



HAL
open science

A Signal Dependent Analysis of Orthogonal Least Squares Based on the Restricted Isometry Property

Jonathan Bosse

► **To cite this version:**

Jonathan Bosse. A Signal Dependent Analysis of Orthogonal Least Squares Based on the Restricted Isometry Property. IEEE Transactions on Signal Processing, 2023, 71, pp.1574-1585. <10.1109/TSP.2023.3267009>. <hal-04147370>

HAL Id: hal-04147370

<https://hal.science/hal-04147370v1>

Submitted on 30 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A Signal dependent analysis of Orthogonal Least Squares based on the Restricted Isometry Property

Jonathan Bosse

DEMR, ONERA, Université Paris-Saclay F-91123 Palaiseau - France

Abstract—The Orthogonal Least Squares (OLS) is a widespread greedy algorithm for sparse signal approximation or reconstruction. We here address the problem of sufficient condition for exact signal support recovery via OLS. Based on the generalized Wielandt inequality, new bounds on the OLS selection rule are provided that enable a simple approach of the Restricted Isometry Property based analysis of OLS. Since these new bounds naturally take the signal magnitude distribution into account, they allow a derivation of a new relaxed exact support recovery condition in the noise-free case. Moreover, we study their influence on the design of such new conditions in bounded noise and Gaussian noise cases.

Index Terms—Compressed sensing, greedy algorithms, orthogonal least squares (OLS), Restricted Isometry Property (RIP), sparse reconstruction

I. INTRODUCTION

THE recovery of signals that admit a sparse representation has been raising a lot of interest, especially since the emergence of compressed sensing (CS) [1], [2]. The goal of sparse recovery is, given a measured $\mathbf{y} \in \mathbb{C}^m$, to recover the K -sparse signal vector (that contains at most K nonzero entries) $\mathbf{x} \in \mathbb{C}^N$ from \mathbf{y} (usually $m < N$) such that

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (1)$$

where $\mathbf{A} \in \mathbb{C}^{m \times N}$ is called the dictionary and its columns are often called atoms. The observation \mathbf{y} may also be corrupted by additive noise.

Among the popular methods, popularity probably due to the fact that they are very intuitive and exhibit good performances, greedy algorithms such as Matching Pursuit (MP) [3], Orthogonal Matching Pursuit (OMP) [4] or Orthogonal Least Squares (OLS) [5] have been extensively used and studied. MP is often outperformed by both OMP and OLS. It has been observed in [6] that OLS usually performs better than OMP at the price of an additional computational cost.

Performance studies in sparse recovery are often conducted through the coherence (the maximal absolute value of the scalar product between dictionary columns) or exact recovery conditions [4], [6] in a noise free case. Among the first OMP analysis in presence of noise, one can cite [7]. The coherence is a simple, convenient tool but often leads to severe bounds that limit the analysis to small sparsity levels [8]. Although it is often harder to evaluate, the Restricted Isometry Property (RIP) [2] is a very popular mean to conduct a finer analysis.

A matrix is said to satisfy the RIP of order K if, for every K -sparse vector \mathbf{x} , we have

$$(1 - \delta_K) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_K) \|\mathbf{x}\|_2^2.$$

The smallest such constant $\delta_K \in [0, 1[$ is called the Restricted Isometry Constant (RIC).

The first RIP based performance study of OMP is due to [9]. It establishes conditions on the RIC for which OMP succeeds in exact support recovery. Since then, these results have been improved, so that the recovery conditions became tighter and tighter [10-15] both in noise free and noisy cases. Conditions for support recovery in presence of noise usually consists in requirements on the minimal signal magnitude, assuming that conditions on the RIC for noiseless support recovery hold true [12-15].

Improving the result [16], an optimal sufficient condition on the RIC of order $K + 1$ for exact support recovery of OLS was recently proposed in [17], [18]

$$\delta_{K+1} < C_K = \begin{cases} \frac{1}{\sqrt{K+\frac{1}{4}}} & K = 2 \\ \frac{1}{\sqrt{K+\frac{1}{16}}} & K = 3 \\ \frac{1}{\sqrt{K}} & K = 1, K \geq 4 \end{cases}. \quad (2)$$

This condition is optimal in the sense that if it is violated, there exist dictionaries and signals such that the support recovery via OLS fails. One can remark that the counterexamples given in [17] for which recovery fails, all deal with signals of equal amplitudes. It should be emphasized that these approaches do not assume any prior knowledge on the signal and can be seen as a worst case analysis. As such, they may be overly pessimistic. It is indeed widely observed that greedy algorithms often perform very well even if those sufficient conditions are violated.

It has been acknowledged that “flat” signals that possess the same magnitude are usually seen as a worst case [19]. But in many applications (e.g. Radar, Sonar), the signals have no reason to be of equal amplitude: once they are reordered (what can always be done without loss of generality), they often obey some decay. Thus, relaxed bounds that exploit the signal magnitude decay are helpful to understand why and how in practice greedy algorithms might still perform well even if worst case (flat signals) sufficient conditions are violated. In [20] it has been proved that OMP indeed performs very well when the signal is decaying. In [19], decay of the signal’s coefficients has been exploited to relax usual condition on the coherence for support recovery for both OMP and OLS. The decay of the signal’s coefficients has also been exploited to relax the condition on the RIC for exact support recovery on OMP [9], [15], [21], block-OMP [22] or orthogonal multi-matching pursuit [23]. But to the best of our knowledge,

such study for RIP based analysis of OLS has not yet been addressed by the literature.

In this paper, we provide new bounds on the OLS selection rule that are useful for a RIP based analysis of OLS. These bounds are based on the generalized Wielandt inequality [24], [25] and enable a quite simple analysis of OLS. The virtue of the proposed bounds is to naturally take the signal coefficient's distribution into account, contrary to existing recovery conditions in the literature [16-18]. We show that, as soon as the signal's coefficients obey some sufficient decay, the proposed bounds can improve the state of the art. It enables the derivation of a new relaxed condition for OLS support recovery in the noiseless scenario as well as improved sufficient condition on signal's magnitude for support recovery in the presence of noise. Unlike the state of the art, the condition on δ_{K+1} for noise free signal recovery depends both on δ_K and on the signal amplitude's distribution and becomes naturally sharper than existing ones as soon as the signal coefficients obey a sufficient decay. Moreover, the proposed recovery conditions allow to give sufficient condition on the signal magnitude decay to ensure perfect recovery given RIC of order $K + 1$ and K .

Note that in [13], an upper bound on the OMP selection rule for atoms that are not in the signal support is proposed, based on Wielandt inequality. As we will show, Wielandt's bounds are tighter for OLS and thus, suit the OLS analysis better. Moreover, the generalized Wielandt inequality is needed to provide a lower bound on the OLS selection rule for atoms that are in the signal support, in order to derive the proposed analysis.

The paper is organized as follows. In section II, we present the basis of the OLS algorithm. Then, in section III, the generalized Wielandt inequality is introduced, from which we deduce new bounds for the OLS selection rule. These bounds are compared to the state of the art. Then, in section IV, we provide, thanks to these bounds, a new RIP based analysis of OLS in the noise free case. Finally, section V deals with sufficient signal's condition for OLS support recovery in both bounded and Gaussian noise scenarios.

Notations : The following notations will be used in the paper. $\langle \cdot, \cdot \rangle$ refers to the classic inner product between complex vectors. Vectors are denoted by lowercases whereas matrices are denoted by capital letters. \mathbf{I} denotes the identity matrix. \mathbf{a}_i denotes the i -th column of \mathbf{A} , and x_i denotes the i -th coefficient of \mathbf{x} . \mathbf{A}_Λ denotes the matrix restricted to the columns that belong to the set Λ . \mathbf{x}_Λ denotes the vector whose index are restricted to the set Λ . \mathbf{P}_Λ denotes the orthogonal projection on $\text{span}(\mathbf{A}_\Lambda)$ and $\mathbf{P}_\Lambda^\perp = \mathbf{I} - \mathbf{P}_\Lambda$ its orthogonal complement. $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the classic ℓ_1 and ℓ_2 vector norms on \mathbb{C}^N . $S \subset \Omega = \{1, 2, \dots, N\}$ will be denoted as the true support of the signal to be recovered \mathbf{x} . $|S|$ denotes here the cardinal of the set S . $\Lambda \setminus S$ denotes the set Λ where the subset S is excluded. Without loss of generality (since the columns of the dictionary \mathbf{A} can always be reordered accordingly), we assume that the nonzero coefficients of \mathbf{x} are decreasing : $|x_{S(1)}| \geq |x_{S(2)}| \geq \dots \geq |x_{S(K)}|$ where $S(k)$ denotes the k -th element of the set S . We also assume, without loss of generality (since the signal \mathbf{x} can always be

re-scaled accordingly), that the columns of \mathbf{A} are of unit norm. $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ correspond to the minimal and maximal eigenvalue of a matrix, respectively. \bar{E} denotes the complement of any event E .

II. THE OLS ALGORITHM

The OLS algorithm is summarized in Algorithm 1. This iterative algorithm aims at finding a sparse approximation in a certain dictionary of a given discrete signal \mathbf{y} . For that, it selects relevant atoms according to a selection rule and then gets rid of their contribution thanks to an orthogonal projection (cleaning). If the stopping rule is not satisfied, then the algorithm continues. This basic principle is common to other greedy approaches : MP and OMP. The difference between them lies in the way the projection is done and in the choice of the selection rule. These iterative algorithms select at each step an atom that best describes the signal residual, according to a dedicated selection rule, and then remove an estimated component of this atom to update the signal residual.

The stopping rule usually depends on the context (e.g. noise free, Gaussian noise, bounded noise) and will consequently be precised each time it is necessary in the sequel.

At the k -th step, given an estimated support S_{k-1} , the OLS selects an atom \mathbf{a}_{i_k} that minimizes the norm of the residual $\mathbf{r}_{k-1 \cup i} = \mathbf{P}_{S_{k-1} \cup i}^\perp \mathbf{y}$,

$$i_k = \arg \min_{i \notin S_{k-1}} \|\mathbf{r}_{k-1 \cup i}\|_2,$$

which can be rewritten, after some calculation [26], as

$$i_k = \arg \max_{i \notin S_{k-1}} \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1} \rangle|}{\left\| \mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i \right\|_2}. \quad (3)$$

As it has been observed earlier, the OLS selection rule (3) can be interpreted as a cosine of an angle [6], [27]. Indeed, as $\|\mathbf{r}_{k-1}\|_2$ does not depend on i , one can always divide (3) by this quantity. Since $\mathbf{P}_{S_{k-1}}^\perp$ is an orthogonal projector, it is idempotent and Hermitian, so that $|\langle \mathbf{a}_i, \mathbf{r}_{k-1} \rangle| = \left| \langle \mathbf{a}_i, \mathbf{P}_{S_{k-1}}^\perp \mathbf{r}_{k-1} \rangle \right| = \left| \langle \mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i, \mathbf{r}_{k-1} \rangle \right|$. Then, we can write

$$i_k = \arg \max_{i \notin S_{k-1}} \cos \Theta \left(\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i, \mathbf{r}_{k-1} \right), \quad (4)$$

where for any pair of complex vectors (\mathbf{x}, \mathbf{y}) we define as in [25]

$$\cos \Theta(\mathbf{x}, \mathbf{y}) = \frac{|\langle \mathbf{x}, \mathbf{y} \rangle|}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}.$$

The angle $\Theta(\mathbf{x}, \mathbf{y})$ defined this way is then obviously such that $0 \leq \Theta(\mathbf{x}, \mathbf{y}) \leq \pi/2$.

Consequently, the OLS selection rule amounts to comparing the cosine of the angle between each projected atom $\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i$ and the residual \mathbf{r}_{k-1} and selecting the least angle. This angular interpretation of the OLS selection rule is at the basis of the present paper, where generalized Wielandt's bounds between pairs of complex vectors allowed us to derive new relevant bounds for the OLS selection rule.

Data: \mathbf{y}, \mathbf{A}
Result: $\hat{S} = S_k, \hat{\mathbf{x}} = \mathbf{x}_k$
 $\mathbf{r}_0 \leftarrow \mathbf{y}, \mathbf{x}_0 \leftarrow \mathbf{0}, S_0 \leftarrow \emptyset, k \leftarrow 0;$
while *Stopping rule is not met* **do**
 $k \leftarrow k + 1$
 $i_k \leftarrow \arg \max_{i \notin S_{k-1}} \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1} \rangle|}{\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i\|_2}$
 $S_k \leftarrow i_k \cup S_{k-1}$
 $\mathbf{P}_{S_k} \leftarrow \mathbf{A}_{S_k} (\mathbf{A}_{S_k}^H \mathbf{A}_{S_k})^{-1} \mathbf{A}_{S_k}^H$
 $\mathbf{r}_k \leftarrow \mathbf{y} - \mathbf{P}_{S_k} \mathbf{y} = \mathbf{P}_{S_k}^\perp \mathbf{y}$
 $\mathbf{x}_k \leftarrow (\mathbf{A}_{S_k}^H \mathbf{A}_{S_k})^{-1} \mathbf{A}_{S_k}^H \mathbf{y}$
end

Algorithm 1: the OLS algorithm

III. NEW BOUNDS ON THE NOISELESS OLS SELECTION RULES

Greedy algorithms studies are often conducted comparing lower bounds on the corresponding noiseless selection rule for atoms belonging to the support S to upper bounds on the noiseless selection rule for atoms that are not in the support. As a consequence, bounds on the noiseless OLS selection rule are crucial, as they are at the basis of the algorithm analysis both in noiseless and noisy cases. In this section we propose such new bounds based on the generalized Wielandt inequality that will be at the basis of the new recovery conditions derived in section IV and V.

We now state the generalized Wielandt inequality that quantifies how $\cos \Theta(\mathbf{x}, \mathbf{y})$ is affected, when an invertible matrix \mathbf{B} is applied both on \mathbf{x} and \mathbf{y} . Whereas Wielandt inequality [28] deals with orthogonal signals, the generalized Wielandt inequality deals with signals of arbitrary angle. Initially introduced in [24], the study [25, Theorem 3.2] later gave equality conditions as well as a symmetric formulation of the bound that appears to be very useful for our study. This is why we retain the formulation of [25]:

Theorem 1. [25, Theorem 3.2] *For any invertible matrix \mathbf{B} , if we define $\cos \phi = \cos \Theta(\mathbf{x}, \mathbf{y})$ and $\cos \psi = \cos \Theta(\mathbf{B}\mathbf{x}, \mathbf{B}\mathbf{y})$, then*

$$\frac{-\chi + \cos \phi}{1 - \chi \cos \phi} \leq \cos \psi \leq \frac{\chi + \cos \phi}{1 + \chi \cos \phi} \quad (5)$$

with

$$\chi = \frac{\kappa(\mathbf{B})^2 - 1}{\kappa(\mathbf{B})^2 + 1},$$

where $\kappa(\mathbf{B})$ is the condition number of the matrix \mathbf{B} . If \mathbf{B} is Hermitian and positive definite $\kappa(\mathbf{B}) = \frac{\lambda_{\max}(\mathbf{B})}{\lambda_{\min}(\mathbf{B})}$.

Let α_1 and α_2 be the solutions to $|\langle \mathbf{x}, \mathbf{y} \rangle| = \alpha_1 \langle \mathbf{x}, \mathbf{y} \rangle$ and $|\langle \mathbf{B}\mathbf{x}, \mathbf{B}\mathbf{y} \rangle| = \alpha_2 \langle \mathbf{B}\mathbf{x}, \mathbf{B}\mathbf{y} \rangle$. Equality in the right-hand side is met if and only if $\frac{\alpha_1 \mathbf{x}}{\|\alpha_1 \mathbf{x}\|_2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2}$ is in the eigensubspace of $\mathbf{B}^H \mathbf{B}$ corresponding to $\lambda_{\max}(\mathbf{B}^H \mathbf{B})$ and $\frac{\alpha_1 \mathbf{x}}{\|\alpha_1 \mathbf{x}\|_2} + \frac{\mathbf{y}}{\|\mathbf{y}\|_2}$ is in the eigensubspace corresponding to $\lambda_{\min}(\mathbf{B}^H \mathbf{B})$ and either $\alpha_1 = \alpha_2$ or $\langle \mathbf{B}\mathbf{x}, \mathbf{B}\mathbf{y} \rangle = 0$. Equality is met in the left-hand side if and only if $\frac{\alpha_2 \mathbf{x}}{\|\alpha_2 \mathbf{x}\|_2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2}$ is in the eigensubspace corresponding to $\lambda_{\max}(\mathbf{B}^H \mathbf{B})$ and $\frac{\alpha_2 \mathbf{x}}{\|\alpha_2 \mathbf{x}\|_2} + \frac{\mathbf{y}}{\|\mathbf{y}\|_2}$ is in the

eigensubspace corresponding to $\lambda_{\min}(\mathbf{B}^H \mathbf{B})$ and either $\alpha_1 = \alpha_2$ or $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

As we will see, the generalized Wielandt inequality allows us to link the OLS selection rule to the eigenvalues of a certain matrix at any iteration k . These eigenvalues can then be linked to the RIC. Thus, one can deduce the following new bounds on the noiseless OLS selection rule:

Lemma 1. *For any given set $S_k \subset S$, we have the following upper and lower bounds in absence of noise*

$$\max_{i \notin S} \cos \Theta(\mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k) \leq \delta_{K+1} \quad (6)$$

$$\max_{i \in S \setminus S_k} \cos \Theta(\mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k) \geq \frac{\cos \phi_k - \delta_K}{1 - \delta_K \cos \phi_k}, \quad (7)$$

where

$$\cos \phi_k = \max_{i \in S \setminus S_k} \frac{|x_i|}{\|\mathbf{x}_{S \setminus S_k}\|_2}. \quad (8)$$

Proof. See Appendix A. \square

Of course, these bounds can immediately be rewritten in terms of the OLS selection rule

$$\max_{i \notin S} \frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2} \leq \delta_{K+1} \|\mathbf{r}_k\|_2,$$

and

$$\max_{i \in S \setminus S_k} \frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2} \geq \frac{\cos \phi_k - \delta_K}{1 - \delta_K \cos \phi_k} \|\mathbf{r}_k\|_2.$$

A Wielandt's inequality based upper bound has been established in [13] for atoms that do not belong to the support using Wielandt's inequality for the analysis of OMP. The main difference between OMP and OLS lies in the absence of the denominator in the selection rule (3). Since $\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2 \leq 1$,

$$\max_{i \notin S} |\langle \mathbf{a}_i, \mathbf{r}_k \rangle| \leq \max_{i \notin S} \frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2},$$

so that one can see that the bound (6) is tighter for the OLS selection rule. As we can see, this bound is well suited to the analysis of OLS. To our best knowledge, (7) is new and only applies to OLS.

If we have a look at the conditions in Theorem 1 that have to be fulfilled for the proposed bounds to be achieved, one can see that they are not easily verified for any dictionary \mathbf{A} and signal \mathbf{x} . Nevertheless, the proposed bounds naturally take the signal's amplitude distribution into account through the term $\cos \phi_k$, what can lead to great improvements, as we will see.

Note that unlike the state of the art bounds for OLS selection rule that depend on $\|\mathbf{r}_k\|_2^2$, Wielandt's inequality based bounds depend on $\|\mathbf{r}_k\|_2$.

The study [17], [18] exploits different bounds to prove (2). The interested reader may consult Appendix B that provides a comparison of the bounds (6) and (7) to the state of the art. It is shown that the bound (6) is the tightest to date, whereas (7) is looser than the bound exploited in [17], [18] when the signal exhibit constant amplitudes and can be tighter when the signal components are decaying.

In the sequel, two cases are studied where the proposed bounds are useful to obtain improved results when the signal's

coefficient exhibit some decay. First, one can see that the signal decay allows some relaxation on the condition on δ_{K+1} . Then, the proposed bounds are used in a noisy context to obtain improved recovery conditions.

IV. EXACT SUPPORT RECOVERY IN THE NOISE FREE CASE

One of the consequences of the proposed bounds consists in relaxed OLS recovery conditions. This is achieved when the signal's coefficients are decaying through the parameter $\cos \phi_k$ in (7). As we will see, the way the signal's coefficient are distributed directly influences the values of this parameter.

We have the following result

Theorem 2. *If a K -sparse signal \mathbf{x} of support S follows the model (1), then OLS with stopping rule $\|\mathbf{r}_k\|_2 = 0$, recovers any such \mathbf{x} in K iterations provided*

$$\delta_{K+1} < \frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi}, \quad (9)$$

where

$$\begin{aligned} \cos \Phi &= \min_{\Lambda} \left\{ \frac{\max_{i \in S \setminus \Lambda} |x_i|}{\|\mathbf{x}_{S \setminus \Lambda}\|_2}, \Lambda \subset S \right\}, \\ &= \min_{1 \leq k \leq K} \frac{|x_{S(k)}|}{\sqrt{\sum_{i=k}^K |x_{S(i)}|^2}}. \end{aligned} \quad (10)$$

Proof. This is a consequence of Lemma 1. For every step k , given S_{k-1} composed of correct atoms, OLS selects a correct atom when

$$\max_{i \notin S} \cos \Theta \left(\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i, \mathbf{r}_{k-1} \right) < \max_{i \in S \setminus S_{k-1}} \cos \Theta \left(\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i, \mathbf{r}_{k-1} \right)$$

That is to say, using (6) and (7), when

$$\delta_{K+1} < \frac{\cos \phi_{k-1} - \delta_K}{1 - \delta_K \cos \phi_{k-1}}, \quad (11)$$

where $\cos \phi_{k-1}$ is defined in (8).

The righthand-side of (11) is monotonically increasing with respect to $\cos \phi_{k-1}$ (see Lemma 7 in Appendix A). As a consequence, the condition (9) implies (11) for every $k \in \{1, \dots, K\}$, so that OLS selects a correct atom at each step k .

As soon as $k \leq K$, using Lemma 6 (in Appendix A) in conjunction with [29, Remark1]

$$\|\mathbf{r}_k\|_2 \geq \sqrt{1 - \delta_K} \|\mathbf{x}_{S \setminus S_k}\|_2,$$

so that $\|\mathbf{r}_k\|_2 > 0$. We have $\|\mathbf{r}_K\|_2 = 0$, so that the algorithm stops at iteration K .

Equation (10) remains to be proved. Let us define $S_k^0 = \{S(1), \dots, S(k)\}$ an ordered subset of the true support S and recall that, without loss of generality, we assumed $|x_{S(1)}| \geq |x_{S(2)}| \geq \dots \geq |x_{S(K)}|$. For any fixed k , let $S_{k-1} \subset S$ be any set of cardinality $k-1$. Then, let $j = \arg \max_{i \in S \setminus S_{k-1}} |x_i|$.

If $j \in S_{k-1}^0$, denoting j_0 such that $j = S(j_0)$, then $\|\mathbf{x}_{S \setminus S_{k-1}}\|_2 \leq \|\mathbf{x}_{S \setminus S_{j_0-1}^0}\|_2$, and

$$\begin{aligned} \max_{i \in S \setminus S_{k-1}} \frac{|x_i|}{\|\mathbf{x}_{S \setminus S_{k-1}}\|_2} &\geq \frac{|x_{S(j_0)}|}{\|\mathbf{x}_{S \setminus S_{j_0-1}^0}\|_2} \\ &\geq \min_{i \in \{1, \dots, K\}} \frac{|x_{S(i)}|}{\|\mathbf{x}_{S \setminus S_{i-1}^0}\|_2}. \end{aligned}$$

If $j \notin S_{k-1}^0$, then $S_{k-1} = S_{k-1}^0$ and by definition

$$\max_{i \in S \setminus S_{k-1}} \frac{|x_i|}{\|\mathbf{x}_{S \setminus S_{k-1}}\|_2} = \frac{|x_{S(k)}|}{\|\mathbf{x}_{S \setminus S_{k-1}^0}\|_2}.$$

Finally, we have for any $S_{k-1} \subset S$, with $1 \leq k \leq K$,

$$\max_{i \in S \setminus S_k} \frac{|x_i|}{\|\mathbf{x}_{S \setminus S_{k-1}}\|_2} \geq \min_{i \in \{1, \dots, K\}} \frac{|x_{S(i)}|}{\|\mathbf{x}_{S \setminus S_{i-1}^0}\|_2},$$

which implies (10). \square

The expression for $\cos \Phi$ in (10) is convenient in order to avoid a combinatorial search.

Interestingly, the worst case (i.e. least value of $\cos \phi_{k-1}$ in (7)) is achieved when OLS selects the support of the strongest modulus of $\mathbf{x}_{S \setminus S_{k-1}}$ at each step k . This suggests that weak greedy strategy for OLS [4] could benefit from not picking necessarily the atom corresponding to the strongest OLS selection rule at each iteration.

Comparing the condition (9) that involves both δ_K and $\cos \Phi$ with the state of the art [17], [18] is not very convenient, because the latter only involve δ_{K+1} and K . However, note that evaluating δ_K is usually less expensive than δ_{K+1} that is already required to verify the common sufficient conditions. Let us examine the following bounds on the condition imposed on δ_{K+1} in (9):

Lemma 2. *The condition (9) becomes*

$$\delta_{K+1} < \frac{1 - \sin \Phi}{\cos \Phi}. \quad (12)$$

in the worst case ($\delta_K = \delta_{K+1}$) and

$$\delta_{K+1} < \cos \Phi, \quad (13)$$

in the best case ($\delta_K = 0$).

Proof. By definition $0 \leq \delta_K \leq \delta_{K+1}$. Since the quantity $\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi}$ is monotonically decreasing with respect to δ_K (see Lemma 7 in Appendix A), the best case corresponds to the case $\delta_K = 0$ where the constraint imposed on δ_{K+1} in (9) is the lightest and the worst case corresponds to $\delta_K = \delta_{K+1}$ where the constraint imposed on δ_{K+1} in (9) is the strongest. The best case is immediate. The worst case corresponds to the lowest δ_{K+1} such that, combining (9) with $\delta_K = \delta_{K+1}$,

$$\delta_{K+1}^2 \cos \Phi - 2\delta_{K+1} + \cos \Phi > 0.$$

Finding the desired root gives the corresponding result (12). \square

The relation (13) can be seen as an upper bound of (9), whereas (12) can be seen as a lower bound. As a consequence,

(12) is a simpler but necessarily looser sufficient recovery condition that does not depend on δ_K .

Lemma 3. *If a K -sparse signal \mathbf{x} of support S follows the model (1), then OLS with stopping rule $\|\mathbf{r}_k\|_2 = 0$, recovers any such \mathbf{x} in K iterations provided*

$$\delta_{K+1} < \frac{1 - \sin \Phi}{\cos \Phi}. \quad (14)$$

Now, let us examine how the signal amplitude distribution impacts $\cos \Phi$. By definition, we have

$$\frac{1}{\sqrt{K}} \leq \cos \Phi \leq 1, \quad (15)$$

where the lower bound corresponds to the case where the signal's components have the same modulus (i.e. "flat" signal) and the upper bound corresponds to the case of strongly decaying signal, where we can make the following approximation

$$\sqrt{\sum_{i=k}^K |x_{S(i)}|^2} \approx |x_{S(k)}|, \text{ so that } \cos \Phi \approx 1.$$

Note that, implicitly, one assumes in (9) that $\cos \Phi \geq \delta_K$ because δ_{K+1} is positive. In view of (12), since $\delta_K \leq \delta_{K+1}$, this condition is met.

A. Flat signals ($\cos \Phi = \frac{1}{\sqrt{K}}$)

The lower bound in (15) is achieved for flat signals. Since the quantity $\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi}$ decreases as $\cos \Phi$ decreases (see Lemma 7 in Appendix A), it proves that the case of flat signals represents a worst case in our analysis: it leads to a stronger constraint imposed on δ_{K+1} in (9). In such case, using Lemma 2 and (15), one sees that for flat signals, the equation (9) becomes

$$\delta_{K+1} < \sqrt{K} - \sqrt{K-1},$$

in the worst case ($\delta_K = \delta_{K+1}$) and

$$\delta_{K+1} < \frac{1}{\sqrt{K}},$$

in the best case ($\delta_K = 0$). This shows, as expected, that the proposed bound (12) cannot improve upon the bound (2) in this context.

B. Signals of uneven amplitudes (decaying signals: $\cos \Phi > \frac{1}{\sqrt{K}}$)

The quantity $\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi}$ increases as $\cos \Phi$ increases, so that the condition imposed on δ_{K+1} in (9) becomes looser and looser. Note that when $\cos \Phi$ tends to one, irrespective of the value of δ_K , the constraint (9) vanishes for δ_{K+1} as it tends to $\delta_{K+1} < 1$. It can also be seen in (14), where in this context $\sin \Phi$ decreases.

The case of increasing $\cos \Phi$ values corresponds to signals exhibiting some decay. At a certain point, the lower bound in (12) becomes superior to C_K , and relaxed conditions with respect to (2) can be envisaged for δ_{K+1} . More precisely, as soon as

$$\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} \geq C_K,$$

that is to say, when

$$\cos \Phi \geq \frac{\delta_K + C_K}{1 + \delta_K C_K}, \quad (16)$$

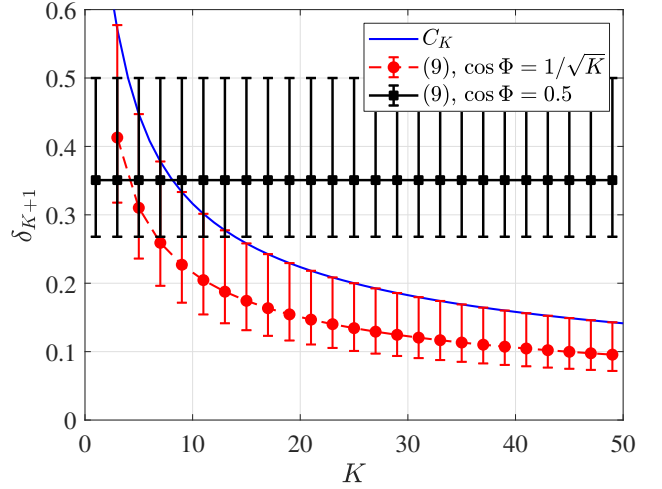


Figure 1. Comparison of the bounds (2) and (9). Upper, lower and center of the bar corresponds to the best ($\delta_K = 0$), worst ($\delta_K = \delta_{K+1}$) and an intermediate case ($\delta_K = \delta_{K+1}/2$), respectively.

the proposed bound (9) improves upon the condition (2).

A sufficient condition that does not depend on δ_K can be obtained using (14):

$$\frac{1 - \sin \Phi}{\cos \Phi} \geq C_K.$$

This is true, after straightforward calculation, when

$$(C_K^2 + 1) \cos^2 \Phi - 2C_K \cos \Phi \geq 0,$$

that implies

$$\cos \Phi \geq \frac{2C_K}{1 + C_K^2} = \frac{2\sqrt{K}}{K + \sqrt{K}}, \quad (17)$$

for $K \geq 4$.

This condition corresponds to the case where δ_K tends to δ_{K+1} . Thus, it stands for a worst case. An upper bound for $\cos \Phi$ can be achieved considering the case $\delta_K = 0$ in (16), giving

$$\cos \Phi \geq C_K, \quad (18)$$

which is always verified.

Figure 1 illustrates the fact that, when the signal has constant magnitude, (9) does not improve on (2) as it only equals this result in the best case ($\delta_K = 0$). When the signal exhibit some decay larger than approximately $\frac{2}{\sqrt{K}}$ (e.g. that does not depend on K), the proposed condition becomes quickly less restrictive than (2).

Whereas state of the art sufficient recovery conditions state what conditions on the dictionary allows a perfect recovery of the signal with OLS, our approach allows also to answer a related question : "given a dictionary and its RIC of order $K+1$ and K , what type of signals can be perfectly reconstructed?". This is a direct consequence of (9):

Lemma 4. *Given a dictionary and its RIC δ_{K+1} and δ_K , every K -sparse signal \mathbf{x} of support S following the model (1),*

is perfectly recovered by OLS with stopping rule $\|\mathbf{r}_k\|_2 = 0$ in K iterations, provided

$$\cos \Phi > \frac{\delta_{K+1} + \delta_K}{1 + \delta_K \delta_{K+1}}, \quad (19)$$

where $\cos \Phi$ is defined in (10).

The parameter $\cos \Phi$ is not always convenient to interpret. This is why it is interesting to make the link with the decay rate of the signals which is easier prone to interpretations.

Let α be the minimal decay rate between the amplitudes of the signal as in [9]:

$$\frac{|x_{S(i)}|}{|x_{S(i+1)}|} > \alpha, \forall i \in S.$$

We have the following result

Lemma 5. *If a K -sparse signal \mathbf{x} of support S follows the model (1) with decay rate α , then OLS with stopping rule $\|\mathbf{r}_k\|_2 > 0$, recovers such \mathbf{x} in K iterations provided*

$$\alpha > \sqrt{1 + \frac{(\delta_K + \delta_{K+1})^2}{1 + \delta_K^2 \delta_{K+1}^2 - \delta_K^2 - \delta_{K+1}^2}}. \quad (20)$$

Proof. We have

$$\begin{aligned} \sqrt{\sum_{i=k}^K |x_{S(i)}|^2} &\leq |x_{S(k)}| \sqrt{\sum_{i=0}^{K-k} \left(\frac{1}{\alpha^2}\right)^i} \\ &= |x_{S(k)}| \sqrt{\frac{1 - \left(\frac{1}{\alpha^2}\right)^{K-k}}{1 - \frac{1}{\alpha^2}}} \\ &\leq |x_{S(k)}| \frac{1}{\sqrt{1 - \frac{1}{\alpha^2}}}, \end{aligned}$$

from which we deduce

$$\cos \Phi \geq \sqrt{1 - \frac{1}{\alpha^2}}.$$

The result follows combining this result with (9). \square

One can see that, contrary to [9], no condition on δ_{K+1} is required other than $\delta_{K+1} < 1$. The larger δ_{K+1} and δ_K are, the larger α should be. As a consequence, when α verifies (20), it is possible to recover the support of \mathbf{x} even if $\delta_{K+1} > C_K$. When δ_{K+1} tends to one, α can take an arbitrary large value, but since δ_K and δ_{K+1} are usually required not to be too large constants compared to one, the required value for α in (20) is usually not large.

Again, the condition (20) depends on δ_K . A sufficient condition that does not depend on δ_{K+1} is obtained as we did previously, when δ_K tends to δ_{K+1} , leading to

$$\alpha > \sqrt{1 + \frac{4\delta_{K+1}^2}{(1 - \delta_{K+1}^2)^2}}. \quad (21)$$

Contrary to the studies [9], [15], where δ_{K+1} needs to satisfy some constraint (respectively $\delta_{K+1} < 1/3$ and $\delta_{K+1} < \sqrt{2}/2$) for support recovery with OMP in presence of strongly decaying signal, the proposed result for OLS does not require

such condition. Potentially, even if δ_{K+1} is close to one, OLS should be able to succeed in exact support recovery, provided the decay rate α is sufficient.

Example. Let us examine a real example from [17]: Let $K = 2$, and let us consider the dictionary $\mathbf{A} \in \mathbb{R}^{3 \times 3}$ such that

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & \frac{\delta}{2} & \frac{\delta}{2} \\ \frac{\delta}{2} & 1 & -\frac{\delta}{2} \\ \frac{\delta}{2} & -\frac{\delta}{2} & 1 \end{bmatrix}. \quad (22)$$

In this case, $\delta_{K+1} = \delta_3 = \delta$. Since $\delta_K = \delta_2$ is equal to the coherence [8] (i.e. the largest non diagonal coefficient modulus of the Gram matrix $\mathbf{A}^T \mathbf{A}$) we have $\delta_2 = \frac{\delta}{2}$. The paper [17] states that the recovery occurs by OLS as soon as

$$\delta_3 < \frac{1}{\sqrt{\frac{1}{4} + 2}}, \quad (23)$$

irrespective of the signal value. The limiting case is the signal $\mathbf{x} = [0 \ 1 \ 1]^T$ for which the reconstruction fails if (23) is no true.

Let us consider the signal $\mathbf{x} = [0 \ 1 \ x]^T$ with $x \in \mathbb{R}$, $x \geq 1$. The limiting value $\bar{\delta}_3$ for which (9) does not hold in this context, is such that

$$\bar{\delta}_3 = \frac{\cos \Phi - \frac{\bar{\delta}_3}{2}}{1 - \frac{\bar{\delta}_3}{2} \cos \Phi},$$

with $\cos \Phi = \frac{x}{\sqrt{1+x^2}}$. After some calculus, one can find

$$\bar{\delta}_3 = \frac{3 - \sqrt{1 + 8 \sin^2 \Phi}}{2 \cos \Phi}.$$

In Figure 2, we compare the value of δ for which OLS fails and call it $\delta_{failure}$. We see that when the signal is flat (i.e. $x = 1$) the condition (23) is less restrictive than our proposed bound $\bar{\delta}_3$. When x grows, we can observe that, empirically, OLS still continues to provide exact support recovery for values of δ that are larger than the bound (23). When x grows, then $\cos \Phi$ is increasing and so is the proposed bound $\bar{\delta}_3$, contrary to (23). As a consequence, the proposed bound becomes more informative when $\cos \Phi$ is not too low. This illustrates the benefit of the condition (9) that naturally takes the signal amplitude into account: when the signal amplitude is decaying, the condition on δ_{K+1} can be relaxed.

C. Case of Gaussian random amplitudes

Let us now assume that the signal is Gaussian distributed: the x_i are now independent complex circular Gaussian random variables: $x_i \sim \mathcal{CN}(0, \alpha^{2i})$, where α is the decay rate of the variances of the x_i 's. This corresponds to a large variety of signal models in the literature (radar, communications ...). Even if we assume that all signal entries have the same variance (i.e. $\alpha = 1$), which is again a worst case, we can see in Figure 3 that the proposed condition is often less restrictive than the state of the art bound, depending on the value of δ_K . This is due to the fact that for each realization of the signal the ordered amplitude always exhibit some small decay that allows our condition to be less severe on average.

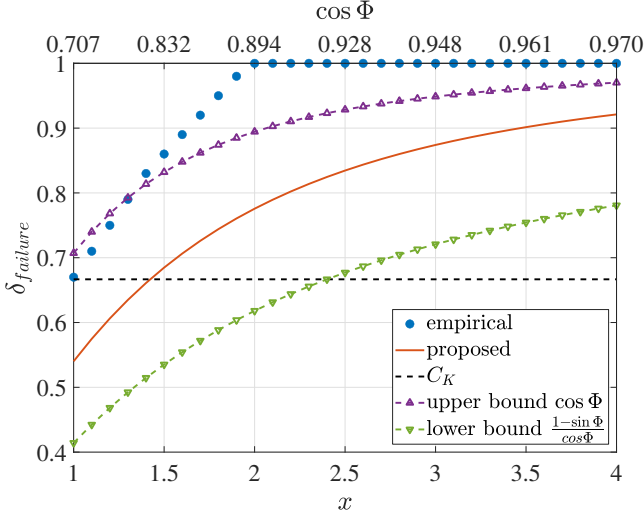


Figure 2. Value $\delta_{failure}$ of δ above which OLS fails to recover the support of the signal $\mathbf{x} = [0 \ 1 \ x]^T$ with the dictionary (22).

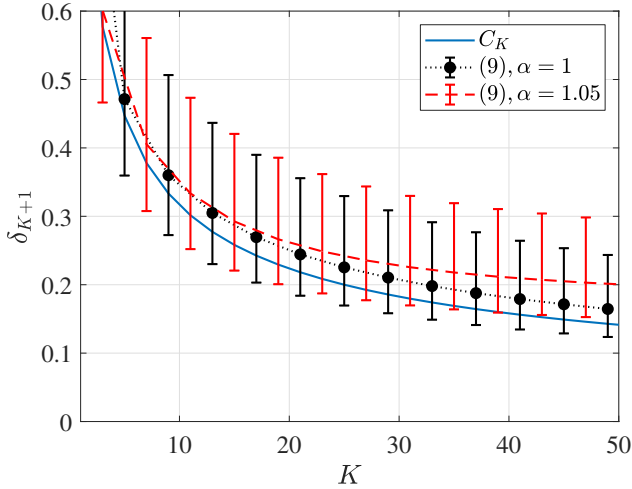


Figure 3. Sufficient recovery conditions on δ_{K+1} are plotted. The case of circular iid Gaussian random variables: $x_i \sim \mathcal{CN}(0, \alpha^{2i})$ is considered and averaged in (9) over 10000 Monte carlo trials. Upper, lower and center of the bar corresponds to the best ($\delta_K = 0$), worst ($\delta_K = \delta_{K+1}$) and intermediate case ($\delta_K = \delta_{K+1}/2$), respectively.

The result strongly depend on the value of δ_K (i.e. on the structure of the dictionary \mathbf{A}): this is why we consider the upper and lower bounds (13) and (12) in Figure 3. When the standards deviation are decaying (here by 5%) then our recovery condition becomes appealing, whatever the value of δ_K . This suggests that in practice, signals are quite likely to obey a sufficient decay that makes the proposed relaxed condition less restrictive than the state of the art (even if the signal coefficients have the same distribution).

D. Link with coherence bounds

It is also possible to obtain generalized Wielandt's inequality based bounds on the OLS selection rules that exploits the coherence rather than the RIC. For that, it is known that the

link between the eigenvalues of $\mathbf{A}_S^H \mathbf{A}_S$ and the coherence

$$\mu = \max_{i \neq j} \frac{|\langle \mathbf{a}_i, \mathbf{a}_j \rangle|}{\|\mathbf{a}_i\|_2 \|\mathbf{a}_j\|_2},$$

can be made through the Gershgorin disks theorem [4], [30]

$$1 - (K-1)\mu \leq \lambda_{\min}(\mathbf{A}_S^H \mathbf{A}_S) \leq \lambda_{\max}(\mathbf{A}_S^H \mathbf{A}_S) \leq 1 + (K-1)\mu,$$

assuming that the columns of the matrix \mathbf{A} are of unitary norm. As a consequence, provided $\mu < \frac{1}{K-1}$, one can formulate a RIP with a constant $(K-1)\mu$. Note that, by definition $\delta_K \leq (K-1)\mu$, except for $K=2$, where $\delta_2 = \mu$ [8]. Consequently, it is possible to convert our RIP bounds (6) and (7) into coherence bounds, replacing δ_K by $(K-1)\mu$.

Then, a sufficient condition for support recovery with OLS in absence of noise becomes analogously,

$$K\mu < \frac{\cos \Phi - (K-1)\mu}{1 - (K-1)\mu \cos \Phi},$$

which can be rewritten as

$$K(K-1) \cos \Phi \mu^2 - (2K-1)\mu + \cos \Phi > 0.$$

Finding the relevant root of this equation leads to the following condition on μ for support recovery

$$\mu < \frac{2K-1 - \sqrt{1 + 4K(K-1) \sin^2 \Phi}}{2K(K-1) \cos \Phi}. \quad (24)$$

Interestingly, note that when $\cos \Phi$ tends to one, that is to say when the signal amplitudes are quickly decaying, the condition (24) becomes $\mu < \frac{1}{K}$ which is the best relaxed bound achieved in [19]. But as soon as $\cos \Phi$ decreases, the condition (24) becomes stronger than the classical bound $\frac{1}{2K-1}$. Indeed, in the worst case, when $\cos \Phi = \frac{1}{\sqrt{K}}$, the condition (24) becomes

$$\mu < \frac{2K-1 - \sqrt{1 + 4(K-1)^2}}{2\sqrt{K}(K-1)},$$

whose right-hand side can be shown to be lower than $\frac{1}{2\sqrt{K}(K-1)}$. This suggests that the proposed bounds are not well suited to a coherence based study of OLS. The link between the eigenvalues $\lambda(\mathbf{A}_S^H \mathbf{A}_S)$ and μ is made through the Gershgorin Theorem, whereas, for the RIC the connection with $\lambda(\mathbf{A}_S^H \mathbf{A}_S)$ seems more direct since [8]

$$\delta_K = \max_{\Lambda/|\Lambda| \leq K} \lambda_{\max}(\mathbf{A}_\Lambda^H \mathbf{A}_\Lambda - \mathbf{I}).$$

V. EXACT SUPPORT RECOVERY CONDITIONS IN PRESENCE OF NOISE

Now, we consider the following signal in noise model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (25)$$

where the signal is still assumed to be K -sparse and where the noise \mathbf{n} can be either bounded or white circular complex Gaussian.

A. Bounded noise case

In this section, we consider the case of bounded noise. That is to say, there exists $\epsilon > 0$ such that

$$\|\mathbf{n}\|_2 < \epsilon.$$

We have the following result

Theorem 3. *If a K -sparse signal \mathbf{x} of support S follows the model (25) and if the condition (9) is fulfilled, OLS with the stopping rule $\|\mathbf{r}_k\|_2 < \epsilon$, recovers the support of \mathbf{x} in exactly K iterations and then stops, provided*

$$\min_{i \in S} |x_i| > \frac{2\epsilon}{\sqrt{1 - \delta_K} \left(\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} - \delta_{K+1} \right)}, \quad (26)$$

where $\cos \Phi$ is defined in (10).

Proof. The proof, inspired from [7], can be found in Appendix C. \square

The current sharpest condition for support recovery (for real signals) via OLS can be found in [18, Theorem 4], and states that

$$\min_{i \in S} |x_i| \geq \left(\frac{1}{1 - \delta_{K+1}/C_K} + \frac{1}{\sqrt{1 - \delta_{K+1}}} \right) \epsilon, \quad (27)$$

is a sufficient condition for OLS to recover the support in exactly K iterations, provided $\delta_{K+1} < C_K$. It can easily be shown that (27) also ensures that the stopping rule $\|\mathbf{r}_k\|_2 < \epsilon$ is not met until $k = K + 1$, so that it can be compared with the proposed condition (26). Note that when the signal amplitude do not possess any decay, this result is sharper than the proposed result (26). This is due to the fact that (27) is based on tighter bounds on the noisy selection rule. Since the relevant usage of the generalized Wielandt inequality requires in our case to investigate the angle between vectors than should be sparse, their approach does not seem to be directly applicable with the proposed bounds. Instead we based our approach on [12].

Contrary to (27), (26) depends on δ_K . A lower bound of (26) can be achieved when $\delta_K = 0$, whereas $\delta_K = \delta_{K+1}$ leads to an upper bound (worst case). In Figure 4, we compare both sufficient conditions (27) and (26) for different values of the parameter $\cos \Phi$. As we can see, when the signal is “flat”, the bound (27) is tighter than the proposed bound. But when the parameter $\cos \Phi$ is high enough (i.e. the decay rate is sufficient), the proposed bound is lower than (27). Note also, that in this case the condition imposed on δ_{K+1} is also weaker, as we have seen in the previous section and this can be observed in Figure 4. Consequently, it allows to extend the domain of validity of the proposed sufficient condition.

B. Gaussian noise case

We assume here that the noise is complex centered circular white Gaussian: $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$. In the Gaussian case, we adopt the following stopping criterion

$$\max_{i \notin S_k} \frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2} < \lambda. \quad (28)$$

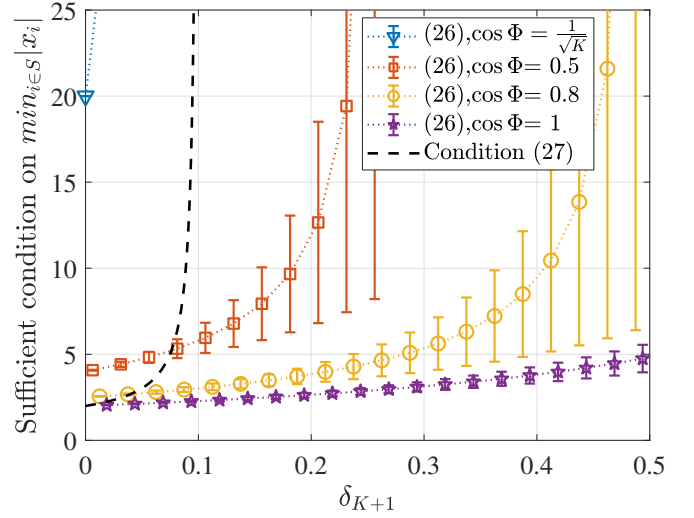


Figure 4. Sufficient condition (27) compared to (26) for different values of $\cos \Phi$. For (26), the upper bar corresponds to the upper bound obtained for $\delta_K = \delta_{K+1}$ whereas the lower bar stands for the case $\delta_K = 0$. $K = 100$, $\epsilon = 1$.

The threshold λ is set such that when the residual \mathbf{r}_k contains only noise, for any $i \notin S_k$

$$\frac{|\langle \mathbf{a}_i, \mathbf{P}_{S_k}^\perp \mathbf{n} \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2} \geq \lambda, \quad (29)$$

occurs with probability P_{fa} (tuned false alarm rate), for a given (non random) S_k .

Assuming a fixed set S_k (taking the true randomness nature of S_k into account is much more involved), $2 \frac{|\langle \mathbf{a}_i, \mathbf{P}_{S_k}^\perp \mathbf{n} \rangle|^2}{\sigma^2 \|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2^2}$ follows a centered χ_2^2 distribution of degree of freedom 2. Then $\lambda = \sigma \sqrt{\frac{1}{2} F_{\chi_2^2}^{-1}(1 - P_{fa})}$, where $F_{\chi_2^2}^{-1}(\cdot)$ is the inverse cumulative distribution function of the corresponding central Chi square distribution. The test (29) corresponds to the Generalized Likelihood Ratio Test when the previously estimated atoms in S_k have already been selected and are seen as interference [31].

Theorem 4. *If a K -sparse signal \mathbf{x} of support S follows the model (25) and if the condition (9) is fulfilled, OLS with termination rule (28) recovers the support of \mathbf{x} in exactly K iterations and then stops, provided*

$$\min_{i \in S} \frac{|x_i|}{\sigma} > \frac{\sqrt{2 F_{\chi_2^2}^{-1}(1 - P_{fa})}}{\sqrt{1 - \delta_K} \left(\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} - \delta_{K+1} \right)}, \quad (30)$$

where $\cos \Phi$ is defined in (10), with probability at least $1 - \sum_{k=0}^K (1 - (1 - P_{fa})^{N-k})$.

Proof. See Appendix D. \square

When we compare the proposed bound to the state of the art, the conclusions remain similar to the bounded noise case.

VI. CONCLUSION

In this paper, a new approach to Restricted Isometry Property based analysis of the Orthogonal Least Squares is

proposed. The generalized Wielandt inequality allowed us to derive new bounds on the algorithm's selection rule, that naturally take the signal's coefficient distribution into account. This enabled the derivation of a new sufficient condition for exact support recovery for Orthogonal Least Squares in absence of noise. This condition improves state of the art results, as soon as the signal's coefficients exhibit some sufficient decay. In this case, conditions on the Restricted Isometry Constant can be relaxed, as well as the minimal requirement on the signal amplitude for exact support recovery in noisy scenarios.

APPENDIX A
PROOF OF LEMMA 1

Let us first remark that for any $i \notin S$

$$\left| \langle \mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k \rangle \right| = \left| \left\langle \mathbf{P}_{S_k}^\perp \mathbf{A}_{i \cup S \setminus S_k} \mathbf{e}_1, \mathbf{P}_{S_k}^\perp \mathbf{A}_{i \cup S \setminus S_k} \begin{bmatrix} 0 \\ \mathbf{x}_S \end{bmatrix} \right\rangle \right|,$$

where \mathbf{e}_1 is a $(K+1) \times 1$ vector that contains zero everywhere except for the first component equal to 1. Applying the generalized Wielandt inequality (5) with $\mathbf{x} = \mathbf{e}_1$ and $\mathbf{y} = [0, \mathbf{x}_S^T]^T$, $\mathbf{B} = \left(\mathbf{A}_{i \cup S \setminus S_k}^H \mathbf{P}_{S_k}^\perp \mathbf{A}_{i \cup S \setminus S_k} \right)^{\frac{1}{2}}$ one finds that

$$\cos \phi = \frac{|\langle \mathbf{x}, \mathbf{y} \rangle|}{\|\mathbf{y}\|_2 \|\mathbf{x}\|_2} = 0,$$

and consequently since $|\mathbf{x}^H \mathbf{B}^H \mathbf{B} \mathbf{y}| = |\langle \mathbf{a}_i, \mathbf{r}_k \rangle|$, $\|\mathbf{B} \mathbf{x}\|_2 = \|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2$ and $\|\mathbf{B} \mathbf{y}\|_2 = \|\mathbf{r}_k\|_2$ with

$$\chi_1 = \frac{\kappa(\mathbf{B})^2 - 1}{\kappa(\mathbf{B})^2 + 1}, \quad (31)$$

$$\kappa(\mathbf{B})^2 = \frac{\lambda_{\max}(\mathbf{B}^H \mathbf{B})}{\lambda_{\min}(\mathbf{B}^H \mathbf{B})}, \quad (32)$$

we have that

$$\frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2 \|\mathbf{r}_k\|_2} \leq \chi_1.$$

The matrix \mathbf{B} is invertible, because its eigenvalues are comprised between the square root of the maximal and minimal eigenvalues of $\mathbf{A}_{i \cup S}^H \mathbf{A}_{i \cup S}$ (see Lemma 6), which is positive definite if $\delta_{K+1} < 1$.

Lemma 6. [7, Lemma 5] Let S_1 and S_2 be two sets such that $S_1 \cap S_2 = \emptyset$, then

$$\lambda_{\min}(\mathbf{A}_{S_1 \cup S_2}^H \mathbf{A}_{S_1 \cup S_2}) \leq \lambda_{\min}(\mathbf{A}_{S_1}^H \mathbf{P}_{S_2}^\perp \mathbf{A}_{S_1}),$$

$$\lambda_{\max}(\mathbf{A}_{S_1}^H \mathbf{P}_{S_2}^\perp \mathbf{A}_{S_1}) \leq \lambda_{\max}(\mathbf{A}_{S_1 \cup S_2}^H \mathbf{A}_{S_1 \cup S_2}).$$

As noted in [29, Remark 1], the link between the RIC and the eigenvalues of $\mathbf{A}_S^H \mathbf{A}_S$ can be made noticing that for any subset $S \subset \{1, \dots, N\}$ of cardinal K

$$1 - \delta_K \leq \lambda_{\min}(\mathbf{A}_S^H \mathbf{A}_S) \leq \lambda_{\max}(\mathbf{A}_S^H \mathbf{A}_S) \leq 1 + \delta_K. \quad (33)$$

Using Lemma 6, we have

$$\lambda_{\max}(\mathbf{B}^H \mathbf{B}) \leq \lambda_{\max}(\mathbf{A}_{i \cup S}^H \mathbf{A}_{i \cup S}) \leq 1 + \delta_{K+1},$$

and

$$\lambda_{\min}(\mathbf{B}^H \mathbf{B}) \geq \lambda_{\min}(\mathbf{A}_{i \cup S}^H \mathbf{A}_{i \cup S}) \geq 1 - \delta_{K+1},$$

which gives

$$\kappa(\mathbf{B})^2 = \frac{\lambda_{\max}(\mathbf{B}^H \mathbf{B})}{\lambda_{\min}(\mathbf{B}^H \mathbf{B})} \leq \frac{1 + \delta_{K+1}}{1 - \delta_{K+1}}. \quad (34)$$

Since χ_1 is increasing with the positive variable $\kappa(\mathbf{B})^2$, inserting (34) in (31) gives $\chi_1 \leq \frac{1 + \delta_{K+1} - 1}{1 + \delta_{K+1} + 1} = \delta_{K+1}$. This concludes the proof for the upper bound (6).

Now, for any $i \in S \setminus S_k$, let us remark that

$$|\langle \mathbf{a}_i, \mathbf{r}_k \rangle| = \left| \langle \mathbf{P}_{S_k}^\perp \mathbf{A}_{S \setminus S_k} \mathbf{e}_i, \mathbf{P}_{S_k}^\perp \mathbf{A}_{S \setminus S_k} \mathbf{x}_{S \setminus S_k} \rangle \right|,$$

where \mathbf{e}_i is now a $K \times 1$ vector that contains zero everywhere except for the column index of $\mathbf{A}_{S \setminus S_k}$ that corresponds to \mathbf{a}_i , where its component equals 1. Applying the generalized Wielandt inequality (5) with $\mathbf{x} = \mathbf{e}_i$, $\mathbf{y} = \mathbf{x}_{S \setminus S_k}$ and the following invertible matrix $\mathbf{B} = \left(\mathbf{A}_{S \setminus S_k}^H \mathbf{P}_{S_k}^\perp \mathbf{A}_{S \setminus S_k} \right)^{\frac{1}{2}}$ one finds that

$$\frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2 \|\mathbf{r}_k\|_2} \geq \frac{-\chi_2 + \cos \varphi_i}{1 - \chi_2 \cos \varphi_i},$$

with

$$\chi_2 = \frac{\kappa(\mathbf{B})^2 - 1}{\kappa(\mathbf{B})^2 + 1},$$

where

$$\kappa(\mathbf{B})^2 = \frac{\lambda_{\max}(\mathbf{A}_{S \setminus S_k}^H \mathbf{P}_{S_k}^\perp \mathbf{A}_{S \setminus S_k})}{\lambda_{\min}(\mathbf{A}_{S \setminus S_k}^H \mathbf{P}_{S_k}^\perp \mathbf{A}_{S \setminus S_k})}, \quad (35)$$

and

$$\cos \varphi_i = \frac{|\langle \mathbf{x}, \mathbf{y} \rangle|}{\|\mathbf{y}\|_2 \|\mathbf{x}\|_2} = \frac{|x_i|}{\|\mathbf{x}_{S \setminus S_k}\|_2}.$$

Then, it comes

$$\max_{i \in S \setminus S_k} \frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2 \|\mathbf{r}_k\|_2} \geq \max_{i \in S \setminus S_k} \frac{-\chi_2 + \cos \varphi_i}{1 - \chi_2 \cos \varphi_i}$$

where $\cos \phi_k$ is defined in (8). Using Lemma 7, it comes

$$\max_{i \in S \setminus S_k} \frac{-\chi_2 + \cos \varphi_i}{1 - \chi_2 \cos \varphi_i} = \frac{-\chi_2 + \cos \phi_k}{1 - \chi_2 \cos \phi_k}. \quad (36)$$

Lemma 7. The function $f(c, x) = \frac{-x+c}{1-cx}$ is respectively monotonically increasing and decreasing with respect to c and x , positive reals inferior to one.

Proof. This is straightforward, since the partial derivatives are such that $\frac{\partial f}{\partial c}(c, x) = \frac{1-x^2}{(1-cx)^2} \geq 0$ and $\frac{\partial f}{\partial x}(c, x) = \frac{c^2-1}{(1-cx)^2} \leq 0$ for $(c, x) \in [0, 1] \times [0, 1]$. \square

Using then Lemma 6, we have

$$\frac{\lambda_{\max}(\mathbf{A}_{S \setminus S_k}^H \mathbf{P}_{S_k}^\perp \mathbf{A}_{S \setminus S_k})}{\lambda_{\min}(\mathbf{A}_{S \setminus S_k}^H \mathbf{P}_{S_k}^\perp \mathbf{A}_{S \setminus S_k})} \leq \frac{1 + \delta_K}{1 - \delta_K},$$

that implies $\chi_2 \leq \delta_K$. The result (7) follows using Lemma 7.

APPENDIX B

COMPARISON OF THE BOUNDS (6) AND (7) TO THE STATE OF THE ART

A. *Upper bounds on $\cos \Theta(\mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k)$ for $i \notin S$ in the literature*

In the context of real signal and dictionary, the study [16] proposed the following bound

$$\max_{i \notin S} \frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{a}_i^H \mathbf{P}_{S_k}^\perp\|_2} < \frac{\|\mathbf{r}_k\|_2^2}{\sqrt{K-k} \|\mathbf{x}_{S \setminus S_k}\|_2}, \quad (37)$$

which has been improved in [17] by

$$\max_{i \notin S} \frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{a}_i^H \mathbf{P}_{S_k}^\perp\|_2} \leq \frac{\delta_{K+1} \|\mathbf{r}_k\|_2^2}{\|\mathbf{x}_{S \setminus S_k}\|_2}, \quad (38)$$

provided $\delta_{K+1} < \frac{1}{2}$. The bound is said to be tight in the sense that it is possible to provide a particular dictionary \mathbf{A} and signal \mathbf{x} (with all positive amplitude being equal to one) for which (38) is an equality [17]. We have

$$\begin{aligned} \|\mathbf{r}_k\|_2 &= \|\mathbf{P}_{S_k}^\perp \mathbf{A}_{S \setminus S_k} \mathbf{x}_{S \setminus S_k}\|_2 \\ &\stackrel{(a)}{\leq} \sqrt{\lambda_{\max}(\mathbf{A}_{S \setminus S_k}^H \mathbf{P}_{S_k}^\perp \mathbf{A}_{S \setminus S_k})} \|\mathbf{x}_{S \setminus S_k}\|_2 \\ &\stackrel{(b)}{\leq} \sqrt{\lambda_{\max}(\mathbf{A}_{S \setminus S_k}^H \mathbf{A}_{S \setminus S_k})} \|\mathbf{x}_{S \setminus S_k}\|_2 \\ &\leq \sqrt{1 + \delta_K} \|\mathbf{x}_{S \setminus S_k}\|_2, \end{aligned} \quad (39)$$

where (a) follows from standard results on Rayleigh quotient, (b) from the Lemma 6 in Appendix A and the last inequality from (33).

Inserting (39) in (37) and (38) gives respectively

$$\max_{i \notin S} \cos \Theta(\mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k) \leq \frac{\sqrt{1 + \delta_K}}{\sqrt{K-k}}, \quad (40)$$

and

$$\max_{i \notin S} \cos \Theta(\mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k) \leq \sqrt{1 + \delta_K} \delta_{K+1}. \quad (41)$$

Under the RIP condition $\delta_{K+1} < \frac{1}{\sqrt{K}}$ [17], (41) is tighter than (40). Both are clearly less informative than (6).

B. *Lower bounds on $\cos \Theta(\mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k)$ for $i \in S \setminus S_k$ in the literature*

The study [16] exploited the following lower bound

$$\max_{i \in S \setminus S_k} \frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{a}_i^H \mathbf{P}_{S_k}^\perp\|_2} \geq \frac{\|\mathbf{r}_k\|_2^2}{\sqrt{K-k} \|\mathbf{x}_{S \setminus S_k}\|_2}. \quad (42)$$

This bound is due to the fact that $\|\mathbf{a}_i^H \mathbf{P}_{S_k}^\perp\|_2 \leq 1$ and

$$\begin{aligned} \max_{i \in S \setminus S_k} |\langle \mathbf{a}_i^H \mathbf{r}_k \rangle| &\stackrel{(a)}{\geq} \frac{\|\mathbf{x}_{S \setminus S_k}\|_1}{\sqrt{K-k} \|\mathbf{x}_{S \setminus S_k}\|_2} \max_{i \in S \setminus S_k} |\langle \mathbf{a}_i, \mathbf{r}_k \rangle| \\ &= \frac{\sum_{l \in S \setminus S_k} |x_l| \max_{i \in S \setminus S_k} |\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\sqrt{K-k} \|\mathbf{x}_{S \setminus S_k}\|_2} \\ &\stackrel{(b)}{\geq} \frac{1}{\sqrt{K-k} \|\mathbf{x}_{S \setminus S_k}\|_2} \sum_{l \in S \setminus S_k} |x_l \mathbf{a}_i, \mathbf{r}_k \rangle| \\ &= \frac{\|\mathbf{r}_k\|_2^2}{\sqrt{K-k} \|\mathbf{x}_{S \setminus S_k}\|_2}, \end{aligned}$$

where (a) is due to the fact that for any vector of length N , $\|\mathbf{x}\|_1 \leq \sqrt{N} \|\mathbf{x}\|_2$ with equality if and only if all entries in \mathbf{x} have constant modulus. The equality in (b) does not seem to be easily reachable for any dictionary \mathbf{A} . This shows us that this bound should be rather tight when all signal's entries have the same magnitude.

With a similar reasoning as in (39), one has $\|\mathbf{r}_k\|_2 \geq \sqrt{1 - \delta_K} \|\mathbf{x}_{S \setminus S_k}\|_2$. One can then rewrite (42) as a cosine

$$\max_{i \in S \setminus S_k} \cos \Theta(\mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k) \geq \sqrt{\frac{1 - \delta_K}{K-k}}. \quad (43)$$

As we can see, if the signal is ‘‘flat’’ ($|x_i| = |x_j|$ for every $(i, j) \in S$), then $\cos \phi_k = \frac{1}{\sqrt{K-k}}$ and the bound (7) writes

$$\max_{i \in S \setminus S_k} \cos \Theta(\mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k) \geq \frac{1 - \sqrt{K-k} \delta_K}{\sqrt{K-k} - \delta_K}.$$

Since δ_K is usually much smaller than $\sqrt{K-k}$, the latter bound is usually a worse bound than (43) in the case of signals of even amplitudes.

When $\cos \phi_k$ is large, that is to say when the signals exhibit some decay, the proposed bound increases (see Lemma 7 in Appendix A) and becomes more informative than (42) which does not depend on $\cos \phi_k$ (i.e. the signal magnitude distribution).

APPENDIX C
PROOF OF THEOREM 3

OLS succeeds in reconstruction at iteration k if

$$\max_{i \notin S} \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1} \rangle|}{\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i\|_2} < \max_{i \in S \setminus S_k} \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1} \rangle|}{\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i\|_2}. \quad (44)$$

By definition,

$$\begin{aligned} \mathbf{r}_{k-1} &= \mathbf{P}_{S_{k-1}}^\perp (\mathbf{A} \mathbf{x} + \mathbf{n}) \\ &= \mathbf{r}_{k-1}^0 + \mathbf{P}_{S_{k-1}}^\perp \mathbf{n}, \end{aligned}$$

where $\mathbf{r}_{k-1}^0 = \mathbf{P}_{S_{k-1}}^\perp \mathbf{A} \mathbf{x}$ is the noiseless residual at iteration k .

Since one can write, using the triangular and Cauchy-Schwartz inequalities

$$\begin{aligned} \left| \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1} \rangle|}{\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i\|_2} - \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1}^0 \rangle|}{\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i\|_2} \right| &\leq \frac{|\langle \mathbf{a}_i, \mathbf{P}_{S_{k-1}}^\perp \mathbf{n} \rangle|}{\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i\|_2} \\ &\leq \|\mathbf{n}\|_2 \\ &< \epsilon. \end{aligned}$$

Thus, combining this with (44), OLS succeeds if

$$\max_{i \in S \setminus S_k} \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1}^0 \rangle|}{\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i\|_2} - \max_{i \notin S} \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1}^0 \rangle|}{\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i\|_2} > 2\epsilon,$$

using (6) and (7), it is sufficient that

$$\|\mathbf{r}_{k-1}^0\|_2 \left(\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} - \delta_{K+1} \right) > 2\epsilon. \quad (45)$$

Since

$$\begin{aligned} \|\mathbf{r}_{k-1}^0\|_2 &\geq \sqrt{1-\delta_K} \|\mathbf{x}_{S \setminus S_{k-1}}\|_2 \\ &\geq \sqrt{1-\delta_K} \sqrt{K-k+1} \min_{i \in S \setminus S_{k-1}} |x_i|, \end{aligned} \quad (46)$$

it is sufficient to have

$$\min_{i \in S \setminus S_{k-1}} |x_i| \left(\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} - \delta_{K+1} \right) > \frac{2\epsilon}{\sqrt{1-\delta_K} \sqrt{K-k+1}},$$

for every k . A sufficient condition is then (26) for the OLS to select a correct atom of the support at each iteration k .

At the iteration $1 \leq k \leq K$, we also have

$$\begin{aligned} \|\mathbf{r}_k\|_2 &= \|\mathbf{P}_{S_k}^\perp \mathbf{A} \mathbf{x} + \mathbf{P}_{S_k}^\perp \mathbf{n}\|_2 \\ &\geq \|\mathbf{r}_k^0\|_2 - \|\mathbf{P}_{S_k}^\perp \mathbf{n}\|_2, \end{aligned}$$

if (26) holds true then, so is (45). We finally have

$$\begin{aligned} \|\mathbf{r}_k\|_2 &\geq \frac{2\epsilon}{\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} - \delta_{K+1}} - \epsilon \\ &\geq \epsilon, \end{aligned}$$

so that the algorithm does not stop until $k = K$, where $\|\mathbf{r}_K^0\|_2 = 0$ and $\|\mathbf{r}_K\|_2 = \|\mathbf{P}_{S_K}^\perp \mathbf{n}\|_2 \leq \|\mathbf{n}\|_2 < \epsilon$. Consequently, the OLS algorithm stops after iteration K .

APPENDIX D PROOF OF THEOREM 4

Let us denote, for any set S_k composed of correct atoms

$$N_k = \max_{i \notin S_k} \frac{|\langle \mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{n} \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2}.$$

Following the same reasoning as for the bounded noise case, we have that OLS selects a correct atom at step k , as soon as

$$\max_{i \in S \setminus S_{k-1}} \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1}^0 \rangle|}{\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i\|_2} - \max_{i \notin S} \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1}^0 \rangle|}{\|\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i\|_2} > 2N_{k-1}.$$

Following the same lines that led to (45) and (46), one can deduce the following sufficient condition for OLS to select a correct atom

$$\min_{i \in S \setminus S_{k-1}} |x_i| > \frac{2N_{k-1}}{\sqrt{1-\delta_K} \sqrt{K-k+1} \left(\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} - \delta_{K+1} \right)}.$$

The latter is true when

$$\min_{i \in S \setminus S_{k-1}} |x_i| > \frac{2N_{k-1}}{\sqrt{1-\delta_K} \left(\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} - \delta_{K+1} \right)}. \quad (47)$$

Now, if (30) holds true, (47) is valid as soon as $N_{k-1} \leq \lambda$ and OLS selects a correct atom at each step k .

We now have to check that the stopping rule is not met until $k = K$. For $k \leq K$

$$\begin{aligned} \max_i \frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2} &\geq \frac{|\langle \mathbf{a}_i, \mathbf{r}_k^0 \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2} - N_k \\ &\geq \frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} \|\mathbf{r}_k^0\|_2 - N_k, \end{aligned}$$

when (28) holds true, then

$$\|\mathbf{r}_k^0\|_2 \geq \frac{2\lambda}{\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} - \delta_{K+1}}.$$

Consequently, if $N_k < \lambda$,

$$\max_i \frac{|\langle \mathbf{a}_i, \mathbf{r}_k \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2} \leq \lambda,$$

and the stopping rule is not met. When $k = K$ then $\max_i \frac{|\langle \mathbf{a}_i^H \mathbf{r}_k \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2} = N_K$ so that the stopping rule is met if $N_K \leq \lambda$.

Finally, when (30) is true, the probability for OLS to succeed at each step and then stop is at least the probability of the event $E = \bigcap_{0 \leq i \leq K} N_k < \lambda$. We have, using the classical union bound,

$$\begin{aligned} Pr(E) &= 1 - Pr \left(\bigcup_{0 \leq i \leq K} \bar{E}_i \right), \\ &\geq 1 - \sum_{i=0}^K Pr(N_k \geq \lambda). \end{aligned}$$

The proof is concluded using the fact that

$$\begin{aligned} Pr(N_k < \lambda) &= Pr \left(\frac{|\langle \mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{n} \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2} < \lambda, \forall i \notin S_k \right) \\ &\geq \prod_{i=1, i \notin S_k}^N Pr \left(\frac{|\langle \mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{n} \rangle|}{\|\mathbf{P}_{S_k}^\perp \mathbf{a}_i\|_2} < \lambda \right) \\ &= (1 - P_{fa})^{N-k}, \end{aligned}$$

where the inequality is due to [32].

REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [3] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, 1993.
- [4] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [5] 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, 1989.
- [6] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, "Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 3158–3174, 2013.
- [7] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information theory*, vol. 57, no. 7, pp. 4680–4688, 2011.
- [8] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.
- [9] M. A. Davenport and M. B. Wakin, "Analysis of orthogonal matching pursuit using the restricted isometry property," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4395–4401, 2010.
- [10] Q. Mo and Y. Shen, "A remark on the restricted isometry property in orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3654–3656, 2012.

- [11] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4973–4976, 2012.
- [12] R. Wu, W. Huang, and D.-R. Chen, "The exact support recovery of sparse signals with noise via orthogonal matching pursuit," *IEEE signal processing letters*, vol. 20, no. 4, pp. 403–406, 2012.
- [13] L.-H. Chang and J.-Y. Wu, "An improved rip-based performance guarantee for sparse signal recovery via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5702–5715, 2014.
- [14] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, "A sharp condition for exact support recovery with orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1370–1382, 2016.
- [15] C. Liu, Y. Fang, and J. Liu, "Some new results about sufficient conditions for exact support recovery of sparse signals via orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4511–4524, 2017.
- [16] J. Wen, J. Wang, and Q. Zhang, "Nearly optimal bounds for orthogonal least squares," *IEEE Transactions on Signal Processing*, vol. 65, no. 20, pp. 5347–5356, 2017.
- [17] J. Kim and B. Shim, "Optimal restricted isometry condition for exact sparse recovery with orthogonal least squares," *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 1385–1390, 2020.
- [18] J. Kim, J. Wang, and B. Shim, "Optimal restricted isometry condition of normalized sampling matrices for exact sparse recovery with orthogonal least squares," *IEEE Transactions on Signal Processing*, vol. 69, pp. 1521–1536, 2021.
- [19] C. Herzet, A. Drémeau, and C. Soussen, "Relaxed recovery conditions for omp/ols by exploiting both coherence and decay," *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 459–470, 2015.
- [20] K. Schnass, "Average performance of orthogonal matching pursuit (omp) for sparse approximation," *IEEE Signal Processing Letters*, vol. 25, no. 12, pp. 1865–1869, 2018.
- [21] J. Ding, L. Chen, and Y. Gu, "Perturbation analysis of orthogonal matching pursuit," *IEEE Transactions on Signal processing*, vol. 61, no. 2, pp. 398–410, 2012.
- [22] J. Wen, Z. Zhou, Z. Liu, M.-J. Lai, and X. Tang, "Sharp sufficient conditions for stable recovery of block sparse signals by block orthogonal matching pursuit," *Applied and Computational Harmonic Analysis*, vol. 47, no. 3, pp. 948–974, 2019.
- [23] Z. Xu, "The performance of orthogonal multi-matching pursuit under the restricted isometry property," *Journal of Computational Mathematics*, pp. 495–516, 2015.
- [24] F. L. Bauer and A. S. Householder, "Some inequalities involving the euclidean condition of a matrix," *Numerische Mathematik*, vol. 2, no. 1, pp. 308–311, 1960.
- [25] M. Lin and G. Sinnamon, "The generalized wielandt inequality in inner product spaces," *Eurasian Mathematical Journal*, vol. 3, no. 1, pp. 72–85, 2012.
- [26] R. Behrens and L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Transactions on Signal Processing*, vol. 42, no. 6, pp. 1413–1424, 1994.
- [27] T. Blumensath and M. E. Davies, "On the difference between orthogonal matching pursuit and orthogonal least squares," 2007.
- [28] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, Second Edition, 2013.
- [29] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [30] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed sensing: theory and applications*, Y. C. Eldar and G. Kutyniok, Eds. Cambridge University Press Cambridge, 2012, ch. 1.
- [31] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Transactions on signal processing*, vol. 42, no. 8, pp. 2146–2157, 1994.
- [32] Z. Šidák, "Rectangular confidence regions for the means of multivariate normal distributions," *Journal of the American Statistical Association*, vol. 62, no. 318, pp. 626–633, 1967.