On the Noise Robustness of Simultaneous Orthogonal Matching Pursuit

Jean-François Determe* Jérôme Louveaux† Laurent Jacques† François Horlin*

February 20, 2017

IEEE copyright notice – published paper: J. Determe, J. Louveaux, L. Jacques and F. Horlin, "On the Noise Robustness of Simultaneous Orthogonal Matching Pursuit," in *IEEE Transactions on Signal Processing*, vol. 65, no. 4, pp. 864-875, Feb. 15 2017. doi: 10.1109/TSP.2016.2626244. `http://ieeexplore.ieee.org/document/7738592/`

arXiv:1602.03400v3 [cs.IT] 17 Feb 2017

Abstract

In this paper, the joint support recovery of several sparse signals whose supports exhibit similarities is examined. Each sparse signal is acquired using the same noisy linear measurement process, which returns fewer observations than the dimension of the sparse signals. The measurement noise is assumed additive, Gaussian, and admits different variances for each sparse signal that is measured. Using the theory of compressed sensing, the performance of simultaneous orthogonal matching pursuit (SOMP) is analyzed for the envisioned signal model. The cornerstone of this paper is a novel analysis method upper bounding the probability that SOMP recovers at least one incorrect entry of the joint support during a prescribed number of iterations. Furthermore, the probability of SOMP failing is investigated whenever the number of sparse signals being recovered simultaneously increases and tends to infinity. In particular, convincing observations and theoretical results suggest that SOMP committing no mistake in the noiseless case does not guarantee the absence of error in the noisy case whenever the number of acquired sparse signals scales to infinity. Finally, simulation results confirm the validity of the theoretical results.

1 Introduction

The recovery of sparse signals, *i.e.*, signals exhibiting a low number of non-zero entries, acquired through noisy linear measurements is a central problem in digital signal processing. This task is further involved when the number of measurements is lower than the dimension of the signal to be recovered. Recently, researchers have paid special attention to this class of problems due to the emergence of the *compressed sensing* (CS) field of research [15], which aims at providing reliable recovery methods of sparse signals for which the number of measurements is low.

Before describing our problem, we wish to introduce key notions. The cardinality of a set A is denoted by $|A|$. The notation $[n]$ denotes the set $\{1, 2, \ldots, n\}$ and x_j refers to the j-th

*Jean-François Determe and François Horlin are with the OPERA Wireless Communications Group, Université libre de Bruxelles, 1050 Brussels, Belgium. E-mail: jdeterme@ulb.ac.be, fhorlin@ulb.ac.be. Jean-François Determe is funded by the Belgian National Science Foundation (F.R.S.-FNRS).

†Laurent Jacques, Jérôme Louveaux, and Jean-François Determe are with the ICTEAM departement, Université catholique de Louvain. E-mail: laurent.jacques@uclouvain.be, jerome.louveaux@uclouvain.be. Laurent Jacques is funded by the Belgian National Science Foundation (F.R.S.-FNRS).

entry of $\boldsymbol{x}$. The support of any vector $\boldsymbol{x} \in \mathbb{R}^n$ is defined as $\text{supp}(\boldsymbol{x}) := \{j \in [n] : x_j \neq 0\}$. We define s-sparse signals as vectors whose supports exhibit a cardinality equal to or less than s. Loosely speaking, a signal $\boldsymbol{x} \in \mathbb{R}^n$ is said to be sparse whenever its support is significantly smaller than the dimension of its space, *i.e.*, $|\text{supp}(\boldsymbol{x})| \ll n$. Finally, the notation $A \gtrsim B$ means that $\exists\, c > 0 : A \geq cB$.

1.1 Objective and signal model

We focus on a scenario where the objective is recovering the joint support $\mathcal{S}$ of K sparse signals $\boldsymbol{x}_k \in \mathbb{R}^n$ $(1 \leq k \leq K)$, *i.e.*, $\mathcal{S} := \text{supp}(\boldsymbol{X}) := \cup_{k\in[K]}\text{supp}(\boldsymbol{x}_k)$, observed by means of a common measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ for which $m < n$. The resulting measurement vectors $\boldsymbol{y}_k \in \mathbb{R}^m$ $(1 \leq k \leq K)$ gather the measurements of each sparse signal: $\boldsymbol{y}_k = \boldsymbol{\Phi}\boldsymbol{x}_k + \boldsymbol{e}_k$ where $\boldsymbol{e}_k$ is an additive noise term. It is assumed that $\boldsymbol{e}_k$ is distributed as $\mathcal{N}(0, \sigma_k^2\boldsymbol{I}_{m\times m})$ and that, for $k_1 \neq k_2$, $\boldsymbol{e}_{k_1}$ and $\boldsymbol{e}_{k_2}$ are statistically independent. The vector of the noise standard deviations is denoted by $\boldsymbol{\sigma} := (\sigma_1, \ldots, \sigma_K)^{\mathrm{T}}$.

For the sake of simplicity, Equation (1) aggregates the K equations $\boldsymbol{y}_k = \boldsymbol{\Phi}\boldsymbol{x}_k + \boldsymbol{e}_k$ into a single relationship:

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{X} + \boldsymbol{E} \tag{1}$$

where $\boldsymbol{Y} = \big(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_K\big) \in \mathbb{R}^{m\times K}$, $\boldsymbol{X} = \big(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K\big) \in \mathbb{R}^{n\times K}$ and $\boldsymbol{E} = \big(\boldsymbol{e}_1, \ldots, \boldsymbol{e}_K\big) \in \mathbb{R}^{m\times K}$. Note that the signal models incorporating only one sparse vector are called single measurement vector (SMV) models while those where $K > 1$ sparse signals are measured are referred to as multiple measurement vector (MMV) models [17].

The MMV signal model above applies to several scenarios. For example, in [24, Section IV.B], the source localization problem is studied when using measurements at different time instants. The signal model describing their problem is equivalent to Equation (1), where each measurement vector in $\boldsymbol{Y}$ corresponds to one time instant. In [1, Section 3], the authors describe jointly sparse signal models comparable to ours that typically occur in networks of sensors where a possibly large number of sensing nodes exchange their measurements of the same object to reconstruct it. In this case, each individual measurement vector within $\boldsymbol{Y}$ is generated by one sensor. Other applications are described in the survey [27].

In this paper, the *atoms* should be understood as being the columns of $\boldsymbol{\Phi}$, *i.e.*, $\{\boldsymbol{\phi}_j\}_{j\in[n]}$ where $\boldsymbol{\phi}_j$ denotes the j-th column of $\boldsymbol{\Phi}$. Although the notion of atom is typically used when dealing with dictionaries, we use it here as it simplifies our discussions. Note that most of the results presented in this paper assume that all the atoms of $\boldsymbol{\Phi}$ have a unit ℓ_2 norm.

1.2 Detailed contribution

Our main contribution is a method to analyze SOMP in a noisy setting. The principal technical contribution is an upper bound on the probability that SOMP identifies incorrect entries, *i.e.*, entries not belonging to $\mathcal{S}$, during a prescribed number of iterations. Using quantities describing the reliability of SOMP in the noiseless case, we then show that the probability of incorrect identification decreases exponentially with the number of sparse signals K under the condition that the amplitudes of the sparse signals are sufficiently high when compared to the noise variances.

We also establish that the guarantee that SOMP correctly identifies all the entries of the joint support in the absence of noise is not adequate in noisy scenarios when K tends to ∞. The development presented in this paper sheds light on this phenomenon by providing two convincing explanations for its existence.

As discussed at the end of the paper, an interesting corollary of our upper bound is the asymptotic exact recovery condition (AERC) for SOMP, *i.e.*, a condition ensuring that the probability of SOMP failing falls to 0 as $K \to \infty$. The condition we derive actually also guarantees an arbitrarily small probability of error for a sufficiently high value of K, which is a stronger result than the AERC. In particular, our very final result involves four fundamental quantities:

- The number of sparse signals K, which defines how many independent measurement channels are available for the joint recovery.
- The quantity $\mu_X(K) := \min_{j\in\mathcal{S}} \frac{1}{K}\sum_{k=1}^{K} |X_{j,k}|$, which sets the minimal averaged amplitudes of the coefficients associated with each atom indexed by the support $\mathcal{S}$.
- The quantity $\sigma(K)^2 := (1/K)\sum_{k=1}^{K}\sigma_k^2$ is the average noise power on all the K sparse signals (or measurement channels).
- The quantity $\omega_\sigma(K) := (1/\sqrt{K})\|\boldsymbol{\sigma}\|_1/\|\boldsymbol{\sigma}\|_2 \in [1/\sqrt{K}; 1]$ quantifies how close to a sparse vector $\boldsymbol{\sigma}$ is. In particular, if $\boldsymbol{\sigma}$ is 1-sparse, then $\omega_\sigma(K) = 1/\sqrt{K}$ while $\omega_\sigma = 1$ when its entries are identical. In practice, we prefer working with the upper bound $\omega_\sigma = \max_{1\leq K<\infty}\omega_\sigma(K)$ belonging to $(0; 1]$.

Given the four quantities above, the minimum signal-to-mean-noise ratio $\mathrm{SNR}_{\min}$ is defined as $\min_{1\leq K<\infty}\frac{\mu_X(K)}{\sigma(K)}$. It is shown that the probability of SOMP committing at least one error during its first $s+1$ iterations is no more than p_{err} if

$$K \gtrsim \frac{1}{(\alpha\,\mathrm{SNR}_{\min} - \beta\omega_\sigma)^2}\left(\log n + s\log\frac{2e|\mathcal{S}|}{s} - \log p_{\mathrm{err}}\right) \tag{2}$$

where the condition is valid only if $\mathrm{SNR}_{\min} > \frac{\beta\omega_\sigma}{\alpha}$. In particular, if $\mathrm{SNR}_{\min} \gg \frac{\beta\omega_\sigma}{\alpha}$, then $\alpha\mathrm{SNR}_{\min} - \beta\omega_\sigma \simeq \alpha\mathrm{SNR}_{\min}$. As clarified later on, the term β is upper bounded by $\sqrt{2/\pi}$ while α quantifies the sensing properties of the matrix $\boldsymbol{\Phi}$ as well as the reliability of SOMP decisions in the noiseless case. Note that the condition $\mathrm{SNR}_{\min} > \frac{\beta\omega_\sigma}{\alpha}$ testifies to the impossibility of performing correct decisions unless the SNR is above a certain minimal floor level.

Finally, some numerical simulations validate our methodology by showing that the expression appearing in Equation (2) is accurate up to some numerical adjustments aiming at tightening large theoretical constants.

1.3 Related work

In a SMV setting, full support recovery guarantees for OMP with bounded noise signals as well as with Gaussian noises have been proposed in [2]. This work also provides criteria on the stopping criteria to guarantee that OMP terminates after having picked all the correct atoms. This contribution has then been slightly refined in [8] to provide conditions independent of the particular support that is to be recovered.

Gribonval *et al.* have investigated the performance of SOMP for a problem resembling ours in [21]. Their contribution has been to provide a lower bound on the probability of correct full support recovery when the signal to be estimated is sparse and its non-zero entries are statistically independent mean-zero Gaussian random variables.

1.4 Outline

First of all, Section 2 describes SOMP and related quantities. Technical prerequisites are delivered to the reader in Section 3. In Section 4, we present some results on SOMP in the noiseless case that will be used afterwards. Then, Section 5 provides upper bounds on the probability that SOMP fails to identify a correct atom at a fixed iteration. These results are finally exploited in Section 6 to deliver usable and easily interpretable upper bounds on the probability that SOMP includes at least one incorrect entry to the estimated support for a prescribed number of iterations. Numerical results presented in Section 7 confirm the validity of our results. The conclusion then follows. Most of the technicalities are reported in the Appendix to simplify the presentation in the core of the paper.

1.5 Conventions

We find useful to introduce the main notations used in this paper. For $1 \leq p < \infty$ and $\boldsymbol{x} \in \mathbb{R}^n$, we have $\|\boldsymbol{x}\|_p := (\sum_{j=1}^n |x_j|^p)^{1/p}$ and $\|\boldsymbol{x}\|_\infty := \max_{j\in[n]} |x_j|$. With $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$, we define $\|\boldsymbol{\Phi}\|_{p\to q}$ as [20, Equation A.8] $\|\boldsymbol{\Phi}\|_{p\to q} := \sup_{\|\boldsymbol{\phi}\|_p=1} \|\boldsymbol{\Phi}\boldsymbol{\phi}\|_q$ where $1 \leq p, q \leq \infty$. In particular, with $\boldsymbol{A} \in \mathbb{R}^{n\times K}$, $\|\boldsymbol{A}\|_{\infty\to\infty}$ is equal to $\max_{j\in[n]} \sum_{k=1}^K |A_{j,k}|$ [20, Lemma A.5]. Unless otherwise specified, every vector is to be understood as a column vector. Also, for $\mathcal{S} \subseteq [n]$, $\boldsymbol{x}_\mathcal{S}$ denotes the vector formed by the entries of $\boldsymbol{x}$ indexed within $\mathcal{S}$. In a likewise fashion, for $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$, we define $\boldsymbol{\Phi}_\mathcal{S}$ as the matrix formed by the columns of $\boldsymbol{\Phi}$ whose indexes belong to $\mathcal{S}$. The notation $\overline{\mathcal{S}}$ refers to the relative complement of $\mathcal{S}$ with respect to $[n]$. The Moore-Penrose pseudoinverse of any matrix $\boldsymbol{\Phi}$ is given by $\boldsymbol{\Phi}^+$ while its transpose is denoted by $\boldsymbol{\Phi}^\mathrm{T}$. The range of any matrix $\boldsymbol{\Phi}$, *i.e.*, the space spanned by its columns, is denoted by $\mathcal{R}(\boldsymbol{\Phi})$. The inner product of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, written as $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$, is given by $\boldsymbol{x}^\mathrm{T}\boldsymbol{y}$. The minimum and maximum eigenvalues of a matrix $\boldsymbol{\Phi}$ are denoted by $\lambda_{\min}(\boldsymbol{\Phi})$ and $\lambda_{\max}(\boldsymbol{\Phi})$, respectively. The probability measure is given by $\mathbb{P}$ while the mathematical expectation is denoted by $\mathbb{E}$.

2 Simultaneous orthogonal matching pursuit

Many algorithms have been proposed to solve the joint support recovery problem associated with Equation (1). In a SMV setting, canonical algorithms include ℓ_1-minimization [5, 16], matching pursuit (MP) [25] and orthogonal matching pursuit (OMP) [10, 26]. While algorithms relying on ℓ_1-minimization are probably among the most reliable algorithms designed for compressed sensing problems, they often exhibit a higher computational complexity than their greedy counterparts, such as MP and OMP [30]. Greedy algorithms are thus more suited to real time applications. The well-known algorithms described above for SMV problems admit several extensions within the MMV framework. Specifically, one of the most natural generalization of OMP is SOMP [31], which is described in Algorithm 1.

As described in Algorithm 1, SOMP updates the support at each iteration t by including in the current estimated support a single atom $\boldsymbol{\phi}_{j_t}$, whose index is denoted by j_t, which maximizes SOMP metric

$$\|(\boldsymbol{R}^{(t)})^\mathrm{T}\boldsymbol{\phi}_j\|_1 = \sum_{k=1}^K |\langle \boldsymbol{\phi}_j, \boldsymbol{r}_k^{(t)} \rangle|$$

ALGORITHM 1:
Simultaneous orthogonal matching pursuit (SOMP)

Require: $\boldsymbol{Y} \in \mathbb{R}^{m\times K}$, $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$, $s \geq 1$
1: Initialization: $\boldsymbol{R}^{(0)} \leftarrow \boldsymbol{Y}$ and $\mathcal{S}_0 \leftarrow \emptyset$
2: $t \leftarrow 0$
3: **while** $t < s$ **do**
4: Determine the atom of $\boldsymbol{\Phi}$ to be included in the support:
$j_t \leftarrow \operatorname{argmax}_{j\in[n]}(\|(\boldsymbol{R}^{(t)})^{\mathrm{T}}\boldsymbol{\phi}_j\|_1)$
5: Update the support : $\mathcal{S}_{t+1} \leftarrow \mathcal{S}_t \cup \{j_t\}$
6: Projection of each measurement vector onto $\mathcal{R}(\boldsymbol{\Phi}_{S_{t+1}})$:
$\boldsymbol{Y}^{(t+1)} \leftarrow \boldsymbol{\Phi}_{\mathcal{S}_{t+1}}\boldsymbol{\Phi}^{+}_{\mathcal{S}_{t+1}}\boldsymbol{Y}$
7: Projection of each measurement vector onto $\mathcal{R}(\boldsymbol{\Phi}_{\mathcal{S}_{t+1}})^{\perp}$:
$\boldsymbol{R}^{(t+1)} \leftarrow \boldsymbol{Y} - \boldsymbol{Y}^{(t+1)}$
8: $t \leftarrow t+1$
9: **end while**
10: **return** $\mathcal{S}_s$ {Support at last step}

(steps 4 and 5). In this description of SOMP, $\boldsymbol{r}_k^{(t)}$ denotes the k-th column of the residual matrix $\boldsymbol{R}^{(t)}$ at iteration t. During steps 6 and 7, each measurement vector $\boldsymbol{y}_k$ is projected onto the orthogonal complement of $\mathcal{R}(\boldsymbol{\Phi}_{\mathcal{S}_{t+1}})$, denoted by $\mathcal{R}(\boldsymbol{\Phi}_{\mathcal{S}_{t+1}})^{\perp}$. In this way, an atom cannot be picked twice since, once included in the support, the projection onto $\mathcal{R}(\boldsymbol{\Phi}_{\mathcal{S}_{t+1}})^{\perp}$ ensures that $\langle\boldsymbol{\phi}_j, \boldsymbol{r}_k^{(t+1)}\rangle = 0$ if $j \in \mathcal{S}_t$. The algorithm finishes when the prescribed number of iterations, $s+1$, is attained. We now turn to the description of useful quantities related to SOMP.

The atoms indexed by the joint support $\mathcal{S}$ are referred to as the correct atoms while the incorrect atoms are those not indexed by $\mathcal{S}$. The set $\mathcal{P}^{(t)}$ contains the $\binom{|\mathcal{S}|}{t}$ orthogonal projectors $\boldsymbol{\Phi}_{\mathcal{S}_t}\boldsymbol{\Phi}^{+}_{S_t}$ such that $\mathcal{S}_t \subseteq \mathcal{S}$ and $|\mathcal{S}_t| = t$. $\boldsymbol{P}^{(0)}$ is defined as the zero matrix. Loosely speaking, $\mathcal{P}^{(t)}$ is the set of all the possible orthogonal projectors $\boldsymbol{P}^{(t)}$ at iteration t assuming that only atoms belonging to $\mathcal{S}$ have been picked previously. Enumerating all the possible orthogonal projection matrices at each iteration by means of $\mathcal{P}^{(t)}$ will be necessary later on since not knowing the sequence of the atoms picked by SOMP beforehand requires to consider all the possible projectors. For each iteration t, we also define

$$\beta_{j,k}^{(t,\boldsymbol{P})} := |\langle\boldsymbol{\phi}_j, (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\Phi}\boldsymbol{x}_k\rangle| \tag{3}$$

where $\boldsymbol{P}$ belongs to $\mathcal{P}^{(t)}$. The value of $\beta_{j,k}^{(t,\boldsymbol{P})}$ consequently is the inner product of the j-th atom with the k-th column of the residual matrix whenever the orthogonal projector at iteration t is $\boldsymbol{P} \in \mathcal{P}^{(t)}$ and no noise is present.

We define $\gamma_c^{(t,\boldsymbol{P})}$ as the highest SOMP metric obtained for the correct atoms at iteration t without noise and, similarly, $\gamma_i^{(t,\boldsymbol{P})}$ is the counterpart quantity for the incorrect atoms instead:

$$\gamma_c^{(t,\boldsymbol{P})} := \max_{j\in S}\sum_{k=1}^{K}\beta_{j,k}^{(t,\boldsymbol{P})} \text{ and } \gamma_i^{(t,\boldsymbol{P})} := \max_{j\in\overline{\mathcal{S}}}\sum_{k=1}^{K}\beta_{j,k}^{(t,\boldsymbol{P})}. \tag{4}$$

It is also convenient to define two quantities identifying which are the best correct atom and incorrect atom in the noiseless case, at iteration t and for the orthogonal projection matrix $\boldsymbol{P}$:

$$j_c^{(t,\boldsymbol{P})} = \arg\max_{j\in S}\sum_{k=1}^{K}\beta_{j,k}^{(t,\boldsymbol{P})} \text{ and } j_i^{(t,\boldsymbol{P})} = \arg\max_{j\in\overline{\mathcal{S}}}\sum_{k=1}^{K}\beta_{j,k}^{(t,\boldsymbol{P})}$$

where the subscripts c and i refer to correct and incorrect atoms, respectively.

3 Technical prerequisites & Notations

We now wish to settle key theoretical notions to be used later on for the actual analysis of SOMP.

3.1 Restricted isometry property and related concepts

Any $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$ exhibits the restricted isometry property (RIP) [5] of order s if there exists a constant $\delta_s < 1$ such that

$$(1 - \delta_s)\|\boldsymbol{u}\|_2^2 \leq \|\boldsymbol{\Phi}\boldsymbol{u}\|_2^2 \leq (1 + \delta_s)\|\boldsymbol{u}\|_2^2 \tag{5}$$

for all s-sparse vectors $\boldsymbol{u}$. The smallest δ_s for which Equation (5) holds is the Rectricted Isometry Constant (RIC) of order s. The RIC of order s is theoretically given by $\delta_s = \max(U_s, L_s)$ where $U_s = \max_{\mathcal{S} \subseteq [n], |\mathcal{S}|=s} \lambda_{\max}(\boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}}) - 1$ and $L_s = 1 - \min_{S \subseteq [n], |\mathcal{S}|=s} \lambda_{\min}(\boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}})$. The RIC therefore provides an upper bound on the alteration of the ℓ_2 norm of sparse vectors after multiplication by $\boldsymbol{\Phi}$. Interestingly enough, $\delta_s < 1$ implies that $\boldsymbol{\Phi}_{\mathcal{S}}$ has full column rank for every support $\mathcal{S}$ of size $|\mathcal{S}| \leq s$ since $\min_{S \subseteq [n], |\mathcal{S}|=s} \lambda_{\min}(\boldsymbol{\Phi}_{\mathcal{S}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}}) > 0$.

3.2 Lipschitz functions

A function $f : \mathbb{R}^K \to \mathbb{R} : \boldsymbol{g} \mapsto f(\boldsymbol{g})$ is called a Lipschitz function with regard to the ℓ_2-norm if and only if

$$\exists L > 0 : \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^K, \, |f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L \, \|\boldsymbol{x} - \boldsymbol{y}\|_2 \tag{6}$$

holds. The constant L is then referred to as the Lipschitz constant. An interesting property of Lipschitz functions is the following concentration inequality for standard Gaussian random vectors $\boldsymbol{g}$ [20, Theorem 8.40.]:

$$\mathbb{P}\left(f(\boldsymbol{g}) - \mathbb{E}[f(\boldsymbol{g})] \geq \varepsilon\right) \leq \exp\left(\frac{-\varepsilon^2}{2L^2}\right). \tag{7}$$

This result shows that, for $\boldsymbol{g} \sim \mathcal{N}(0, \boldsymbol{I}_{K \times K})$, $f(\boldsymbol{g})$ concentrates around its expectation and the concentration improves as L decreases. The similar inequality $\mathbb{P}(-f(\boldsymbol{g}) + \mathbb{E}[f(\boldsymbol{g})] \geq \varepsilon) \leq \exp(-\varepsilon^2/(2L^2))$, where $\varepsilon > 0$, results from the observation that, if f is Lipschitz, then so is $-f$.

3.3 On the folded normal distribution

A recurring distribution in this paper is the folded normal distribution, which is the absolute value of a normal random variable. If $X \sim \mathcal{N}(\beta, \sigma^2)$, then $Y = |X|$ is the associated folded normal random variable. We often refer to X as the underlying normal variable. Similarly, the underlying mean and variances are those of the underlying normal variable. Note that the expectation of the folded random variable is always higher than that of its underlying normal random variable because of the folding of the probability density function (PDF) occurring for values lower than 0.

4 Results on SOMP without noise

As discussed later on, the reliability of SOMP in the noiseless case determines which noise levels are unlikely to make SOMP detect incorrect atoms. In this section, we thus provide a lower bound on the measure of the noiseless reliability given by $\gamma_c^{(t,\boldsymbol{P})} - \gamma_i^{(t,\boldsymbol{P})}$, *i.e.*, the quantity that fixes the minimum gap of the SOMP metrics for correct and incorrect atoms in the noiseless case. To do so, we assume the existence of three quantities. The first quantity, denoted by Γ, quantifies the minimum relative reliability of SOMP decisions:

$$\frac{\gamma_c^{(t,\boldsymbol{P})}}{\gamma_i^{(t,\boldsymbol{P})}} \geq \Gamma > 1 \text{ for all } t \in \big[0, s\big] \text{ and all } \boldsymbol{P} \in \mathcal{P}^{(t)}. \tag{8}$$

The remaining quantities ψ and τ_X are entwined and provide a lower bound on the SOMP metric for the best correct atom:

$$\gamma_c^{(t,\boldsymbol{P})} \geq \psi\tau_X \text{ for all } t \in \big[0, s\big] \text{ and all } \boldsymbol{P} \in \mathcal{P}^{(t)}. \tag{9}$$

The quantity $\psi = \psi(\delta_{|\mathcal{S}|}, |\mathcal{S}|)$ typically increases as $\delta_{|\mathcal{S}|}$ and $|\mathcal{S}|$ dwindle. This property conveys the idea that a small support size $|\mathcal{S}|$ and a measurement matrix $\boldsymbol{\Phi}$ endowed with good CS properties tend to increase the value of $\gamma_c^{(t,\boldsymbol{P})}$. Finally, τ_X is related to the intrinsic energy of the sparse signals $\boldsymbol{x}_k$ prior to their projection by $\boldsymbol{\Phi}$. Therefore, τ_X only depends on $\boldsymbol{X}$. Combining Γ, ψ, and τ_X yields the desired lower bound on the absolute reliability of SOMP in the noiseless case:

Lemma 1. *For every $t \in \big[0, s\big]$ and $\boldsymbol{P} \in \mathcal{P}^{(t)}$, we have*

$$\gamma_c^{(t,\boldsymbol{P})} - \gamma_i^{(t,\boldsymbol{P})} \geq \left(1 - \frac{1}{\Gamma}\right) \psi\tau_X.$$

where Γ, ψ, and τ_X are defined in Equations (8)-(9).

The following subsections aim at providing evidence for the existence of Γ, ψ, and τ_X. In particular, valid expressions are provided for these parameters in light of previous works in the literature.

4.1 A lower bound on SOMP relative reliability

In this section, we briefly discuss the existence of lower bounds for $\gamma_c^{(t,\boldsymbol{P})}/\gamma_i^{(t,\boldsymbol{P})}$, *i.e.*, possible expressions for Γ, quantifying how reliably SOMP distinguishes correct and incorrect atoms in the noiseless case. In [13], we have proposed several expressions of Γ that use the RIC and sometimes the restricted orthogonality constant (ROC), thereby only requiring to know of the residual support size or, equivalently, the current iteration number t. For example, we have shown that $\frac{\gamma_c^{(t,\boldsymbol{P})}}{\gamma_i^{(t,\boldsymbol{P})}} \geq \frac{(1-\delta_{|\mathcal{S}|})\sqrt{|\mathcal{S}|-1}}{\delta_{|\mathcal{S}|}|\mathcal{S}|}$ for all $t < |\mathcal{S}|$ and for $\delta_{|\mathcal{S}|} < 1$. Similarly, a bound using the support itself can be obtained by a straightforward adaptation of the proof of [7, Theorem 4.5], *i.e.*, $\frac{\gamma_c^{(t,\boldsymbol{P})}}{\gamma_i^{(t,\boldsymbol{P})}} \geq \frac{1}{\left\|\boldsymbol{\Phi}_{\mathcal{S}}^{+}\boldsymbol{\Phi}_{\overline{\mathcal{S}}}\right\|_{1\to 1}}$ if $\boldsymbol{\Phi}_{\mathcal{S}}$ has full column rank. Note that in both papers, the bounds hold only if correct decisions have been made during all the iterations preceding t. We will not get further into the details of such matters as our only objective here is to provide evidence for the existence of Γ.

4.2 A lower bound on $\gamma_c^{(t,\boldsymbol{P})}$

We now present a convenient lower bound for $\gamma_c^{(t,\boldsymbol{P})}$ and the associated expressions for ψ and τ_X.

Lemma 2. *Adapted from [12]. If $\boldsymbol{\Phi}$ satisfies the RIP with $|\mathcal{S}|$-th RIC $\delta_{|\mathcal{S}|} < 1$, then*

$$\gamma_c^{(t,\boldsymbol{P})} \geq \frac{(1-\delta_{|\mathcal{S}|})(1+\delta_{|\mathcal{S}|})}{1+\sqrt{|\mathcal{S}|}\delta_{|\mathcal{S}|}} \min_{j\in\mathcal{S}} \sum_{k=1}^{K} |X_{j,k}|$$

where $\gamma_c^{(t,\boldsymbol{P})}$ is defined in Section 2.

This lemma shows that $\gamma_c^{(t,\boldsymbol{P})}$ depends on the CS properties of $\boldsymbol{\Phi}$ through the RIC $\delta_{|\mathcal{S}|}$ while $\min_{j\in\mathcal{S}} \sum_{k=1}^{K} |X_{j,k}|$ indicates that the sum of the absolute coefficients of $\boldsymbol{X}$ associated with each atom also influences $\gamma_c^{(t,\boldsymbol{P})}$. Lemma 2 thus provides $\tau_X = \min_{j\in\mathcal{S}} \sum_{k=1}^{K} |X_{j,k}|$ and $\psi = (1-\delta_{|\mathcal{S}|})(1+\delta_{|\mathcal{S}|})/(1+\sqrt{|\mathcal{S}|}\delta_{|\mathcal{S}|})$.

Notice that, without restrictions on the sparse signal matrix $\boldsymbol{X}$, nothing can be said about the non-maximum SOMP metrics for the atoms belonging to the correct support $\mathcal{S}$, *i.e.*, it might happen that all the SOMP metrics for the correct atoms are zero except for the highest one among them. As a result, we only focus on the correct atom index by $j_c^{(t,\boldsymbol{P})}$ and the associated noiseless SOMP metric $\gamma_c^{(t,\boldsymbol{P})}$. A short example is available in Section .1 to prove this statement.

5 Upper bounds on the probability of SOMP failing at iteration t

This section provides an upper bound on the probability that SOMP picks an incorrect atom at iteration t given a fixed orthogonal projector $\boldsymbol{P} \in \mathcal{P}^{(t)}$. This time, the derived results include the noise. First of all, we examine the statistical distribution of SOMP metric for a single atom in Section 5.1. The desired upper bound will then be derived.

5.1 On the distribution of $\|(\boldsymbol{R}^{(t)})^{\mathrm{T}}\boldsymbol{\phi}_j\|_1$

The quantity $\|(\boldsymbol{R}^{(t)})^{\mathrm{T}}\boldsymbol{\phi}_j\|_1 = \sum_{k=1}^{K} |\langle \boldsymbol{\phi}_j, \boldsymbol{r}_k^{(t)} \rangle|$ ultimately defines which atom is picked at each iteration. It is therefore interesting to determine its statistical distribution. We have

$$\begin{aligned}\sum_{k=1}^{K} |\langle \boldsymbol{\phi}_j, \boldsymbol{r}_k^{(t)} \rangle| &= \sum_{k=1}^{K} |\langle \boldsymbol{\phi}_j, (\boldsymbol{I}-\boldsymbol{P}^{(t)})(\boldsymbol{\Phi}\boldsymbol{x}_k + \boldsymbol{e}_k) \rangle| \\ &= \sum_{k=1}^{K} |\langle \boldsymbol{\phi}_j^{(t)}, \boldsymbol{\Phi}\boldsymbol{x}_k \rangle + \langle \boldsymbol{\phi}_j^{(t)}, \boldsymbol{e}_k \rangle|\end{aligned}$$

where $\boldsymbol{\phi}_j^{(t)} := (\boldsymbol{I}-\boldsymbol{P}^{(t)})\boldsymbol{\phi}_j$. Let us consider a fixed projection matrix $\boldsymbol{P}^{(t)} = \boldsymbol{P}$. It is easy to prove that $\langle (\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\phi}_j, \boldsymbol{e}_k \rangle$ is distributed as $\mathcal{N}(0, (\sigma_{j,k}^{(\boldsymbol{P})})^2)$ where

$$\sigma_{j,k}^{(\boldsymbol{P})} := \|(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\phi}_j\|_2 \sigma_k \leq \sigma_k \tag{10}$$

provided that $\|\boldsymbol{\phi}_j\|_2 = 1$ .We define the related noise standard deviation vectors

$$\boldsymbol{\sigma}_j^{(\boldsymbol{P})} := \|(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{\phi}_j\|_2 \boldsymbol{\sigma} \tag{11}$$

for each atom. It is thereby possible to replace the term $\langle \boldsymbol{\phi}_j^{(t)}, \boldsymbol{e}_k \rangle$ by $\sigma_{j,k}^{(\boldsymbol{P})} g_k$ where $g_k \sim \mathcal{N}(0,1)$. Moreover, $|\langle \boldsymbol{\phi}_j^{(t)}, \boldsymbol{\Phi}\boldsymbol{x}_k \rangle + \sigma_{j,k}^{(\boldsymbol{P})} g_k|$ is distributed as $||\langle \boldsymbol{\phi}_j^{(t)}, \boldsymbol{\Phi}\boldsymbol{x}_k \rangle| + \sigma_{j,k}^{(\boldsymbol{P})} g_k|$. Hence, $\|(\boldsymbol{R}^{(t)})^{\mathrm{T}} \boldsymbol{\phi}_j\|_1$ is a sum of K folded normal random variables and, since $|\langle \boldsymbol{\phi}_j^{(t)}, \boldsymbol{\Phi}\boldsymbol{x}_k \rangle| = \beta_{j,k}^{(t,\boldsymbol{P})}$, it admits the same distribution as:

$$f_j^{(t,\boldsymbol{P})} : \mathbb{R}^K \to \mathbb{R} : \boldsymbol{g} \mapsto \sum_{k=1}^{K} \left| \beta_{j,k}^{(t,\boldsymbol{P})} + \sigma_{j,k}^{(\boldsymbol{P})} g_k \right|. \tag{12}$$

5.2 On the probability of SOMP picking an incorrect atom at iteration t

In this section, we provide a general upper bound on the probability that SOMP picks an incorrect atom at iteration t and for a fixed orthogonal projector $\boldsymbol{P} \in \mathcal{P}^{(t)}$. The idea of the proof is to notice that if the metrics associated with every incorrect atom is lower than a real positive number α and the metric associated with one of the correct atom is higher than α, then a correct decision will necessarily be made. This approach is pessimistic in the sense that more than one specific correct atoms could be picked. Note that, among all the correct atoms, only the best atom in the noiseless case is considered, *i.e.*, the atom indexed by $j_c^{(t,\boldsymbol{P})}$. The proof of Theorem 1 is available in the Appendix.

Theorem 1. *For a fixed iteration t, let $\boldsymbol{P} \in \mathcal{P}^{(t)}$ and $\boldsymbol{R} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$, i.e., $\boldsymbol{R}$ is one of the residuals that could be generated by SOMP on the basis of $\boldsymbol{Y}$ at iteration $t-1$ assuming that only correct atoms have been identified. For $\boldsymbol{g} \sim \mathcal{N}(0, \boldsymbol{I}_{K\times K})$ and for all $\alpha > 0$, the probability of SOMP picking an incorrect atom when running one iteration on $\boldsymbol{R}$ is upper bounded by*

$$\mathbb{P}\left[f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})}(\boldsymbol{g}) \leq \alpha \right] + \sum_{j \in \overline{\mathcal{S}}} \mathbb{P}\left[f_j^{(t,\boldsymbol{P})}(\boldsymbol{g}) \geq \alpha \right]. \tag{13}$$

In Theorem 1, we do not assume that $\boldsymbol{R}$ has been generated on the basis of past iterations of SOMP, *i.e.*, the upper bound we derive is independent of the way $\boldsymbol{R}$ has been obtained. Ignoring this precaution would imply that we were able to upper bound the conditional probability given the event that SOMP succeeded during the previous iterations, which would be more involved than what Theorem 1 establishes.

Now that only probabilities of the form $\mathbb{P}\big[f_j^{(t,\boldsymbol{P})}(\boldsymbol{g}) \geq \alpha\big]$ and $\mathbb{P}\big[f_j^{(t,\boldsymbol{P})}(\boldsymbol{g}) \leq \alpha\big]$ intervene, it is appropriate to find upper bounds for these probabilities and use them to produce a more easily interpretable result, *i.e.*, Theorem 2. The idea of the proof of Theorem 2 is to set the convex combination

$$\alpha = \lambda \mathbb{E}\big[f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})}(\boldsymbol{g})\big] + (1-\lambda) \max_{j \in \overline{\mathcal{S}}} \mathbb{E}[f_j^{(t,\boldsymbol{P})}(\boldsymbol{g})], \tag{14}$$

i.e., express α in Theorem 1 relatively to $\mathbb{E}\big[f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})}(\boldsymbol{g})\big]$ and $\max_{j\in\overline{\mathcal{S}}} \mathbb{E}[f_j^{(t,\boldsymbol{P})}(\boldsymbol{g})]$, and then use the concentration inequalities in Section 3.2 in conjunction with the fact that $f_j^{(t,\boldsymbol{P})}(\boldsymbol{g})$ is Lipschitz (see Lemma 2.1). It has been chosen to set $\lambda = 0.5$ to simplify the final result. A visual interpretation of Theorem 2 is depicted in Figure 1. For the sake of simplicity, the figure considers identical noise levels for each atom (remember that $\boldsymbol{\sigma}_j^{(\boldsymbol{P})} := \|(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{\phi}_j\|_2 \boldsymbol{\sigma}$ so that the noise may exhibit different powers for each atom in the general case). The full proofs are available in the Appendix.

Lemma 2.1. *Each function $f_j^{(t,\boldsymbol{P})}$ defined in Equation (12) is a Lipschitz function whose best Lipschitz constant is $\|\boldsymbol{\sigma}_j^{(\boldsymbol{P})}\|_2$.*

Theorem 2. *Let $\boldsymbol{g} \sim \mathcal{N}(0, \boldsymbol{I}_{K\times K})$. For a fixed iteration t, let $\boldsymbol{P} \in \mathcal{P}^{(t)}$ and $\boldsymbol{R} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{Y}$, i.e., $\boldsymbol{R}$ is one of the residuals that could be generated by SOMP on the basis of $\boldsymbol{Y}$ at iteration $t-1$ assuming that only correct atoms have been identified so far. Let*

$$\Delta\mathbb{E}^{(t,\boldsymbol{P})} := \mathbb{E}\left[f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})}(\boldsymbol{g})\right] - \max_{j\in\overline{\mathcal{S}}} \mathbb{E}\left[f_j^{(t,\boldsymbol{P})}(\boldsymbol{g})\right] \tag{15}$$

and assume $\|\boldsymbol{\phi}_j\|_2 = 1$ for $j \in [n]$. If $\Delta\mathbb{E}^{(t,\boldsymbol{P})} > 0$, then the probability of SOMP making an incorrect decision when executing one iteration on $\boldsymbol{R}$ is upper bounded by

$$(n - |\mathcal{S}| + 1)\exp\left[-\frac{1}{8\|\boldsymbol{\sigma}\|_2^2}(\Delta\mathbb{E}^{(t,\boldsymbol{P})})^2\right]. \tag{16}$$

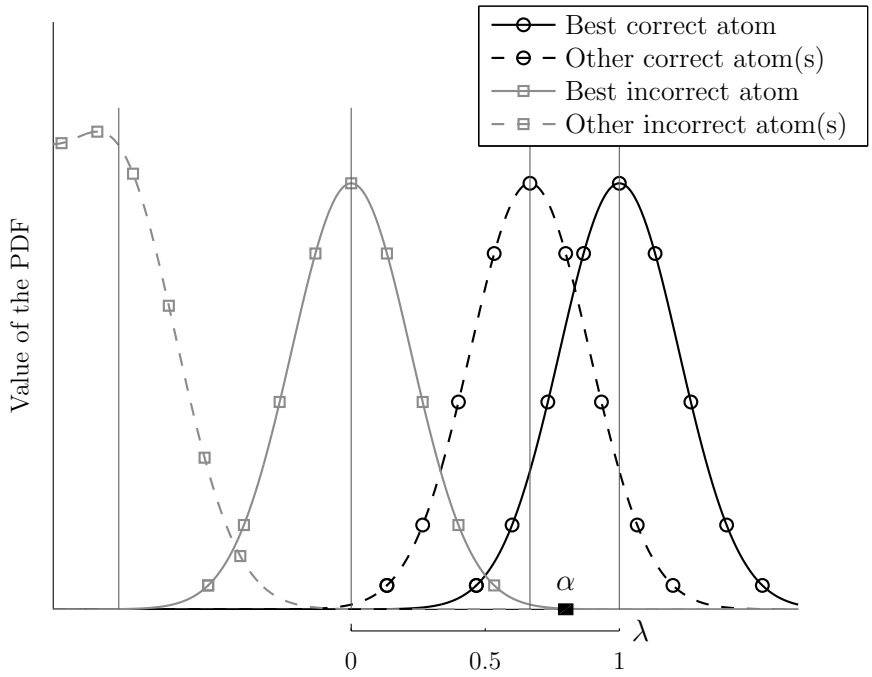


Figure 1: Explanation of Theorem 2 – Probability density function of $f_j^{(t,\boldsymbol{P})}(\boldsymbol{g})$ for $\boldsymbol{g} \sim \mathcal{N}(0, \boldsymbol{I}_{K\times K})$ – $\boldsymbol{\sigma}_{j_1}^{(\boldsymbol{P})} = \boldsymbol{\sigma}_{j_2}^{(\boldsymbol{P})}$ for $j_1, j_2 \in [n]$ – The vertical lines represent the mean of the folded normal distribution associated with each atom, *i.e.* $\mathbb{E}\big[f_j^{(t,\boldsymbol{P})}(\boldsymbol{g})\big]$ for the j-th atom – The position of α is obtained for $\lambda = 0.8$ – Only two atoms belonging to $\mathcal{S}$ (including the best one) and two atoms belonging to $\overline{\mathcal{S}}$ (including the best one) are represented for the sake of clarity.

In Figure 1, it is observed that the mean of the folded normal distribution and the underlying normal distribution are virtually equal for sufficiently high values of $\sum_{k=1}^{K} \beta_{j,k}^{(t,\boldsymbol{P})}$ while a discrepancy is observed for low values, as for the second incorrect atom in Figure 1. As explained in Section 3.3, taking the absolute value of a normal distribution yields a folding of the PDF and thereby increases the mean with regard to the underlying normal distribution. It implies that, for two atom indexes j_1 and j_2, $\gamma_{j_1}^{(t,\boldsymbol{P})} > \gamma_{j_2}^{(t,\boldsymbol{P})}$ is no guarantee for $\mathbb{E}\big[f_{j_1}^{(t,\boldsymbol{P})}(\boldsymbol{g})\big] > \mathbb{E}\big[f_{j_2}^{(t,\boldsymbol{P})}(\boldsymbol{g})\big]$ where $\gamma_j^{(t,\boldsymbol{P})} := \sum_{k=1}^{K} \beta_{j,k}^{(t,\boldsymbol{P})}$ denotes the value of SOMP metric for the j-th atom. We provide two explanations for this phenomenon:

- The noise vectors $\boldsymbol{\sigma}_j^{(\boldsymbol{P})}$ may exhibit different ℓ_2 norms for each atom. Thus, even for $K = 1$, a possibly higher noise level for the j_2-th atom and a sufficiently small gap $\gamma_{j_1}^{(t,\boldsymbol{P})} - \gamma_{j_2}^{(t,\boldsymbol{P})}$ might yield $\mathbb{E}[f_{j_1}^{(t,\boldsymbol{P})}(\boldsymbol{g})] < \mathbb{E}[f_{j_2}^{(t,\boldsymbol{P})}(\boldsymbol{g})]$ even though $\gamma_{j_1}^{(t,\boldsymbol{P})} > \gamma_{j_2}^{(t,\boldsymbol{P})}$. Note that this phenomenon cannot occur for iteration 0 as the noise variances are necessarily equal in that case since $\boldsymbol{P} = \boldsymbol{0}$.

- For $K > 1$, and for possibly identical noise vectors $\boldsymbol{\sigma}_j^{(\boldsymbol{P})}$, the way the $\beta_{j,k}^{(t,\boldsymbol{P})}$ are distributed for k plays a significant role. For example, if only one of the K quantities $\beta_{j_2,k}^{(t,\boldsymbol{P})}$ is overwhelmingly greater than the others then very limited folding will occur on this entry while significant folding is present on the other ones. Conversely, if the entries of $\beta_{j_1,k}^{(t,\boldsymbol{P})}$ are more uniformly distributed, then the overall increase in expectation due to the folding is decreased. Thus, depending on the way the K terms $\beta_{j,k}^{(t,\boldsymbol{P})}$ are distributed for j_1 and j_2, $\mathbb{E}[f_{j_1}^{(t,\boldsymbol{P})}(\boldsymbol{g})] < \mathbb{E}[f_{j_2}^{(t,\boldsymbol{P})}(\boldsymbol{g})]$ might hold despite having $\gamma_{j_1}^{(t,\boldsymbol{P})} > \gamma_{j_2}^{(t,\boldsymbol{P})}$.

Note that the reason why $\max_{j\in\overline{\mathcal{S}}} \mathbb{E}\big[f_j^{(t,\boldsymbol{P})}(\boldsymbol{g})\big]$ cannot be replaced by $\mathbb{E}\big[f_{j_i^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})}(\boldsymbol{g})\big]$ in the definition of $\Delta\mathbb{E}^{(t,\boldsymbol{P})}$ is similar to the two explanations above.

As stated in the abstract, the two points above provide convincing explanations of why $\gamma_c^{(t,\boldsymbol{P})} > \gamma_i^{(t,\boldsymbol{P})}$, *i.e.*, success in the noiseless case, is no guarantee for correct recovery when $K \to \infty$. Indeed, each random variable $(1/K)\mathbb{E}\big[f_j^{(t,\boldsymbol{P})}(\boldsymbol{g})\big]$ concentrates arbitrarily well around its expectation for a sufficiently high value of K. It is therefore to be expected that asymptotic correct recovery at iteration t and for orthogonal projector $\boldsymbol{P}$, *i.e.*, recovery when $K \to \infty$, is ensured if and only if $\Delta\mathbb{E}^{(t,\boldsymbol{P})} > 0$. The condition $\gamma_c^{(t,\boldsymbol{P})} > \gamma_i^{(t,\boldsymbol{P})}$ is actually neither necessary nor sufficient for correct asymptotic recovery as the phenomena discussed above might benefit the correct atoms, *i.e.*, the condition is not necessary, or the incorrect atoms, *i.e.*, the condition is not sufficient. However, the analyses that follow rely on the conservative assumption that only the incorrect atoms benefit from the increase of expectation due to the folding.

6 Upper bound on the probability of SOMP failing during the first $s + 1$ iterations

The final problem to be addressed is to derive an upper bound on the probability of SOMP picking incorrect atoms during the first $s + 1$ iterations. In particular, the bound should not require to know the sequence of orthogonal projectors intervening at each iteration. To do so, we design a lower bound for $\Delta\mathbb{E}^{(t,\boldsymbol{P})}$ in Theorem 2 independent of the orthogonal projector $\boldsymbol{P} \in \mathcal{P}^{(t)}$ and that conveys in a simple manner the impact on SOMP performance of

- The sensing properties of the matrix $\boldsymbol{\Phi}$.

- The absolute reliability of SOMP in the noiseless case, *i.e.*, $\gamma_c^{(t,\boldsymbol{P})} - \gamma_i^{(t,\boldsymbol{P})}$, and the associated lower bound in Lemma 1.

- The noise variances σ_k^2.

Once this bound is obtained, we will show that Theorem 2 yields an upper bound on the probability of failure of SOMP from iteration 0 to iteration $s < |\mathcal{S}|$ included, *i.e.*, an upper bound on the probability that SOMP picks at least one incorrect atom during the first $s + 1$ iterations.

6.1 Deriving the lower bound of $\Delta\mathbb{E}^{(t,\boldsymbol{P})}$

In this section, we propose a lower bound for $\mathbb{E}[f^{(t,\boldsymbol{P})}_{j_c^{(t,\boldsymbol{P})}}(\boldsymbol{g})]$ and an upper bound for $\max_{j\in\overline{\mathcal{S}}}\mathbb{E}[f^{(t,\boldsymbol{P})}_{j}(\boldsymbol{g})]$, that are both valid for all $\boldsymbol{P}\in\mathcal{P}^{(t)}$, so that a lower bound for $\Delta\mathbb{E}^{(t,\boldsymbol{P})}$ is obtained. Exceptionally, the proof is presented in the core of the paper as doing so gives the reader a better understanding of the derivation method. Regarding $\mathbb{E}[f^{(t,\boldsymbol{P})}_{j_c^{(t,\boldsymbol{P})}}(\boldsymbol{g})]$, we have

$$\begin{aligned}\mathbb{E}\left[f^{(t,\boldsymbol{P})}_{j_c^{(t,\boldsymbol{P})}}(\boldsymbol{g})\right] &= \sum_{k=1}^{K}\mathbb{E}\left[\left|\beta^{(t,\boldsymbol{P})}_{j_c^{(t,\boldsymbol{P})},k}+\sigma^{(\boldsymbol{P})}_{j_c^{(t,\boldsymbol{P})},k}g_k\right|\right]\\ &\geq \sum_{k=1}^{K}\left|\beta^{(t,\boldsymbol{P})}_{j_c^{(t,\boldsymbol{P})},k}+\sigma^{(\boldsymbol{P})}_{j_c^{(t,\boldsymbol{P})},k}\mathbb{E}\left[g_k\right]\right|\\ &= \sum_{k=1}^{K}\left|\beta^{(t,\boldsymbol{P})}_{j_c^{(t,\boldsymbol{P})},k}\right| =: \gamma_c^{(t,\boldsymbol{P})}\end{aligned}$$

where the inequality results from Jensen's inequality for convex functions. Unless $\beta^{(t,\boldsymbol{P})}_{j_c^{(t,\boldsymbol{P})},k}/\sigma^{(\boldsymbol{P})}_{j_c^{(t,\boldsymbol{P})},k}\to 0$, the bound above is not sharp as it discards the gain of expectation resulting from the folding. Moreover, the triangle inequality provides

$$\begin{aligned}\mathbb{E}\left[f^{(t,\boldsymbol{P})}_{j}(\boldsymbol{g})\right] &= \mathbb{E}\left[\sum_{k=1}^{K}\left|\beta^{(t,\boldsymbol{P})}_{j,k}+\sigma^{(\boldsymbol{P})}_{j,k}g_k\right|\right]\\ &\leq \sum_{k=1}^{K}\left|\beta^{(t,\boldsymbol{P})}_{j,k}\right|+\sum_{k=1}^{K}\sigma^{(\boldsymbol{P})}_{j,k}\mathbb{E}\left[|g_k|\right]\end{aligned}$$

where $\mathbb{E}[|g_k|]=\sqrt{2/\pi}$ and $\sum_{k=1}^{K}|\beta^{(t,\boldsymbol{P})}_{j,k}|\leq\sum_{k=1}^{K}|\beta^{(t,\boldsymbol{P})}_{j_i^{(t,\boldsymbol{P})},k}|=\gamma_i^{(t,\boldsymbol{P})}$ for $j\in\overline{\mathcal{S}}$. The inequality is not sharp as it assumes maximum folding, *i.e.*, the folding obtained if the underlying mean is 0, for all the terms $|\beta^{(t,\boldsymbol{P})}_{j,k}+\sigma^{(\boldsymbol{P})}_{j,k}g_k|$ while, in the general case, at least one of the $\beta^{(t,\boldsymbol{P})}_{j,k}$ should be non-zero. Also, the inequality $\sigma^{(\boldsymbol{P})}_{j,k}\leq\sigma_k$ holds (see Section 5.1). As a conclusion, we obtain

$$\Delta\mathbb{E}^{(t,\boldsymbol{P})}\geq(\gamma_c^{(t,\boldsymbol{P})}-\gamma_i^{(t,\boldsymbol{P})})-\sqrt{\frac{2}{\pi}}\|\boldsymbol{\sigma}\|_1 \tag{17}$$

where an expression quantifying the absolute reliability of the decisions in the noiseless case intervenes, *i.e.*, $\gamma_c^{(t,\boldsymbol{P})}-\gamma_i^{(t,\boldsymbol{P})}$, as well as a penalty depending on the noise standard deviations for all the K channels, *i.e.*, $\sqrt{2/\pi}\|\boldsymbol{\sigma}\|_1$. Using Lemma 1 yields

$$\Delta\mathbb{E}^{(t,\boldsymbol{P})}\geq\left(1-\frac{1}{\Gamma}\right)\psi\tau_X-\sqrt{\frac{2}{\pi}}\|\boldsymbol{\sigma}\|_1$$

where theoretical expressions for Γ, ψ, and τ_X are discussed in Section 4. Note that the result only holds whenever $\Gamma>1$ for the theoretical expression that has been chosen. The next step is to combine the derived lower bound with Theorem 2.

6.2 An upper bound on the probability of failure of SOMP during the first $s+1$ iterations

The main difficulty here is to derive the desired upper bound not knowing the sequence of supports that are chosen during the first $s+1$ iterations. To circumvent this issue, all the possible

orthogonal projection matrices $\boldsymbol{P} \in \mathcal{P}^{(t)}$ are tested while only one intervenes in practice. This sub-optimal approach is linked to the fact that $\boldsymbol{P}$ and $\boldsymbol{E}$ are statistically dependent random variables as the noise contributes to determining which atom is picked and thus influences the orthogonal projection matrices. The statistical link between both variables appears to be difficult to capture and we consequently chose to use this workaround. This "trick" has been used in similar theoretical analyses [18, 21] and, to the best of the authors' knowledge, a better solution has yet to be found in the literature. The proof is available in the Appendix.

Theorem 3. *Using the quantities Γ, ψ, and τ_X defined in Equation (8) and (9), let us consider*

$$\Delta\mathbb{E} := \left(1 - \frac{1}{\Gamma}\right)\psi\tau_X - \sqrt{\frac{2}{\pi}}\|\boldsymbol{\sigma}\|_1 \tag{18}$$

and assume $\|\boldsymbol{\phi}_j\|_2 = 1$ for $j \in [n]$. If $\Delta\mathbb{E} > 0$, then the probability of SOMP picking at least one incorrect atom from iteration 0 to iteration s included is upper bounded by

$$n\mathcal{C}_s \exp\left[-\frac{1}{8\|\boldsymbol{\sigma}\|_2^2}(\Delta\mathbb{E})^2\right] \tag{19}$$

where $\mathcal{C}_s := \sum_{t=0}^{s}\binom{|\mathcal{S}|}{t} \overset{(s\geq 1)}{\leq} \left(\frac{e(|\mathcal{S}|+s-1)}{s}\right)^s + 1$.

The expressions appearing in Theorem 3 can be deciphered in the following manner:

- As explained in Section 4, the term $(1 - 1/\Gamma)\psi\tau_X$ conveys the impact of the reliability of the decisions prior to the addition of noise.
- Both terms $\sqrt{2/\pi}\|\boldsymbol{\sigma}\|_1$ and $1/\|\boldsymbol{\sigma}\|_2^2$ convey the negative influence of the noise on SOMP performance. The phenomena they account for are however different in nature. The quantity $1/\|\boldsymbol{\sigma}\|_2^2$ translates the spread of the PDF of each $f_j^{(t,\boldsymbol{P})}$ which is characterized by the spread of the Gaussian-like functions in Figure 1. On the other hand, the expression $\sqrt{2/\pi}\|\boldsymbol{\sigma}\|_1$ describes the negative impact that the noise has on the existing reliability in the noiseless case, *i.e.*, $\gamma_c^{(t,\boldsymbol{P})} - \gamma_i^{(t,\boldsymbol{P})}$ in Equation (17). Our commentary of Theorem 2 thoroughly discusses these matters.

The next section deals with the understanding of SOMP performance whenever K increases. In particular, the AERC mentioned in the introduction will be derived.

6.3 Probability of failure for increasing values of K

To gain further insight into what Theorem 3 implies when K increases, we are using the quantities defined in the introduction, *i.e.*, $\mu_X(K) := \min_{j\in\mathcal{S}} \frac{1}{K}\sum_{k=1}^{K}|X_{j,k}|$, $\sigma(K)^2 = \frac{1}{K}\sum_{k=1}^{K}\sigma_k^2$, $\omega_\sigma = \max_{1\leq K<\infty}(1/\sqrt{K})\|\boldsymbol{\sigma}\|_1/\|\boldsymbol{\sigma}\|_2$ and $\mathrm{SNR}_{\min} := \min_{1\leq K<\infty}\frac{\mu_X(K)}{\sigma(K)}$. The definition of $\sigma(K)^2$ implies that $\|\boldsymbol{\sigma}\|_2^2 = K\sigma(K)^2$. Thus, using the expression of τ_X provided by Lemma 2, we obtain

$$\begin{aligned}\frac{\Delta\mathbb{E}}{\|\boldsymbol{\sigma}\|_2} &= \sqrt{K}\left(\left(1 - \frac{1}{\Gamma}\right)\psi\frac{\mu_X(K)}{\sigma(K)} - \sqrt{\frac{2}{K\pi}}\frac{\|\boldsymbol{\sigma}\|_1}{\|\boldsymbol{\sigma}\|_2}\right)\\ &\geq \sqrt{K}\left(\left(1 - \frac{1}{\Gamma}\right)\psi\mathrm{SNR}_{\min} - \sqrt{\frac{2}{\pi}}\omega_\sigma\right)\end{aligned}$$

Hence, ensuring $\xi > 0$ yields $(\Delta\mathbb{E})^2/\|\boldsymbol{\sigma}\|_2^2 \geq K\xi^2$ where

$$\xi := \left(1 - \frac{1}{\Gamma}\right)\psi\mathrm{SNR}_{\min} - \sqrt{\frac{2}{\pi}}\omega_\sigma. \tag{20}$$

As a consequence, under the condition $\xi > 0$, the probability that SOMP picks at least one incorrect atom during the first $s+1$ iterations is upper bounded by

$$n\mathcal{C}_s \exp\left[-\frac{1}{8}K\xi^2\right] \tag{21}$$

which indicates that the condition $\xi > 0$ is sufficient for asymptotic recovery, *i.e.*, it is a valid AERC. More precisely, the expression of ξ shows that $\mathrm{SNR}_{\min}$ should be sufficiently high to guarantee the asymptotic recovery. It is also worth noticing that, for $\mathrm{SNR}_{\min} \to \infty$, our result is equivalent to the exact recovery criterion (ERC) $1 - 1/\Gamma > 0$ in the noiseless case.

To conclude the technical discussions, we show that, as long as $\xi > 0$, there always exists a value of K ensuring an arbitrary maximum probability of failure for SOMP. Fixing the maximum probability of failure $p_{\mathrm{err}} > 0$, elementary algebraic operations show that satisfying $K > K_{\min}(p_{\mathrm{err}})$ yields a probability of failure inferior to p_{err} where

$$K_{\min}(p_{\mathrm{err}}) := \frac{8}{\xi^2}\log\left(\frac{n\mathcal{C}_s}{p_{\mathrm{err}}}\right). \tag{22}$$

The quantity ξ increases with $\mathrm{SNR}_{\min}$ so that the number of sparse signals K needed to achieve a prescribed maximum probability of failure decreases whenever the SNR improves.

Finally, according to Theorem 3 and for $s \geq 1$, we have $\mathcal{C}_s \leq \left(\frac{e(|\mathcal{S}|+s-1)}{s}\right)^s + 1 \lesssim \left(\frac{e(|\mathcal{S}|+s-1)}{s}\right)^s$ so that $\log \mathcal{C}_s \lesssim s \log e\frac{|\mathcal{S}|+s-1}{s} < s\log\frac{2e|\mathcal{S}|}{s}$. Equation (22) now yields

$$K_{\min}(p_{\mathrm{err}}) \lesssim \frac{1}{\xi^2}\left(\log n + s\log\frac{2e|\mathcal{S}|}{s} - \log p_{\mathrm{err}}\right),$$

which suggests that $K_{\min}$ should scale logarithmically with $1/p_{\mathrm{err}}$ and the number of atoms n and linearly with the number of iterations to be performed $s+1$.

7 Numerical results

7.1 Objective and methodology

Through numerical simulations, we wish to validate that our theoretical developments properly describe SOMP performance. To do so, we show that performing some adjustments in Equation (22) enables us to accurately predict the minimum number of sparse signals K needed to achieve a prescribed probability of error p_{err} when using SOMP to perform the full support recovery. More precisely, we consider three parameters α, β, and γ to be identified so that Equation (22) rewrites

$$K_{\min}(p_{\mathrm{err}}) := \frac{8}{\left(\alpha\ \mathrm{SNR}_{\min} - \omega_\sigma\beta\right)^2}\left(\gamma - \log p_{\mathrm{err}}\right). \tag{23}$$

The parameter $\alpha := (1 - 1/\Gamma)\psi$ is related to the relative reliability of SOMP in the noiseless case. In theory, the quantity β is equal to $\sqrt{2/\pi}$. However, the theoretical development provides

values that are not sharp in the general case (see our discussion in Section 6.1) and possibly even less for more constrained signal models. Finally, the parameter $\gamma = \log(n\mathcal{C}_{|\mathcal{S}|-1})$ conveys the impact of the number of atoms n and the number of iterations, *i.e.*, $|\mathcal{S}|$ in this case. In practice, only a few of the $n - |\mathcal{S}|$ incorrect atoms have a non-negligible probability to be picked so that n should be replaced by $\overline{n} \ll n - |\mathcal{S}|$. Also, as we discussed before, $\mathcal{C}_{|\mathcal{S}|-1}$ is a suboptimal term that results from the fact that we considered all the possible correct supports at each iteration to deduce the probability of error while, in practice, only one support out of the numerous possibilities matters. The identification procedure uses the cost function

$$\sum_{\substack{p_{\text{err}} \in \mathcal{I}_{\text{perr}} \\ \text{SNR}_{\min} \in \mathcal{I}_{\text{SNR}}}} \left[\sqrt{K_{\min}}(\alpha \text{SNR}_{\min} - \beta\omega_\sigma) - \sqrt{8(\gamma - \log p_{\text{err}})} \right]^2$$

where $K_{\min}$ is obtained by means of simulations, $\mathcal{I}_{\text{perr}} := \{0.05; 0.5; 0.9\}$ and $\mathcal{I}_{\text{SNR}}$ is to be specified afterwards. The cost function is evaluated for each 3-tuple $(\alpha, \beta, \gamma) \in \mathcal{I}_\alpha \times \mathcal{I}_\beta \times \mathcal{I}_\gamma$ where each set $\mathcal{I}_\alpha$, $\mathcal{I}_\beta$ and $\mathcal{I}_\gamma$ consists of 500 uniformly distributed points in the intervals $[0.1; 1.4]$, $[0; \sqrt{2/\pi}]$ and $[0; 5]$, respectively. The 3-tuple (α, β, γ) minimizing the cost function is then chosen.

7.2 Simulations signal model

We now describe the signal model we consider to run the simulations. It is more simple than the general model described in Section 1.1 so that $\text{SNR}_{\min}$ can be easily computed. First of all, the noise standard deviation vector $\boldsymbol{\sigma}$ has its odd-indexed (even-indexed) entries identical, *i.e.*, $\boldsymbol{\sigma} = (\sigma_{\text{odd}}, \sigma_{\text{even}}, \sigma_{\text{odd}}, \sigma_{\text{even}}, \dots)$ where we define $r_\sigma = \sigma_{\text{even}}/\sigma_{\text{odd}}$. It is then easy to show that $\omega_\sigma = \frac{1}{\sqrt{2}}(r_\sigma + 1)/\sqrt{r_\sigma^2 + 1}$ for K even and/or $r_\sigma = 1$. Furthermore, we assume $X_{j,k} = \varepsilon_{j,k}\mu_X$ $(j \in \mathcal{S}, k \in [K])$ where μ_X is fixed and the $\{\varepsilon_{j,k}\}_{j\in[n],k\in[K]}$ are statistically independent Rademacher variables, *i.e.*, random variables returning either -1 or 1 with probability 0.5 for both outcomes. Under this assumption, we have $\mu_X(K) = \mu_X$ for every K.

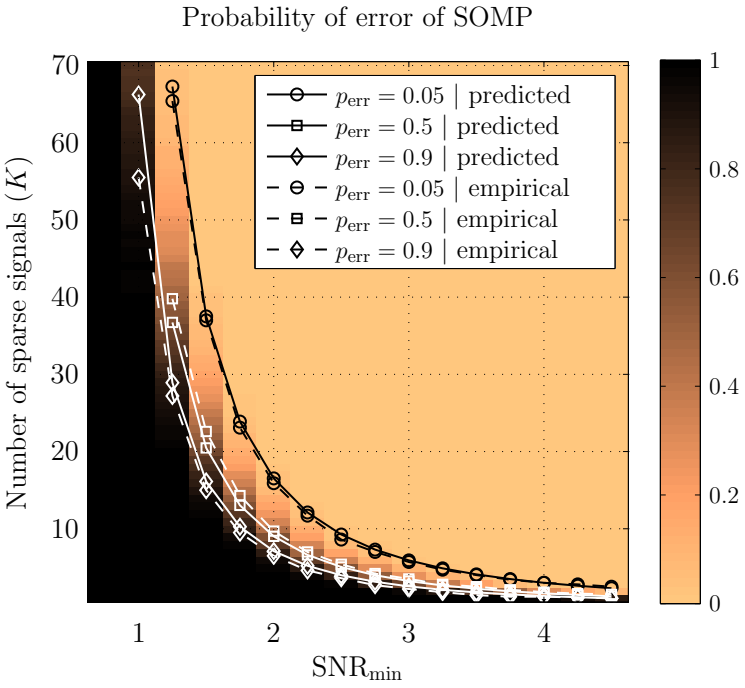
7.3 Simulation setup

The sensing matrix $\boldsymbol{\Phi}$ has been chosen of size 250×1000, *i.e.*, $m = 250$ and $n = 1000$, and its columns are the realizations of statistically independent vectors uniformly distributed on the unit hypersphere $\mathbb{S}^{m-1} := \{\boldsymbol{\phi} \in \mathbb{R}^m : \|\boldsymbol{\phi}\|_2 = 1\}$. The sensing matrix is identical for all the simulations that have been conducted. All the simulations will be such that $\sigma(K) = 1$. The reason is that it is always possible to recast a signal model for which $\sigma(K) = \zeta \neq 1$ as another one satisfying $\sigma(K) = 1$ while maintaining the value of $\text{SNR}_{\min}$ by multiplying $\boldsymbol{Y}$ by $1/\zeta$. This multiplication does not affect SOMP decisions as all the inner products intervening in SOMP decision metric are equally affected (see step 4 in Algorithm 1). Note that our results and our Matlab software are available in [11].

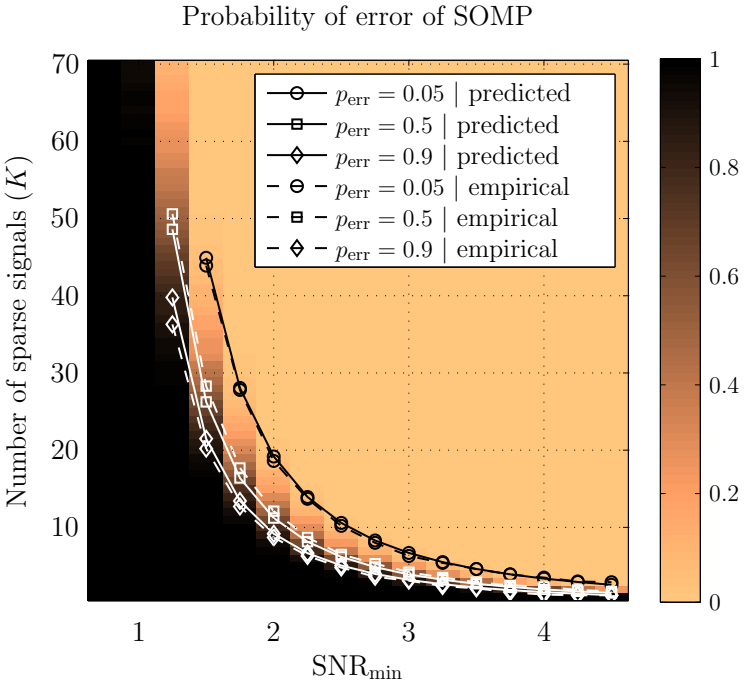
7.4 Results and analysis

We wish to determine whether Equation (23) accurately predicts SOMP performance provided that the parameters α, β, and γ are properly identified. To do so, we begin with Monte-Carlo simulations for the homoscedastic signal model, *i.e.*, for $r_\sigma = 1$, for two different support cardinalities, *i.e.*, $|\mathcal{S}| = 10$ and $|\mathcal{S}| = 20$. For each support cardinality, we identify a 3-tuple of parameters (α, β, γ) as explained in Section 7.1. The predictions based on the identified parameters are then compared against another set of Monte-Carlo simulations for which r_σ varies. For each Monte-Carlo simulation, the support is chosen at random and the random variables $\boldsymbol{X}$

and $\boldsymbol{E}$ are statistically independent. Figure 2 plots the results obtained for the homoscedastic signal model.



(a) Configuration with $|\mathcal{S}| = 10$ – Identified parameters (with $\mathcal{I}_{\text{SNR}} = \{1.25; 1.5; 1.75\}$, see Section 7.1): $\alpha = 1.0535$, $\beta = 0.54045$, and $\gamma = 2.0741$.



(b) Configuration with $|\mathcal{S}| = 20$ – Identified parameters (with $\mathcal{I}_{\text{SNR}} = \{1.5; 1.75; 2\}$, see Section 7.1): $\alpha = 1.0535$, $\beta = 0.58682$, and $\gamma = 2.5451$.

Figure 2: Probability of SOMP comitting at least one error when performing the joint full support recovery – The continuous curves plot the values of K_{min} predicted by Equation (23) for several values of p_{err} when using the identified parameters α, β, and γ. The dashed curves correspond to the numerical level sets for a fixed probability of failure p_{err} corresponding to that of the associated continuous curve. Note that the color of each curve has no particular meaning as its only purpose is constrast enhancement. The number of Monte-carlo cases is equal to 2000 for each 2-tuple $(\text{SNR}_{\text{min}}, K)$.

As observed in Figure 2, the fitting of Equation (23) to the numerical results is satisfactory over the full range of the simulations. The identification procedure has yielded values for α slightly higher than 1, which is incoherent with the theory as the value 1 is obtained whenever $\Gamma \to \infty$ and $\delta_{|\mathcal{S}|} = 0$. Also, the values of β are similar for both cardinalities of $\mathcal{S}$ and are lower than $\sqrt{2/\pi} \simeq 0.7979$, as predicted by the theory. The values of α and β remain high but the constant 8 in Equation (23) might not be sharp. In particular, if 8 is replaced by $8/\theta$, then Equation (23) provides identical values of $K_{\min}$ provided that α becomes $\alpha/\sqrt{\theta}$ and β is replaced by $\beta/\sqrt{\theta}$. The values obtained for γ are equal to 2.0741 and 2.5451, which indicates that $n\mathcal{C}_{|\mathcal{S}|-1}$ should be replaced by $\exp(2.0741) \simeq 7.9574$ and $\exp(2.5451) \simeq 12.7445$, respectively. The higher value of γ for $|\mathcal{S}| = 20$ is coherent with the theory as $\mathcal{C}_{|\mathcal{S}|-1}$ increases with $|\mathcal{S}|$. The values obtained for $n\mathcal{C}_{|\mathcal{S}|-1}$ also suggest that the expression $n\mathcal{C}_{|\mathcal{S}|-1}$ is not sharp.

Figure 3 plots the results obtained for $|\mathcal{S}| = 10$ when r_σ increases. While our theoretical developments appropriately predict that the value of $K_{\min}$ decreases as the noise standard deviation vector $\boldsymbol{\sigma}$ gets sparser, *i.e.*, as r_σ increases, our model is pessimistic with regard to the amplitude of the improvement of $K_{\min}$. We hypothesize that the inequalities used in Section 6.1 are not sharp enough to accurately predict $K_{\min}$ on the basis of r_σ. The identified parameters α, β, and γ might also not be optimal.

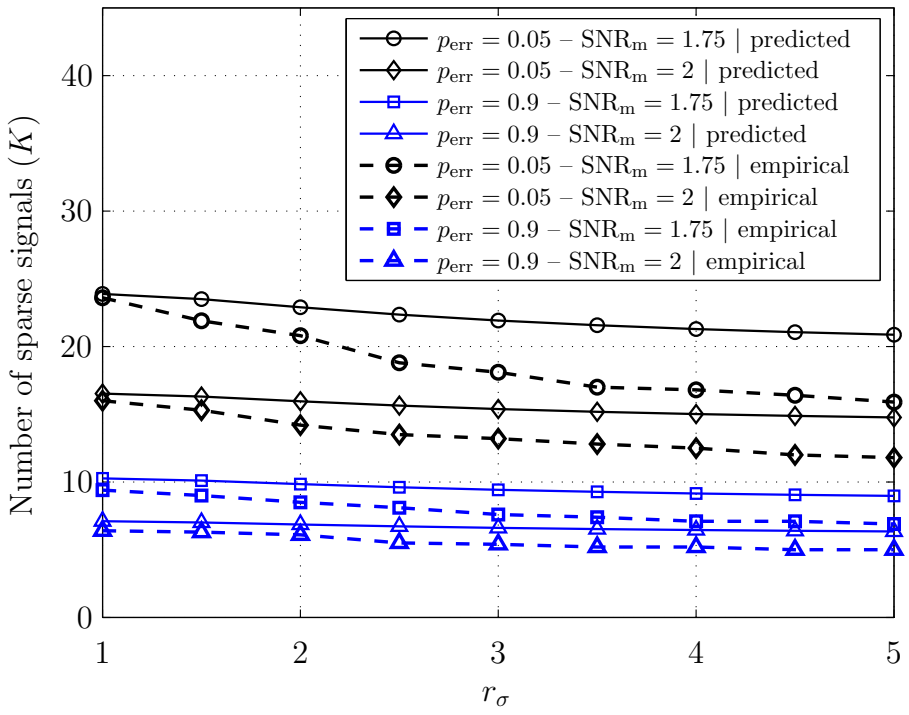


Figure 3: Levels sets of the probability of SOMP comitting at least one error when performing the joint full support recovery – $|\mathcal{S}| = 10$ – The continuous curves plot the values of $K_{\min}$ predicted by Equation (23) for several 2-tuples $(p_{\text{err}}, \text{SNR}_{\text{min}})$ when using the parameters identified in Figure 2a. The dashed curves correspond to the numerical level sets corresponding to that of the associated continuous curve. The number of Monte-carlo cases is equal to 2000 for each 2-tuple (r_σ, K). Only even values of K have been simulated.

8 Conclusion

In this paper, a theoretical analysis of SOMP operating in the presence of Gaussian additive noise has been presented. It has been shown that the signal to be recovered should be sufficiently large when compared to the mean noise power over all the measurement channels to succeed in the support recovery. Assuming this condition is met, the minimum number of sparse signals K to be gathered to achieve a prescribed probability of failure has been derived. An interesting corollary of the aforementioned results is that ensuring SNR values above a threshold allows asymptotic recovery, *i.e.*, the probability of error tends to 0 as K tends to infinity. Finally, numerical results confirmed the validity of the theoretical developments.

[Technical proofs]

.1 Lemma 2 example (Section 4.2)

Let $|\mathcal{S}| \geq 2$, $\eta \in \mathcal{S}$, $K = 1$, and $t = 0$. Let us consider $\chi := \max_{j \in \mathcal{S}\setminus\{\eta\}} |\langle \boldsymbol{\phi}_j, \boldsymbol{\Phi}_{\mathcal{S}} \boldsymbol{x}_{\mathcal{S}} \rangle| = \|\boldsymbol{\Phi}_{\mathcal{S}\setminus\{\eta\}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}} \boldsymbol{x}_{\mathcal{S}}\|_\infty$, which is the maximal value of SOMP metric for all the correct atoms except $\boldsymbol{\phi}_\eta$. As $\boldsymbol{\Phi}_{\mathcal{S}} \boldsymbol{x}_{\mathcal{S}} = \boldsymbol{\Phi}_{\mathcal{S}\setminus\{\eta\}} \boldsymbol{x}_{\mathcal{S}\setminus\{\eta\}} + \boldsymbol{\phi}_\eta x_\eta$, we easily show that

$$\boldsymbol{\Phi}_{\mathcal{S}\setminus\{\eta\}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}\setminus\{\eta\}} \boldsymbol{x}_{\mathcal{S}\setminus\{\eta\}} = -\boldsymbol{\Phi}_{\mathcal{S}\setminus\{\eta\}}^{\mathrm{T}} \boldsymbol{\phi}_\eta x_\eta$$

implies $\chi = 0$. Thus, enforcing $\chi = 0$ is possible by setting $\boldsymbol{x}_{\mathcal{S}\setminus\{\eta\}} = -(\boldsymbol{\Phi}_{\mathcal{S}\setminus\{\eta\}}^{\mathrm{T}} \boldsymbol{\Phi}_{\mathcal{S}\setminus\{\eta\}})^{-1} \boldsymbol{\Phi}_{\mathcal{S}\setminus\{\eta\}}^{\mathrm{T}} \boldsymbol{\phi}_\eta x_\eta = -\boldsymbol{\Phi}_{\mathcal{S}\setminus\{\eta\}}^{+} \boldsymbol{\phi}_\eta x_\eta$, which is a non-zero vector except for pathological cases. As a result, at iteration $t = 0$, the noiseless SOMP metric might be non-zero for a single correct atom only, *i.e.*, $\boldsymbol{\phi}_\eta$ in this particular example.

.2 Proof of Theorem 1

To simplify and shorten notations, we will abbreviate $f_j^{(t,\boldsymbol{P})} = f_j^{(t,\boldsymbol{P})}(\boldsymbol{g})$ in this proof. We will consider that only the correct atom whose index is $j_c^{(t,\boldsymbol{P})}$ has a chance to be picked. All the other correct atoms could be picked but the conservative analysis we present does not take into account this possibility. A sufficient condition to pick an atom belonging to $\mathcal{S}$ at iteration t given the projection matrix $\boldsymbol{P}$ is thus given by $\max_{j \in \overline{\mathcal{S}}} f_j^{(t,\boldsymbol{P})} < f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})}$ which is different from the condition $\max_{j \in \overline{\mathcal{S}}} f_j^{(t,\boldsymbol{P})} < \max_{j \in S} f_j^{(t,\boldsymbol{P})}$ where all the correct atoms could be picked. Denoting by $E_{\mathrm{succ}}^{(t,\boldsymbol{P})}$ the event occurring when SOMP picks a correct atom at iteration t given the projection matrix $\boldsymbol{P} \in \mathcal{P}^{(t)}$, we have

$$\mathbb{P}[E_{\mathrm{succ}}^{(t,\boldsymbol{P})}] \geq \mathbb{P}\left[f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})} > \max_{j \in \overline{\mathcal{S}}} f_j^{(t,\boldsymbol{P})} \right].$$

Let $\alpha > 0$, then the event $f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})} > \max_{j \in \overline{\mathcal{S}}} f_j^{(t,\boldsymbol{P})}$ is implied by $(\max_{j \in \overline{\mathcal{S}}} f_j^{(t,\boldsymbol{P})} < \alpha) \cap (f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})} > \alpha)$ so that a new lower bound is given by

$$\begin{aligned}
\mathbb{P}[E_{\mathrm{succ}}^{(t,\boldsymbol{P})}] &\geq \mathbb{P}[(\max_{j \in \overline{\mathcal{S}}} f_j^{(t,\boldsymbol{P})} < \alpha) \cap (f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})} > \alpha)] \\
&= 1 - \mathbb{P}[(\max_{j \in \overline{\mathcal{S}}} f_j^{(t,\boldsymbol{P})} \geq \alpha) \cup (f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})} \leq \alpha)]
\end{aligned}$$

where the second line is obtained by considering the complementary event of that of the previous line. Using the union bound yields

$$\mathbb{P}[E_{\mathrm{succ}}^{(t,\boldsymbol{P})}] \geq 1 - \left(\mathbb{P}\left[f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})} \leq \alpha \right] + \mathbb{P}\left[\max_{j \in \overline{\mathcal{S}}} f_j^{(t,\boldsymbol{P})} \geq \alpha \right] \right).$$

Since the event $\max_{j\in\overline{\mathcal{S}}} f_j^{(t,\boldsymbol{P})} \geq \alpha$ is equal to $\cup_{j\in\overline{\mathcal{S}}}\big[f_j^{(t,\boldsymbol{P})} \geq \alpha\big]$, using the union bound a second time yields

$$\mathbb{P}\big[\max_{j\in\overline{\mathcal{S}}} f_j^{(t,\boldsymbol{P})} \geq \alpha\big] \leq \sum_{j\in\overline{\mathcal{S}}} \mathbb{P}\big[f_j^{(t,\boldsymbol{P})} \geq \alpha\big].$$

Noticing that the probability of SOMP failing at iteration t is equal to $1 - \mathbb{P}[E_{\text{succ}}^{(t,\boldsymbol{P})}]$ concludes the proof.

.3 Proof of Lemma 2.1

We consider two arbitrary vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^K$. Using sequentially the triangle inequality, the reverse triangle inequality, and the CauchySchwarz inequality yields

$$\begin{aligned}
&\left|f_j^{(t,\boldsymbol{P})}(\boldsymbol{x}) - f_j^{(t,\boldsymbol{P})}(\boldsymbol{y})\right| \\
&= \left|\sum_{k=1}^{K} \left(\left|\beta_{j,k}^{(t,\boldsymbol{P})} + \sigma_{j,k}^{(\boldsymbol{P})} x_k\right| - \left|\beta_{j,k}^{(t,\boldsymbol{P})} + \sigma_{j,k}^{(\boldsymbol{P})} y_k\right|\right)\right| \\
&\leq \sum_{k=1}^{K} \left|\left|\beta_{j,k}^{(t,\boldsymbol{P})} + \sigma_{j,k}^{(\boldsymbol{P})} x_k\right| - \left|\beta_{j,k}^{(t,\boldsymbol{P})} + \sigma_{j,k}^{(\boldsymbol{P})} y_k\right|\right| \\
&\leq \sum_{k=1}^{K} \sigma_{j,k}^{(\boldsymbol{P})} |x_k - y_k| = \left\langle \boldsymbol{\sigma}_j^{(\boldsymbol{P})}, (|x_k - y_k|)_{k\in[K]} \right\rangle \\
&\leq \|\boldsymbol{\sigma}_j^{(\boldsymbol{P})}\|_2 \|(|x_k - y_k|)_{k\in[K]}\|_2 = \|\boldsymbol{\sigma}_j^{(\boldsymbol{P})}\|_2 \|\boldsymbol{x} - \boldsymbol{y}\|_2.
\end{aligned}$$

The quantity $\|\boldsymbol{\sigma}_j^{(\boldsymbol{P})}\|_2$ is thus a valid Lipschitz constant and it is also the best one since setting $\boldsymbol{x} = 1.5\, \boldsymbol{\sigma}_j^{(\boldsymbol{P})}$ and $\boldsymbol{y} = 0.5\, \boldsymbol{\sigma}_j^{(\boldsymbol{P})}$ saturates the inequality as $\beta_{j,k}^{(t,\boldsymbol{P})} \geq 0$ for all $k \in [K]$.

.4 Proof of Theorem 2

For better readability we will omit the dependence on $\boldsymbol{g}$ and abbreviate $f_j^{(t,\boldsymbol{P})} = f_j^{(t,\boldsymbol{P})}(\boldsymbol{g})$. Since $f_j^{(t,\boldsymbol{P})}$ is Lipschitz with a Lipschitz constant L equal to $\|\boldsymbol{\sigma}_j^{(\boldsymbol{P})}\|_2$, the concentration inequalities in Section 3.2 yield, for $\varepsilon > 0$,

$$\mathbb{P}\left(f_j^{(t,\boldsymbol{P})} \geq \mathbb{E}[f_j^{(t,\boldsymbol{P})}] + \varepsilon\right) \leq \exp\left(\frac{-\varepsilon^2}{2L^2}\right),$$

which rewrites (with $\alpha := \mathbb{E}[f_j^{(t,\boldsymbol{P})}] + \varepsilon$)

$$\mathbb{P}\left(f_j^{(t,\boldsymbol{P})} \geq \alpha\right) \leq \exp\left(-\frac{1}{2L^2}\left(\alpha - \mathbb{E}[f_j^{(t,\boldsymbol{P})}]\right)^2\right)$$

provided that $\alpha > \mathbb{E}[f_j^{(t,\boldsymbol{P})}]$. Similarly, using the concentration inequality obtained for $-f_j^{(t,\boldsymbol{P})}$ yields

$$\mathbb{P}\left(f_j^{(t,\boldsymbol{P})} \leq \alpha\right) \leq \exp\left(-\frac{1}{2L^2}\left(\mathbb{E}[f_j^{(t,\boldsymbol{P})}] - \alpha\right)^2\right)$$

where $\alpha := \mathbb{E}[f_j^{(t,\boldsymbol{P})}] - \varepsilon < \mathbb{E}[f_j^{(t,\boldsymbol{P})}]$.

As suggested by Figure 1, the value of α is chosen in between $\max_{j\in\overline{\mathcal{S}}}\mathbb{E}\big[f_j^{(t,\boldsymbol{P})}\big]$ and $\mathbb{E}\big[f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})}\big]$ where the assumption $\Delta\mathbb{E}^{(t,\boldsymbol{P})} > 0$ guarantees that the latter is strictly higher than the former. Consequently, if j corresponds to the best correct atom, then the second concentration inequality is used since $\mathbb{E}[f_j^{(t,\boldsymbol{P})}] > \alpha$. Conversely, the first concentration inequality will be used for incorrect atoms. We now consider the convex combination

$$\begin{aligned}\alpha &= \lambda\mathbb{E}\big[f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})}\big] + (1-\lambda)\max_{j\in\overline{\mathcal{S}}}\mathbb{E}\big[f_j^{(t,\boldsymbol{P})}\big] \\ &= (1-\overline{\lambda})\mathbb{E}\big[f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})}\big] + \overline{\lambda}\max_{j\in\overline{\mathcal{S}}}\mathbb{E}\big[f_j^{(t,\boldsymbol{P})}\big]\end{aligned}$$

where $\overline{\lambda} = 1-\lambda$. Thus, we obtain

$$\begin{aligned}\mathbb{P}\left(f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})} \leq \alpha\right) &\leq \exp\left(-\frac{\Big(\mathbb{E}[f_{j_c^{(t,\boldsymbol{P})}}^{(t,\boldsymbol{P})}] - \alpha\Big)^2}{2\|\boldsymbol{\sigma}_{j_c^{(t,\boldsymbol{P})}}^{(\boldsymbol{P})}\|_2^2}\right) \\ &= \exp\left(-\frac{\overline{\lambda}^2}{2\|\boldsymbol{\sigma}_{j_c^{(t,\boldsymbol{P})}}^{(\boldsymbol{P})}\|_2^2}\left(\Delta\mathbb{E}^{(t,\boldsymbol{P})}\right)^2\right).\end{aligned}$$

For $j \in \overline{\mathcal{S}}$, we obtain

$$\begin{aligned}\mathbb{P}\left(f_j^{(t,\boldsymbol{P})} \geq \alpha\right) &\leq \exp\left(-\frac{\Big(\alpha - \mathbb{E}[f_j^{(t,\boldsymbol{P})}]\Big)^2}{2\|\boldsymbol{\sigma}_j^{(\boldsymbol{P})}\|_2^2}\right) \\ &\leq \exp\left(-\frac{\Big(\alpha - \max_{\tilde{j}\in\overline{\mathcal{S}}}\mathbb{E}[f_{\tilde{j}}^{(t,\boldsymbol{P})}]\Big)^2}{2\|\boldsymbol{\sigma}_j^{(\boldsymbol{P})}\|_2^2}\right) \\ &= \exp\left(-\frac{\lambda^2}{2\|\boldsymbol{\sigma}_j^{(\boldsymbol{P})}\|_2^2}\left(\Delta\mathbb{E}^{(t,\boldsymbol{P})}\right)^2\right)\end{aligned}$$

where the second inequality holds because we have $\alpha > \max_{\tilde{j}\in\overline{\mathcal{S}}}\mathbb{E}[f_{\tilde{j}}^{(t,\boldsymbol{P})}] \geq \mathbb{E}[f_j^{(t,\boldsymbol{P})}]$ for every $j \in \overline{\mathcal{S}}$. A less sharp upper bound is obtained for both inequalities by remembering that $\|\boldsymbol{\sigma}_j^{(\boldsymbol{P})}\|_2 \leq \|\boldsymbol{\sigma}\|_2$. Setting $\lambda = \overline{\lambda} = 0.5$ and combining Theorem 1 with the derived inequalities for $j = j_c^{(t,\boldsymbol{P})}$ and $j \in \overline{\mathcal{S}}$ conclude the proof.

.5 Proof of Theorem 3

The event $E_{\text{succ}}^{(t,\boldsymbol{P})}$ occurs whenever SOMP picks a correct atom at iteration t given the projection matrix $\boldsymbol{P} \in \mathcal{P}^{(t)}$. Considering all the possible orthogonal projectors $\boldsymbol{P} \in \mathcal{P}^{(t)}$ from iteration 0 to iteration s included, we have $\mathcal{C}_s = \sum_{t=0}^{s}\binom{|\mathcal{S}|}{t}$ possible orthogonal projectors. If SOMP succeeds in choosing a correct atom for all the possible orthogonal projectors at each iteration, then we know that correct decisions will occur at each iteration. Thus, defining $E_{\text{succ}}^{(s)}$ as the event occurring whenever SOMP is successful during the first $s+1$ iterations, we have

$$\mathbb{P}[E_{\text{succ}}^{(s)}] \geq \mathbb{P}\left[\bigcap_{t=0}^{s}\bigcap_{\boldsymbol{P}\in\mathcal{P}^{(t)}} E_{\text{succ}}^{(t,\boldsymbol{P})}\right] = 1 - \mathbb{P}\left[\bigcup_{t=0}^{s}\bigcup_{\boldsymbol{P}\in\mathcal{P}^{(t)}} E_{\text{fail}}^{(t,\boldsymbol{P})}\right]$$

where the event intervening in the R.H.S. is the complementary event of that of the preceding expression. In particular, $E_{\text{fail}}^{(t,\boldsymbol{P})}$ is the event occurring when SOMP picks an incorrect atom at iteration t given the orthogonal projector $\boldsymbol{P} \in \mathcal{P}^{(t)}$. The union bound yields

$$\mathbb{P}\left[\bigcup_{t=0}^{s} \bigcup_{\boldsymbol{P}\in\mathcal{P}^{(t)}} E_{\text{fail}}^{(t,\boldsymbol{P})}\right] \leq \sum_{t=0}^{s} \sum_{\boldsymbol{P}\in\mathcal{P}^{(t)}} \mathbb{P}[E_{\text{fail}}^{(t,\boldsymbol{P})}].$$

Therefore, if $E_{\text{fail}}^{(s)}$ is the probability of failure of SOMP during the first $s+1$ iterations, we have $\mathbb{P}[E_{\text{fail}}^{(s)}] = 1 - \mathbb{P}[E_{\text{succ}}^{(s)}] \leq \sum_{t=0}^{s} \sum_{\boldsymbol{P}\in\mathcal{P}^{(t)}} \mathbb{P}[E_{\text{fail}}^{(t,\boldsymbol{P})}]$ where Theorem 2 yields

$$\mathbb{P}\left[E_{\text{fail}}^{(t,\boldsymbol{P})}\right] \leq (n - |\mathcal{S}| + 1) \exp\left[-\frac{1}{8\|\boldsymbol{\sigma}\|_2^2}(\Delta\mathbb{E}^{(t,\boldsymbol{P})})^2\right].$$

The lower bound for $\Delta\mathbb{E}^{(t,\boldsymbol{P})}$ derived in Section 6.1 provides

$$\Delta\mathbb{E}^{(t,\boldsymbol{P})} \geq \left(1 - \frac{1}{\Gamma}\right)\psi\tau_X - \sqrt{\frac{2}{\pi}}\|\boldsymbol{\sigma}\|_1.$$

Using the inequality $n - |\mathcal{S}| + 1 \leq n$, noticing that the upper bound of $\mathbb{P}[E_{\text{fail}}^{(s)}]$ has become independent of t and $\boldsymbol{P}$ and remembering that $|\mathcal{P}^{(t)}| = \binom{|\mathcal{S}|}{t}$ conclude the first part of the proof. Regarding the inequality on $\mathcal{C}_s$, we have [22, Appendix A], $\sum_{t=1}^{s}\binom{|\mathcal{S}|}{t} \leq \binom{|\mathcal{S}|+s-1}{s}$ as well as $\binom{|\mathcal{S}|}{t} \leq \left(\frac{e|\mathcal{S}|}{t}\right)^t$. Thus, we easily obtain

$$\mathcal{C}_s = 1 + \sum_{t=1}^{s}\binom{|\mathcal{S}|}{t} \leq 1 + \left(\frac{e(|\mathcal{S}| + s - 1)}{s}\right)^s.$$

References

[1] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," *Rice University Technical Report*, no. TREE-0612, Nov. 2006.

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

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

[4] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[5] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[6] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, 2006.

[7] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4634–4643, 2006.

[8] W. Dan and R. Wang, "Robustness of orthogonal matching pursuit under restricted isometry property," *Science China Mathematics*, vol. 57, no. 3, pp. 627–634, 2014.

[9] M. Davenport, M. B. Wakin *et al.*, "Analysis of orthogonal matching pursuit using the restricted isometry property," *Information Theory, IEEE Transactions on*, vol. 56, no. 9, pp. 4395–4401, 2010.

[10] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[11] J.-F. Determe, "Somp simulation framework v0.5.3." [Online]. Available: http://homepages.ulb.ac.be/~jdeterme/software.html

[12] J.-F. Determe, J. Louveaux, L. Jacques, and F. Horlin, "Improving the correlation lower bound for simultaneous orthogonal matching pursuit," *Signal Processing Letters, IEEE*, in press.

[13] ——, "On the exact recovery condition of simultaneous orthogonal matching pursuit," *Signal Processing Letters, IEEE*, vol. 23, no. 1, pp. 164–168, 2016.

[14] J. Ding, L. Chen, and Y. Gu, "Robustness of orthogonal matching pursuit for multiple measurement vectors in noisy scenario," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3813–3816.

[15] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[16] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proceedings of the National Academy of Sciences*, vol. 100, no. 5, pp. 2197–2202, 2003.

[17] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *Information Theory, IEEE Transactions on*, vol. 55, no. 11, pp. 5302–5316, 2009.

[18] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 505–519, 2010.

[19] A. K. Fletcher and S. Rangan, "Orthogonal matching pursuit: A brownian motion analysis," *Signal Processing, IEEE Transactions on*, vol. 60, no. 3, pp. 1010–1021, 2012.

[20] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.

[21] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst, "Atoms of all channels, unite! average case analysis of multi-channel sparse recovery using greedy algorithms," *Journal of Fourier analysis and Applications*, vol. 14, no. 5-6, pp. 655–687, 2008.

[22] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *Information Theory, IEEE Transactions on*, vol. 59, no. 4, pp. 2082–2102, 2013.

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

[24] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

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

[26] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.

[27] A. Rakotomamonjy, "Surveying and comparing simultaneous sparse approximation (or group-lasso) algorithms," *Signal processing*, vol. 91, no. 7, pp. 1505–1526, 2011.

[28] S. Rangan and A. K. Fletcher, "Orthogonal matching pursuit from noisy random measurements: A new analysis," in *Advances in Neural Information Processing Systems*, 2009, pp. 540–548.

[29] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 3, pp. 1030–1051, 2006.

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

[31] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[32] R. Wu, W. Huang, and D.-R. Chen, "The exact support recovery of sparse signals with noise via orthogonal matching pursuit," *Signal Processing Letters, IEEE*, vol. 20, no. 4, pp. 403–406, 2013.

[33] W. Xu, Z. Li, Y. Tian, Y. Wang, and J. Lin, "Perturbation analysis of simultaneous orthogonal matching pursuit," *Signal Processing*, vol. 116, pp. 91–100, 2015.

[34] T. Zhang, "Sparse recovery with orthogonal matching pursuit under rip," *Information Theory, IEEE Transactions on*, vol. 57, no. 9, pp. 6215–6221, 2011.