Q}} \neq \mathbf{0}_m$) since $\text{spark}(\mathbf{A}) = 2k-g+b > g+b+1$, which implies that $\mathbf{A}_{\mathcal{Q}\cup\{i\}}$ is full-rank and, in turn, that $\tilde{\mathbf{a}}_i^{\mathcal{Q}} \neq \mathbf{0}_m$. Defining $\mathbf{v} \triangleq \mathbf{1}_{2k-2g}$, we obtain

$$\begin{aligned} \tilde{\mathbf{C}}_{\setminus \mathcal{Q}} \mathbf{v} &= \beta \tilde{\mathbf{A}}_{\setminus \mathcal{Q}} \mathbf{v} \\ &= \beta \tilde{\mathbf{A}} \mathbf{1}_{2k-g+b} = \beta \mathbf{P}_{\mathcal{Q}}^\perp \mathbf{A} \mathbf{1}_{2k-g+b} = \mathbf{0}_{2k-g+b-1} \end{aligned} \tag{48}$$

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

Let us partition the elements of $\mathbf{v} = \mathbf{1}_{2k-2g}$ into two subsets $\mathcal{Q}_1 \cup \mathcal{Q}_2$ with $\mathcal{Q}_1 \cap \mathcal{Q}_2 = \emptyset$ and $|\mathcal{Q}_1| = |\mathcal{Q}_2| = k-g$, and define $\tilde{\mathbf{y}} \triangleq \tilde{\mathbf{C}}_{\mathcal{Q}_1 \setminus \mathcal{Q}} \mathbf{1}_{k-g}$. According to (48), $\tilde{\mathbf{y}}$ rereads $-\tilde{\mathbf{C}}_{\mathcal{Q}_2 \setminus \mathcal{Q}} \mathbf{1}_{k-g}$, therefore $\tilde{\mathbf{y}}$ has two $(k-g)$-sparse representations with disjoint supports in $\tilde{\mathbf{C}}_{\setminus \mathcal{Q}}$. ■

Proof: (*Lemma 9*) Let us first recall that b is set to 0 in this lemma. We prove a result slightly more general than the statement of Lemma 9: for the dictionary defined as in (35), any subset $\mathcal{R}$ with $p \triangleq |\mathcal{R}| \leq 2k-g-2$ can be reached by Oxx. Lemma 9 corresponds to the case $p = g$ ($p \leq 2k - g - 2$ is always satisfied as long as $g < k$).

The result is true for OLS by virtue of [20, 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 $\mathrm{spark}(\mathbf{A}) = 2k - g$ and $|\mathcal{R}| \leq 2k - g - 2$ 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 (36)). For arbitrary values of $\varepsilon_2, \ldots, \varepsilon_p > 0$, we define the following recursive construction:

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

($\mathbf{y}_{p+1}$ implicitly depends on $\varepsilon_2, \ldots, \varepsilon_{p+1}$). We show by recursion that for all $p \in \{1, \ldots, 2k - g - 2\}$, there exist $\varepsilon_2, \ldots, \varepsilon_p > 0$ such that OMP with the dictionary defined as in (35) 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 (4) always yields a unique maximum.

The statement is obviously true for $\mathbf{y}_1 = \mathbf{a}_1$. Assume that it is true for $\mathbf{y}_p$ ($p < 2k - g - 2$) with some $\varepsilon_2, \ldots, \varepsilon_p > 0$ (these parameters will remain fixed in the following). According to [20, Lemma 15], there exists $\varepsilon_{p+1} > 0$ such that OMP with $\mathbf{y}_{p+1} = \mathbf{y}_p + \varepsilon_{p+1} \mathbf{a}_{p+1}$ 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

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

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

$$|\langle \tilde{\mathbf{a}}_i^{\mathcal{R}}, \tilde{\mathbf{a}}_{p+1}^{\mathcal{R}} \rangle| < \|\tilde{\mathbf{a}}_{p+1}^{\mathcal{R}}\|^2 \qquad \forall\, i \neq p + 1. \tag{49}$$

Now, $|\mathcal{R}| = p < 2k - g$ by hypothesis, then Lemma 12 applies (we remind the reader that we assume that $b = 0$). Using (46)-(47), it is easy to see that (49) is equivalent to

$$\mu + 2\mu^2 \mathbf{1}_p^T (\mathbf{A}_{\mathcal{R}}^T \mathbf{A}_{\mathcal{R}})^{-1} \mathbf{1}_p < 1. \tag{50}$$

Since $\mu = \frac{1}{2k-g-1} < \frac{1}{p+1} < \frac{1}{p-1}$, we have $1 - (p-1)\mu > 0$. Then, [23, Lemma 2.3] and $\|\mathbf{1}_p\|^2 = p$ yield:

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

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

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

which proves that the condition (50), and then (49) is met. OMP therefore recovers the subset $\mathcal{R} \cup \{p + 1\} = \{1, \ldots, p + 1\}$.

■

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] M. Elad, *Sparse and redundant representations: From theory to applications in signal and image processing*, Springer, Aug. 2010.

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

[9] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of Computational Mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[10] S. Foucart, "Stability and robustness of weak orthogonal matching pursuits," in *Recent advances in harmonic analysis and applications*, D. Bilyk, L. De Carli, A. Petukhov, A. M. Stokolos, and B. D. Wick, Eds. 2013, vol. 25, pp. 395–405, Springer proceedings in Mathematics & Statistics.

[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. Inf. Theory*, vol. 58, no. 2, pp. 1094–1121, Feb. 2012.

[12] R. F. Von Borries, C. J. Miosso, and C. Potes, "Compressed sensing using prior information," in *2nd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMPSAP 2007*. Dec. 2007, pp. 121–124, IEEE.

[13] M. Amin Khajehnejad, W. Xu, A. Salman Avestimehr, and B. Hassibi, "Weighted ℓ_1 minimization for sparse recovery with prior information," in *IEEE International Symposium on Information Theory (ISIT) 2009*. June 2009, pp. 483–487, IEEE.

[14] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.

[15] L. Jacques, "A short note on compressed sensing with partially known signal support," *Signal Processing*, vol. 90, no. 12, pp. 3308–3312, Dec. 2010.

[16] M. P. Friedlander, H. Mansour, R. Saab, and O. Yilmaz, "Recovering compressively sampled signals using partial support information," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1122–1134, Feb. 2012.

[17] A. S. Bandeira, K. Scheinberg, and L. N. Vicente, "On partially sparse recovery," Tech. Rep., Dept. Mathematics, Univ. Coimbra, 2011.

[18] Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," *SIAM J. Img. Sci.*, vol. 3, no. 3, pp. 462–491, Aug. 2010.

[19] N. B. Karahanoglu and H. Erdogan, "On the theoretical analysis of orthogonal matching pursuit with termination based on the residue," Oct. 2012.

[20] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, "Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3158–3174, May 2013.

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

[22] X. Chen and W. Zhou, "Convergence of reweighted ℓ_1 minimization algorithms and unique solution of truncated ℓ_p minimization," Tech. Rep., The Hong Kong Polytechnic University, Apr. 2010.

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

[24] R. Gribonval and M. Nielsen, "Highly sparse representations from dictionaries are unique and independent of the sparseness measure," *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335–355, May 2007.

[25] 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.

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

[27] 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, 2010.

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

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

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

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

[32] 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.

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

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

[35] Q. Mo and Y. Shen, "A remark on the restricted isometry property in orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, 2012.

[36] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Trans. Signal Processing*, vol. 60, no. 9, pp. 4973–4976, 2012.

[37] Wei Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, May 2009.

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

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

[40] M. Wang, W. Xu, and A. K. Tang, "On the performance of sparse recovery via ℓ_p-minimization," *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7255–7278, Nov. 2011.

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

[42] E. Candes, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sciences, Serie I*, vol. 346, pp. 589–592, 2008.

[43] T. T. Cai, L. Wang, and G. Xu, "Shifting inequality and recovery of sparse signals," *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp. 1300–1308, Mar. 2010.

[44] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

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

[46] 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.