

Sparsity estimation from compressive projections via sparse random matrices

*Original*

Sparsity estimation from compressive projections via sparse random matrices / Ravazzi, Chiara; Fosson, Sophie; Bianchi, Tiziano; Magli, Enrico. - In: EURASIP JOURNAL ON ADVANCES IN SIGNAL PROCESSING. - ISSN 1687-6180. - ELETTRONICO. - 2018:1(2018), pp. 1-18. [10.1186/s13634-018-0578-0]

*Availability:*

This version is available at: 11583/2712583 since: 2018-09-11T09:40:01Z

*Publisher:*

SpringerOpen

*Published*

DOI:10.1186/s13634-018-0578-0

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

RESEARCH

Open Access



# Sparsity estimation from compressive projections via sparse random matrices

Chiara Ravazzi<sup>1\*</sup>, Sophie Fosson<sup>2</sup>, Tiziano Bianchi<sup>3</sup>  and Enrico Magli<sup>3</sup>

## Abstract

The aim of this paper is to develop strategies to estimate the sparsity degree of a signal from compressive projections, without the burden of recovery. We consider both the noise-free and the noisy settings, and we show how to extend the proposed framework to the case of non-exactly sparse signals. The proposed method employs  $\gamma$ -sparsified random matrices and is based on a maximum likelihood (ML) approach, exploiting the property that the acquired measurements are distributed according to a mixture model whose parameters depend on the signal sparsity. In the presence of noise, given the complexity of ML estimation, the probability model is approximated with a two-component Gaussian mixture (2-GMM), which can be easily learned via expectation-maximization.

Besides the design of the method, this paper makes two novel contributions. First, in the absence of noise, sufficient conditions on the number of measurements are provided for almost sure exact estimation in different regimes of behavior, defined by the scaling of the measurements sparsity  $\gamma$  and the signal sparsity. In the presence of noise, our second contribution is to prove that the 2-GMM approximation is accurate in the large system limit for a proper choice of  $\gamma$  parameter. Simulations validate our predictions and show that the proposed algorithms outperform the state-of-the-art methods for sparsity estimation. Finally, the estimation strategy is applied to non-exactly sparse signals. The results are very encouraging, suggesting further extension to more general frameworks.

**Keywords:** Sparsity recovery, Compressed sensing, High-dimensional statistical inference, Gaussian mixture models, Maximum likelihood, Sparse random matrices

## 1 Introduction

Compressed sensing (CS) [1, 2] is a novel signal acquisition technique that recovers an unknown signal from a small set of linear measurements. According to CS, if a signal having dimension  $n$  is known to be sparse, i.e., it has only  $k \ll n$  non-zero entries when represented by a suitable basis, then it can be efficiently recovered using only  $m \ll n$  linear combinations of the signal entries, provided that these linear projections are sufficiently incoherent with respect to the signal basis.

In most of CS applications, it is usually assumed that an upper bound on the sparsity degree  $k$  is known before acquiring the signal. However, some signals may have a time-varying sparsity, as in spectrum sensing [3], or spatially varying sparsity, as in the case of block-based image acquisition [4]. Since the number of linear measurements

required for the recovery depends on the sparsity degree of the signal [5], the knowledge of  $k$  is crucial to fully exploit the potential of CS.

In many recovery algorithms, the optimal tuning of parameters requires the knowledge of the degree of sparsity of the signal. For example, in Lasso techniques [6], a parameter  $\lambda$  related to  $k$  has to be chosen [7], whereas for greedy algorithms, such as orthogonal matching pursuit (OMP) [8] or compressive sampling matching pursuit (CoSaMP) [9], the performance and the number of iterations depend on  $k$ .

The ability to estimate the signal sparsity degree directly from a small number of linear measurements can represent an important tool in several promising applications. One of the most obvious applications is the possibility to dynamically adapt the number of measurements acquired by a CS instrument, e.g., an imager, to the estimated signal sparsity. We can envisage a system that acquires linear measurements in a sequential way and continuously updates the estimated sparsity according to the new

\*Correspondence: [chiara.ravazzi@ieiit.cnr.it](mailto:chiara.ravazzi@ieiit.cnr.it)

<sup>1</sup> National Research Council of Italy, IEIIT-CNR, c/o Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy

Full list of author information is available at the end of the article

measurements. The acquisition can stop as soon as the number of acquired measurements is enough to guarantee the correct reconstruction of a signal based on the estimated sparsity.

Other applications may include the possibility of comparing the support of two sparse signals from their measurements. Due to the linearity of the sparse signal model, the degree of overlap between the supports of two sparse signal can be estimated by measuring the sparsity degree of their sum (or difference) [10]. Finally, sparsity estimation can be used to decide whether a signal can be represented in a sparse way according to a specific basis, which can be used to select the most suitable basis allowing the sparsest representation.

### 1.1 Related work

The problem of estimating the sparsity degree has begun to be recognized as a major gap between theory and practice [11–13], and the literature on the subject is very recent.

In some papers, the joint problem of signal reconstruction and sparsity degree estimation is investigated, in particular for time-varying settings. The following iterative approach is considered: given an initial upper bound for the sparsity degree, at a generic time step  $t$ , the signal is reconstructed and sparsity degree is estimated; such estimation is then used at time  $t + 1$  to assess the number of measurements sufficient for reconstruction. The seminal work [14] investigates the problem in the framework of spectrum sensing for cognitive radios and proposes an iterative method that at each time step performs two operations: (a) the signal is recovered via Lasso, and (b) the sparsity degree is estimated as the number of recovery components with magnitude larger than an empirically set threshold. The efficiency of this procedure is validated via numerical simulations.

Some authors propose sequential acquisition techniques in which the number of measurements is dynamically adapted until a satisfactory reconstruction performance is achieved [15–19]. Even if the reconstruction can take into account the previously recovered signal, these methods require to solve a minimization problem at each newly acquired measurement and may prove too complex when the underlying signal is not sparse, or if one is only interested in assessing the sparsity degree of a signal under a certain basis without reconstructing it.

In other papers, the sparsity degree estimation is only considered, which generally requires less measurements than signal reconstruction. In [13], sparsity degree is estimated through an eigenvalue-based method, for wide-band cognitive radios applications. In this work, the signal reconstruction is not required, while in practice, the used number of measurements was quite large. In [20], the sparsity of the signal is lower-bounded through

the numerical sparsity, i.e., the ratio between the  $\ell_1$  and  $\ell_2$  norms of the signal, where these quantities can be estimated from random projections obtained using Cauchy-distributed and Gaussian-distributed matrices, respectively. A limitation of this approach is that it is not suitable for adaptive acquisition since measurements taken with Cauchy-distributed matrices cannot be used later for signal reconstruction. In [21], this approach is extended to a family of entropy-based sparsity measures of kind  $(\|x\|_q/\|x\|_1)^{q/(1-q)}$  with  $q \in [0, 2]$ , for which estimators are designed and theoretically estimated in terms of limiting distributions. In [22], the authors propose to estimate the sparsity of an image before its acquisition, by calculating the image complexity. However, the proposed method is based on the image pixel values and needs a separate estimation that does not depend on the measurements. Further, in [23], the minimum number of measurements to recovery, the sparsity degree was theoretically investigated.

Finally, we notice that the problem of estimating the sparsity degree of a vector is partially connected to the problem of estimating the number of distinct elements in data streams [24, 25], which has been largely studied in the last decades due to its diverse applications. The analogy lies in the fact that the sparsity degree problem could be seen as the estimation of the number of elements distinct from zero. Moreover, many efficient algorithms to estimate the number of distinct elements are based on random hashing (see [25] for a review) to reduce the storage space, which is our concern as well. However, the problem of distinct elements considers vectors  $a = (a_1, \dots, a_n)$  with  $a_i \in Q$ , where  $Q$  is a finite set, which is intrinsically different from our model where the signal  $x$  has real-valued components. Therefore, the strategies conceived for this problem cannot be applied for our purpose.

### 1.2 Our contribution

In this paper, we propose a technique for directly estimating the sparsity degree of a signal from its linear measurements, without recovery. The method relies on the fact that measurements obtained by projecting the signal according to a  $\gamma$ -sparsified random matrix are distributed according to a mixture density whose parameters depend on  $k$ . This is an extension of the algorithm in [26], which works only in the case of noise-free, exactly  $k$ -sparse signals. First, we analyze the case of noise-free, exactly  $k$ -sparse signals as a special case, and we provide theoretical guarantees regarding the consistency of the proposed estimator and its asymptotic behavior under different regimes of the parameters  $k$  and  $\gamma$ . Then, we analyze the more generic case of noise, including the non-exactly sparse signals, and we propose to approximate the measurement model by a two-component Gaussian mixture model

(2-GMM), whose parameters can be easily estimated via expectation-maximization (EM) techniques. In this case, we prove that there is a regime of behavior, defined by the scaling of the measurement sparsity  $\gamma$  and the sparsity degree  $k$ , where this approximation is accurate. An interesting property of the proposed method is that measurements acquired using a  $\gamma$ -sparsified random matrix also enable signal reconstruction, with only a slight performance degradation with respect to dense matrices [27, 28].

Some preliminary results, limited to the sparsity estimation of noisy, exactly  $k$ -sparse ternary signals, have appeared in [29]. In this paper, we extend the results in [29] from both a theoretical and a practical point of view, by considering any  $k$ -sparse signal and extending the model to non-exactly sparse signals.

### 1.3 Outline of the paper

The paper is organized as follows. Section 2 presents the notation and a brief review of CS-related results. The sparsity estimation problem is formally introduced in Section 3, where we discuss the optimal estimator, whereas the main properties of the optimal estimator in the noise-free setting are outlined in Section 4. In Section 5, we introduce the proposed iterative algorithm for dealing with the noisy setting, together with some approximate performance bounds. Finally, the proposed estimators are experimentally validated in Section 6, while concluding remarks are given in Section 7.

## 2 Preliminaries

In this section, we define some notation, we review the CS fundamentals, and we briefly discuss the use of sparsified matrices in the CS literature.

### 2.1 Notation

Throughout this paper, we use the following notation. We denote column vectors with small letters, and matrices with capital letters. If  $x \in \mathbb{R}^n$ , we denote its  $j$ th element with  $x_j$  and, given  $S \subseteq [n] := \{1, \dots, n\}$ , by  $x|_S$ , the sub-vector of  $x$  corresponding to the indices in  $S$ . The support set of  $x$  is defined by  $\text{supp}(x) = \{i \in [n] : x_i \neq 0\}$  and we use  $\|x\|_0 = |\text{supp}(x)|$ . Finally, the symbol  $\|x\|$  with no subscript has always to be intended as the Euclidean norm of the vector  $x$ .

This paper makes frequent use of the following notation for asymptotics of real sequences  $(a_n)_{n \in \mathbb{N}}$  and  $(b_n)_{n \in \mathbb{N}}$ : (i)  $a_n = O(b_n)$  for  $n \rightarrow \infty$  if there exists a positive constant  $c \in (0, +\infty)$  and  $n_0 \in \mathbb{N}$  such that  $a_n \leq cb_n$  for all  $n > n_0$ , (ii)  $a_n = \Omega(b_n)$  for  $n \rightarrow \infty$  if there exists a constant  $c' \in (0, +\infty)$  and  $n_1 \in \mathbb{N}$  such that  $a_n \geq c'b_n$  for all  $n > n_0$ , (iii)  $a_n = \Theta(b_n)$  for  $n \rightarrow \infty$  if  $a_n = O(b_n)$  and  $a_n = \Omega(b_n)$ , and (iii)  $a_n = o(b_n)$  for  $n \rightarrow \infty$  means that  $\lim_{n \rightarrow \infty} |a_n/b_n| = 0$ .

Given a random variable, we denote the probability density function with  $f$ .

### 2.2 Sparse signal recovery using sparse random projections

Let  $x \in \mathbb{R}^n$  be an unknown deterministic signal. CS [30] aims to recover a signal from a small number of non-adaptive linear measurements of the form

$$y = Ax + \eta, \quad (1)$$

where  $y \in \mathbb{R}^m$  is a vector of observations,  $A \in \mathbb{R}^{m \times n}$  is the sensing matrix with  $m < n$ ,  $\eta \in \mathbb{R}^m$  is an additive Gaussian noise  $\mathcal{N}(0, \sigma^2 I_{m \times m})$ , and  $I_{m \times m}$  is the identity matrix with  $m$  rows, and  $m$  is the columns. Since the solution to (1) is not unique, the signal is typically assumed to be sparse, i.e., it can be represented with  $k$  non-zero coefficients, or compressible, in the sense that it can be well approximated by a vector having only  $k$  non-zero coefficients. In the following, we refer to  $k$  as the signal sparsity degree and we denote the set of signals with exactly  $k$  non-zero components as  $\Sigma_k = \{v \in \mathbb{R}^n : \|v\|_0 \leq k\}$ .

The literature describes a wide variety of approaches to select the sparsest solution to the affine system in (1). In particular, a large amount of work in CS investigates the performance of  $\ell_1$  relaxation for sparse approximation.

The problem of recovery can be analyzed in deterministic settings, where the measurement matrix  $A$  is fixed, or in random settings in which  $A$  is drawn randomly from a sub-Gaussian ensemble. Past work on random designs has focused on matrices drawn from ensemble of dense matrices, i.e., each row of  $A$  has  $n$  non-zero entries with high probability. However, in various applications, sparse sensing matrices are more desirable [31]. Furthermore, sparse measurement matrices require significantly less storage space, and algorithms adapted to such matrices have lower computational complexity [32, 33]. In [27], the authors study what sparsity degree is permitted in the sensing matrices without increasing the number of observations required for support recovery.

In this paper, we consider  $\gamma$ -sparsified matrices [27], in which the entries of the matrix  $A$  are independently and identically distributed according to

$$A_{ij} \sim \begin{cases} \mathcal{N}\left(0, \frac{1}{\gamma}\right) & \text{w.p. } \gamma, \\ \delta_0 & \text{w.p. } 1 - \gamma \end{cases} \quad (2)$$

where  $\delta_0$  denotes a Dirac delta centered at zero.

Since weak signal entries could be confused with noise, in [27], the support recovery is studied also as a function of the minimum (in magnitude) non-zero value of  $x$ :

$$\lambda := \min_{i \in \text{supp}(x)} |x_i|. \quad (3)$$

Consequently, for a fixed  $\lambda > 0$ , let us define:

$$\mathcal{X}_k(\lambda) = \{x \in \Sigma_k : |x_i| \geq \lambda\}. \quad (4)$$

For this class of signals, the following result has been proved.

**Theorem 1** (Corollary 2 in [27]) *Let the measurements matrix  $A \in \mathbb{R}^m$  be drawn with i.i.d. elements from the  $\gamma$ -sparsified Gaussian ensemble. Then, a necessary condition for asymptotically reliable recovery over the signal class  $\mathcal{X}_k(\lambda)$  is*

$$m \geq \max\{g_1(n, k, \lambda, \gamma), g_2(n, k, \lambda, \gamma)\} \quad (5)$$

where

$$g_1(n, k, \lambda, \gamma) \geq \frac{\log \binom{n}{k} - 1}{\frac{1}{2}\gamma k \log(1 + \lambda^2 k / \sigma^2) + \frac{1}{2} \log(2\pi e (\gamma k + \frac{1}{12}))} \quad (6)$$

$$g_2(n, k, \lambda, \gamma) \geq \frac{\log(n - k + 1) - 1}{\frac{1}{2}\gamma \log(1 + \lambda^2 / (\sigma^2 \gamma))}. \quad (7)$$

In particular, Theorem 1 says that if  $\gamma k \rightarrow \infty$  as  $n \rightarrow \infty$ , then the number of measurements is of the same order as that for dense sensing matrices. In sharp contrast, if  $\gamma k \rightarrow 0$  sufficiently fast as  $n \rightarrow \infty$ , then the number of measurements of any decoder increases dramatically. Finally, if  $\gamma k = \Theta(1)$  and  $\lambda^2 k = \Theta(1)$ , then at least  $\max\{\Theta(k \log(n/k)), \Theta(k \log(n-k)/\log k)\}$  measurements are necessary for estimating the support of the signal.

Several recovery algorithms are based on the use of sparse sensing matrices. In particular, count-minute sketch algorithms need about 10 to 15 times more measurements than  $\ell_1$ -decoding and sparse matching pursuit needs about half of the measurements of count-min sketch. Other sketch algorithms include [34] that can be as accurate as  $\ell_1$  decoding with dense matrices under the condition  $\gamma k = \Theta(1)$  with the same order of measurements.

### 3 Sparsity estimation problem: mathematical formulation

Our goal is to estimate  $k$  from the measurements  $y = Ax$ , where  $A$  is a  $\gamma$ -sparsified matrix, without the burden of reconstructing  $x$ . Specifically, we aim at providing conditions on the triplet  $(n, m, k)$  as well as on  $x$  and  $A$  under which the estimation of signal sparsity is accurate with high probability. The theoretical results that we provide also hold true for high dimensional settings, that is,  $(n, m, k)$ , and are allowed to tend to infinity.

Given a rule for computing estimates of the signal sparsity, we will measure the error between the estimate  $\hat{k}(m, n)$  and the true sparsity degree  $k$  using the relative error:

$$e(k, \hat{k}) := |k - \hat{k}|/k. \quad (8)$$

We say that the sparsity estimator  $\hat{k}$  is asymptotically *weakly consistent* when  $e(k, \hat{k})$  converges in probability to 0 as  $m, n \rightarrow \infty$ . If we replace convergence in probability with almost sure convergence, then the estimator is said to be *strongly consistent*.

If the signals are not exactly sparse but compressible, i.e., they admit a representation with few large components in magnitude; in CS literature, the recovery guarantees are expressed in terms of the sparsity degree of the best- $k$  approximation [30, 35] defined as follows

$$\hat{x}_k = \operatorname{argmin}_{z \in \Sigma_k} \|x - z\|. \quad (9)$$

For this reason, the sparsity of a not exactly sparse signal is defined as the number of components containing most of the energy up to a relative error  $\tau$

$$k_\tau = \min\{s \in [n] : \|x - \hat{x}_s\|^2 \leq \tau \|x\|^2\}. \quad (10)$$

Then, defining  $e = x - \hat{x}_{k_\tau}$ , we write

$$y = Ax = A(\hat{x}_{k_\tau} + e) = A\hat{x}_{k_\tau} + \eta \quad (11)$$

where  $\eta = Ae$ . It should be noticed that each component  $\eta_i$  is distributed as a mixture of Gaussians. We make the following approximation:  $\eta_i \sim N(0, \sigma^2)$  with

$$\sigma^2 = \mathbb{E}[\eta_i^2] = \|e\|^2 \leq \tau \|x\|^2. \quad (12)$$

In the noiseless case, by linearity of expectation, we have

$$\begin{aligned} \mathbb{E}\left[\frac{1}{m} \sum_{i=1}^m y_i^2\right] &= \frac{1}{m} \sum_{i=1}^m \mathbb{E}[y_i^2] \\ &= \frac{1}{m} \sum_{i=1}^m \mathbb{E}\left[\sum_{j=1}^n A_{ij}^2 x_j^2 + 2 \sum_{j < k} A_{ij} A_{ik} x_j x_k\right] \end{aligned}$$

from which

$$\begin{aligned} \frac{1}{m} \sum_{i=1}^m \mathbb{E}[y_i^2] &= \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n \mathbb{E}[A_{ij}^2] x_j^2 \\ &\quad + 2 \sum_{j < k} \mathbb{E}[A_{ij}] \mathbb{E}[A_{ik}] x_j x_k = \|x\|^2 \end{aligned}$$

where the last inequality follows from  $\mathbb{E}[A_{ij}] = 0$  and  $\mathbb{E}[A_{ij}^2] = 1$  for all  $i, j$ . Then, the model describing the measurements can be approximated by (1) with  $\sigma^2 \approx \tau \|y\|^2/m$ . We underline that this argument is true for all sensing matrices drawn from the ensemble in (2), and at this time, we do not make any additional assumption on the number of measurements.

Given  $(y, A)$  and assuming that the perturbation is additive Gaussian, the ML estimator of the signal sparsity can be obtained via the following exhaustive search:

$$\operatorname{argmin}_{s \in [n]} \left\{ \min_{T \subseteq [n]: |T|=s, \{x: \operatorname{supp}(x)=T\}} \|Ax - y\|_2^2 \right\}. \quad (13)$$



However, this optimization problem is NP hard and the search of the solution requires an exponential time in the signal length  $n$  (one optimization problem for all subsets of  $[n]$  of size  $s$  and for all  $s$ , which amounts to  $\sum_{s=1}^n \binom{n}{s} = 2^n - 1$ ).

Given  $\text{supp}(x)$ , if  $A$  is chosen from the ensemble of  $\gamma$ -sparsified matrices, any measurement

$$y_i = \sum_{j \in [n]} A_{ij} x_j + \eta_i \quad (14)$$

is a random variable whose density function is a mixture of Gaussians with  $2^k$  components. The result follows from the following argument. Let  $S$  be the overlap between the support of  $i$ th row of  $A$  and  $\text{supp}(x)$ . It is easy to see that given  $S$  then  $y_i \sim \mathcal{N}(0, \alpha_S)$  where  $\alpha_S = \|x_S\|^2 / \gamma + \sigma^2$ . Without any further assumption, taking into account all possible overlaps between the support of the  $i$ th row of  $A$  and support of the signal  $x$  with cardinality  $s \leq k$ , we can have in principle  $\sum_{s \leq k} \binom{k}{s} = 2^k$  different type of Gaussians. We conclude that  $y_i$  is a Gaussian mixture with  $2^k$  components. If the non-zero elements of the signal have all equal values in magnitude, then the number of components of the Gaussian mixture reduces dramatically to  $k$ . Given the set of  $m$  independent and identically distributed samples  $y = (y_1, \dots, y_m)^\top$ , the sparsity estimation can be recast into the problem of evaluating the number of mixture components and parameters. However, also in the simple case where  $k$  is known, the estimation of the finite mixture density function does not admit a closed-form solution, and the computational complexity is practically unfeasible.

#### 4 Method: noise-free setting

In this section, we show that in the absence of noise (i.e.,  $\|\eta\| = 0$ ),  $\|y\|_0$  is a sufficient statistic for the underlying parameter  $k$ . We show that the performance of the proposed estimators of the sparsity degree depends on the SNR  $= \lambda^2 k / \sigma^2$  and that the traditional measure  $\|x\|^2 / \sigma^2$  has no significant effect in the estimation of the sparsity degree.

Even in the absence of noise, since  $A$  is chosen from the ensemble of  $\gamma$ -sparsified matrices, any measurement  $y_i = \sum_{j \in [n]} A_{ij} x_j$  is still a random variable. The ML solution provides the following estimator of the signal sparsity.

**Proposition 1** Let us define

$$\omega_i^* = \mathbb{1}(y_i \neq 0) = \mathbb{1}\left(\sum_{j \in [n]} A_{ij} x_j \neq 0\right). \quad (15)$$

Then, the ML estimate of the signal sparsity is

$$\hat{k}_o = \frac{\log\left(1 - \frac{\|\omega^*\|_0}{m}\right)}{\log(1 - \gamma)}. \quad (16)$$

The estimator derived in proposition 1 has already been proposed in [26] for estimating the degree of sparsity. In the following, we will denote the estimator in (16) as *oracle estimator* since it is equivalent to estimating  $k$  in the presence of an oracle who knows which entries in  $y$  are only due to noise. In our analysis, we prove that the oracle estimator is asymptotically strongly consistent, i.e., with the property that as the number of measurements increases indefinitely, the resulting sequence of estimates converges almost surely to the true sparsity degree (see Theorem 2). This means that the density functions of the estimators become more and more concentrated near the true value of the sparsity degree.

Given a sequence of events  $\{E_m\}_{m \in \mathbb{N}}$ , we denote with  $\limsup_{m \rightarrow \infty} E_m$  the set of outcomes that occur infinitely many times. More formally,

$$\limsup_{m \rightarrow \infty} E_m = \bigcap_{m=1}^{\infty} \bigcup_{\ell \geq m} E_\ell \quad (17)$$

**Theorem 2** Let  $p_k = 1 - (1 - \gamma)^k$ , then

$$\mathbb{P}\left(e(\hat{k}_o, k) > \epsilon\right) \leq 2e^{-2m\xi_k^2}, \quad (18)$$

where  $\xi_k = (1 - p_k)(1 - e^{\epsilon \log(1 - p_k)})$ . Moreover, let  $E_m$  be the sequence of events

$$E_m = \left\{ e(\hat{k}_o, k) \geq \frac{\log\left(1 - \frac{\sqrt{\rho}}{1 - p_k} \sqrt{\frac{\log m}{m}}\right)}{\log(1 - p_k)} \right\} \quad (19)$$

with  $m \in \mathbb{N}$  and  $\rho > 1/2$ , then

$$\mathbb{P}\left(\limsup_{m \rightarrow \infty} E_m\right) = 0. \quad (20)$$

**Remark 1** From Theorem 2, we deduce that almost surely (i.e., with probability 1) the relative error between the estimated sparsity and the true value of the sparsity degree is

$$e(\hat{k}_o, k) = O\left(\sqrt{\frac{\log m}{m}}\right) \quad (21)$$

for all but finitely many  $m$ .

#### 4.1 Asymptotic analysis of ML estimator

We analyze now the performance of the oracle estimator in the large system limit, when  $n, k$ , and  $m$  tend to infinity. Since we are dealing with sparse signals, we assume that the sparsity degree  $k$  scales at most linearly in the signal length, i.e., the relative sparsity  $k/n \leq \rho$  is kept bounded with  $\rho \ll 1$ . The following theorem shows sufficient conditions on the number of measurements for weak consistency in different regimes of behavior, defined by

the scaling of the measurement sparsity  $\gamma$  and the signal sparsity  $k$ .

**Theorem 3** Let  $\psi(k) = \gamma k = o(k)$  as  $k \rightarrow \infty$  and define the function  $g(k)$  as follows:

- a) if  $\psi(k) \rightarrow \infty$  as  $k \rightarrow \infty$ , then  $g(k) = \Omega(e^{2\psi(k)})$ ;
- b) if  $\psi(k) = \Theta(1)$  as  $k \rightarrow \infty$ , then  $g(k) \rightarrow \infty$  for  $k \rightarrow \infty$ ;
- c) if  $\psi(k) = o(1)$  as  $k \rightarrow \infty$ , then  $g(k) = \Omega(\psi(k)^{-2(1+\epsilon)})$ , for any  $\epsilon > 0$ .

If the number of measurements is such that  $\frac{m}{\log m} \geq g(k)$ , then

$$\mathbb{P}\left(e(\widehat{k}_o, k) \geq \epsilon_k\right) \xrightarrow{k \rightarrow \infty} 0, \quad (22)$$

where

$$\epsilon_k = \frac{\log\left(1 - \frac{\sqrt{\rho}}{1-p_k} \sqrt{\frac{\log m}{m}}\right)}{\log(1-p_k)} \xrightarrow{k \rightarrow \infty} 0 \quad (23)$$

for some constant  $\rho > 1/2$ .

In the following theorem, we show that, under stricter conditions, strong consistency is also ensured.

**Theorem 4** Let  $\psi(k) = \gamma k$  and define the function  $g(k)$  as in Theorem 3. If the number of measurements is such that

$$m \geq \max\left\{k, \min\left\{\ell \in \mathbb{N} : \frac{\ell}{\log \ell} \geq g(k)\right\}\right\} \quad (24)$$

then

$$\mathbb{P}\left(\limsup_{k \rightarrow \infty} e(\widehat{k}_o, k) = 0\right) = 1. \quad (25)$$

**Remark 2** Theorems 3 and 4 characterize the regimes in which measurement sparsity begins to improve the estimation of the signal sparsity. The function  $\psi(k) = \gamma k$  represents the average number of non-zeros in each row of  $A$  that align with the support of the signal  $x$ . This analysis reveals three cases of interest, corresponding to whether measurement sparsity has no effect, a small effect, or a significant effect on the number of measurements sufficient for asymptotic consistency. If  $\psi(k) = \Theta(1)$  as  $k \rightarrow \infty$ , then  $m = \Omega(k)$  measurements are sufficient for the concentration result. In sharp contrast, if  $\psi(k) \rightarrow \infty$  as  $k \rightarrow \infty$ , then the number of measurements guaranteeing the asymptotic consistency is exponential in  $\psi(k)$ , meaning that, in order to be sure to get an unbiased estimator with  $k$  measurements, we need  $\psi(k) \leq \frac{1}{2}(\log k - \log(\log k))$ .

If  $\psi(k) \rightarrow 0$ , then the condition  $\psi(k) \geq \frac{2+\epsilon}{\sqrt{\log(k)/k}}$  with  $\epsilon > 0$  is sufficient to get an unbiased estimator with  $k$  measurements.

**Remark 3** Theorems 3 and 4 suggest that in order to obtain a good estimation of the sparsity degree, we need sufficiently sparse matrices, but not too sparse. On the other hand, at the same time, the use of sparser matrices requires more measurements for a successful recovery of the signal. If we combine the results obtained in Theorems 3 and 4 with those provided in Theorem 1, we notice there is a large range for  $\gamma$  (provided by  $c \leq \psi(k) \leq \frac{1}{2}(\log k - \log(\log k))$  with  $c > 0$ ) where both sparsity estimation and the recovery can be successful. We will provide more details on how to choose  $\gamma$  for specific applications in Section 6.

The proofs of Theorems 3 and 4 are postponed to the Appendix.

## 5 Method: noisy setting

As already noticed, in the generic noisy setting, the estimation of signal sparsity via an exhaustive ML is unfeasible. A possible approach is to resort to the well-known EM algorithm [36]. This algorithm can find ML solutions to problems involving observed and hidden variables, and in the general, setting is known to converge to a local maximum of the likelihood.

In this section, we prove that, under suitable conditions, the distribution of the measurements can be well approximated by a two-component Gaussian mixture that can be easily learned by EM algorithm. Finally, we show that the case of not-exactly sparse signals can be well approximated by the same model.

### 5.1 2-GMM approximation for large system limit

Our main goal is to show that the Gaussian mixture model that describes the measurements can be simplified in the large system limit as  $n, k \rightarrow \infty$ . The next theorem reveals that there is a regime of behavior, defined by the scaling of the measurement sparsity  $\gamma$  and the signal sparsity  $k$ , where the measurements can be approximately described by a two-component Gaussian mixture model (2-GMM). We state this fact formally below. Recall that, given two density functions  $f, g$ , their Kolmogorov distance is defined as

$$\|f - g\|_K = \sup_{t \in \mathbb{R}} \left| \int_{-\infty}^t f(\zeta) d\zeta - \int_{-\infty}^t g(\zeta) d\zeta \right|. \quad (26)$$

**Theorem 5** Let  $\text{supp}(x) = k$  and  $\phi(\zeta|\sigma^2)$  be the probability density function of a normally distributed random variable with expected value 0 and variance  $\sigma^2$ , i.e.,

$$\phi(\zeta|\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\zeta^2}{2\sigma^2}}.$$

Given a set  $S$ , let  $\alpha_S = \|x_S\|^2/\gamma + \sigma^2$  and  $p_S = (1 - \gamma)^{k-|S|}\gamma^{|S|}$ . Let us consider the density functions (the subscript is to emphasize the dependence on parameter  $k$ )

$$f_k(\zeta) = \sum_{S \subseteq \text{supp}(x)} p_S \phi(\zeta | \alpha_S) \quad (27)$$

$$f_k^{2-GMM}(\zeta) = (1 - p_k) \phi(\zeta | \sigma^2) + p_k \phi\left(\zeta \left| \sigma^2 + \frac{\|x\|^2}{p_k} \right.\right). \quad (28)$$

Then, there exists a constant  $C \in \mathbb{R}$  such that

$$\|f_k - f_k^{2-GMM}\|_K \leq \frac{C\gamma^2}{\lambda^4} \left( \frac{\sum_{i=1}^n x_i^4}{\gamma} + \sum_{i \neq j} x_i^2 x_j^2 - \frac{\|x\|^4}{p_k} \right). \quad (29)$$

The proof of Theorem 5 is postponed to the Appendix. As a simple consequence, we obtain that, under suitable conditions, the approximation error depends on  $\psi(k)$ .

**Corollary 1** Let  $\psi(k) = \gamma k$  and  $f_k, f_k^{2-GMM}$  be the sequence of density functions defined in (27) and (28). Then, there exists a constant  $C' \in \mathbb{R}$  such that

$$\|f_k - f_k^{2-GMM}\|_K \leq C' \left( \frac{\lambda_{\max}}{\lambda_{\min}} \right)^4 (\psi(k) + \psi(k)^2). \quad (30)$$

with  $C' \approx 0.03$ ,  $\lambda_{\max} = \max_{i: x_i \neq 0} |x_i|$ ,  $\lambda_{\min} = \min_{i: x_i \neq 0} |x_i|$ .

Corollary 1 shows that the error in the approximation can be controlled by parameter  $\psi(k)$ . Some considerations are in order. Consider for example a  $k$ -sparse signal with all non-zero components equal in modulus, i.e., with  $\lambda_{\max} = \lambda_{\min}$ . Then, the bound reduces to  $\|f_k - f_k^{2-GMM}\|_K \leq C(\psi(k) + \psi(k)^2)$ . We can see that if  $\psi(k) = \gamma k \rightarrow 0$ , then the Kolmogorov distance goes to zero. However, as suggested by Theorem 1, we expect to need more measurements  $m$  to perform a good estimation of the sparsity degree. The best regime is when  $\psi(k) = \Theta(1)$  as  $k \rightarrow \infty$ : in that case, the distance remains bounded and we expect that a number of measurements proportional to  $k$  is sufficient for the sparsity estimation (suppose, for example, that  $\gamma = 3/k$ . Then, we expect  $\|f_k - f_k^{2-GMM}\|_K < 0.36$ ). For signals with  $\lambda_{\max} \neq \lambda_{\min}$  similar considerations can be done if  $\lambda_{\max}$  and  $\lambda_{\min}$  scale similarly as a function of  $k$ .

## 5.2 Sparsity estimation via EM

Using the approximation in Theorem 5, we recast the problem of inferring the signal sparsity as the problem of estimating the parameters of a two-component Gaussian mixture, whose joint density function of  $y \in \mathbb{R}^m$  and hidden class variables  $z \in \{0, 1\}^m$  is given by

$$f(y_i, z_i | \alpha, \beta, p) = (1 - p)(1 - z_i) \phi(y_i | \alpha) + p z_i \phi(y_i | \beta) \quad (31)$$

with  $i \in [m]$ . Starting from an initial guess of mixture parameters  $\alpha(0)$ ,  $\beta(0)$ , and  $p(0)$ , the algorithm that we propose (named EM-Sp and summarized in Algorithm 1) computes, at each iteration  $t \in \mathbb{N}$ , the posterior distribution

$$\pi_i(t) = \mathbb{P}(z_i = 1 | \alpha(t), \beta(t), p(t)) \quad (32)$$

(E-Step) and re-estimates the mixture parameters (M-Step) until a stopping criterion is satisfied. Finally, the estimation of the signal sparsity is provided by

$$\hat{k} = \log(1 - p_{\text{final}}) / \log(1 - \gamma). \quad (33)$$

---

### Algorithm 1 EM-Sp [29]

---

**Require:** Measurements  $y \in \mathbb{R}^n$ , parameter  $\gamma$

- 1: Initialization:  $\alpha(0), \beta(0), p(0)$
- 2: **for**  $t = 0, 1, \dots, \text{StopIter}$  **do**
- 3: E-step: for  $i = 1, \dots, m$ , compute the posterior probabilities

$$\pi_i(t+1) = \frac{\frac{p(t)}{\sqrt{\beta(t)}} e^{-y_i^2/(2\beta(t))}}{\frac{1-p(t)}{\sqrt{\alpha(t)}} e^{-y_i^2/(2\alpha(t))} + \frac{p(t)}{\sqrt{\beta(t)}} e^{-y_i^2/(2\beta(t))}} \quad (34)$$

- 4: M-Step: compute the mixture parameters

$$p(t+1) = \frac{\sum_{i=1}^m \pi_i(t+1)}{m}, \quad (35)$$

$$k(t+1) = \frac{\log(1 - p(t+1))}{\log(1 - \gamma)} \quad (36)$$

$$\alpha(t+1) = \frac{\sum_{i=1}^m (1 - \pi_i(t+1)) y_i^2}{\sum_{i=1}^m (1 - \pi_i(t+1))}, \quad (37)$$

$$\beta(t+1) = \frac{\sum_{i=1}^m \pi_i(t+1) y_i^2}{\sum_{i=1}^m \pi_i(t+1)}. \quad (38)$$

- 5: **end for**
- 

The sequence of signal sparsity estimations  $k(t)$  generated by Algorithm 1 converges to a limit point. For brevity, we omit the proof, which can be readily derived from standard convergence arguments for dynamical systems [37]. In the following, we will denote Algorithm 1 as EM-Sparse (EM-Sp).

## 5.3 The Cramér-Rao bound for 2-GMM

The Cramér-Rao (CR) bound is a popular lower bound on the variance of estimators of deterministic parameters. Given a parameter  $\xi$  and an unbiased estimator  $\hat{\xi}$ , let  $f(x; \xi)$  be the likelihood function. The CR bound is given by



$$\text{CR}(\hat{\xi}) = \frac{1}{\mathbb{E} \left( \frac{\partial \log(f(x; \xi))}{\partial \xi} \right)^2} \quad (39)$$

that is, the inverse of the Fisher information.

The EM-Sp algorithm, for measurements that can be exactly modeled as a 2-GMM, and for a large number of measurements, would be asymptotically optimal and unbiased and achieve a performance very close to the CR bound. However, because of the presence of noise in the data and the approximation of the 2-GMM model, we expect that the estimator provided by EM-Sp algorithm will be biased. A theoretical analysis of the bias in terms of these two factors is hard to carry out. In the following, we analyze the performance of EM-Sp in the estimation of the 2-GMM parameters via the CR bound, which gives us an indication of how much the non-idealities of the model affect the performance of the proposed estimator.

Let us consider a 2-GMM framework, in which two zero-mean Gaussians with known variances  $\alpha$  and  $\beta$  are given, and let us call  $p$  the mixture parameter. The likelihood function is  $f(x; p) = (1 - p)\phi(x|\alpha) + p\phi(x|\beta)$ , and

$$\text{CR}(\hat{p}) = \frac{1}{\mathbb{E} \left( \frac{\phi(x|\beta) - \phi(x|\alpha)}{(1-p)\phi(x|\alpha) + p\phi(x|\beta)} \right)^2} \quad (40)$$

is the CR bound for the ML estimator  $\hat{p}$  of  $p$ . The stochastic mean cannot be computed in a closed form, but can be approximated with a Monte Carlo method.

$\text{CR}(\hat{p})$  represents a benchmark to evaluate the accuracy of our estimation of  $p$  via EM-Sp, as will be done in Section 6.2.

## 6 Results and discussion

In this section, we illustrate the performance of the proposed estimators through extensive numerical simulations<sup>1</sup>. We present experiments both in the noise-free setting and in non-ideal settings, where signals are not exactly sparse or measurements are affected by noise.

Finally, an application where sparsity estimation improves the efficiency of signal recovery in CS is proposed.

### 6.1 Noise-free measurements

We start testing signals that are exactly sparse and measurements that are not affected by additive noise.

We evaluate the estimation accuracy in terms of empirical probability of correct estimation: a run is considered successful if  $e(k, \hat{k}) < 5 \times 10^{-2}$  where  $\hat{k}$  is the estimated sparsity. In Fig. 1, we show results averaged over 1000 random instances, obtained by generating different sensing matrices from the  $\gamma$ -sparsified Gaussian ensemble. We underline that the values of the non-zero entries of the signals (which are drawn from a standard Gaussian distribution for this experiment) do not affect the performance in the noise-free case. Similarly, the length  $n$  of the signal plays no role in the estimation (see Proposition 1).

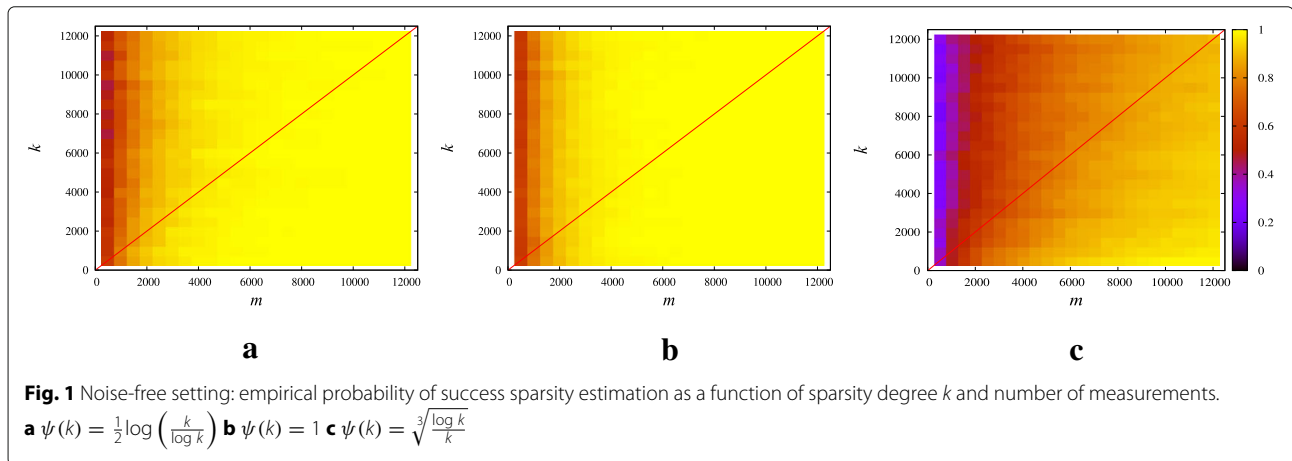
The empirical probability of correct estimation is studied as a function of  $m$  and  $k$  for three different regimes of parameter  $\psi(k)$  defined in Remark 2 (see Fig. 1) :

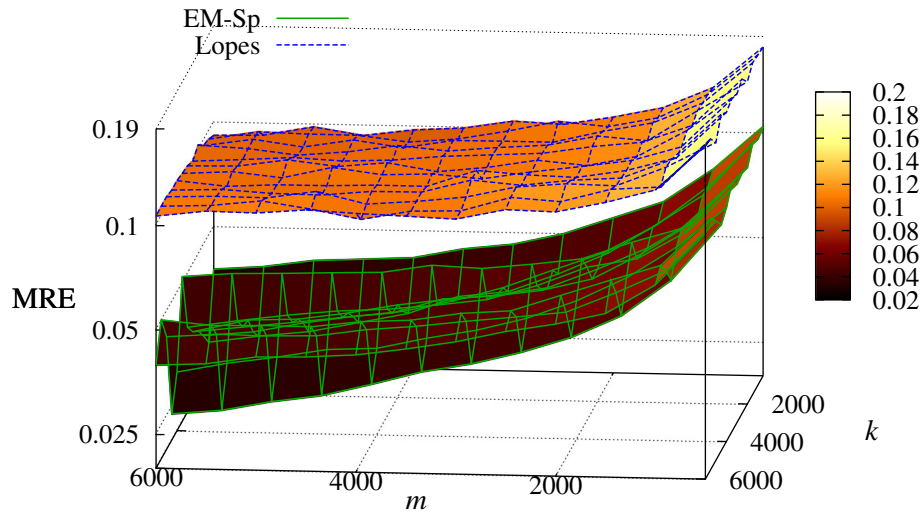
- a)  $\psi(k) = \frac{1}{2}(\log(k) - \log(\log k))$ ;
- b)  $\psi(k) = 1$ ;
- c)  $\psi(k) = \sqrt[3]{\log k / k}$ .

According to Theorem 3 (see also Remark 2), when  $m \geq k$ , the relative error between the estimated sparsity and the true value of the sparsity degree tends to zero almost surely (i.e., with probability 1). This can be appreciated also in the numerical results in Fig. 1, where the line  $m = k$  is drawn for simplicity. Moreover, we can see that for any fixed  $k$ , the error decreases when  $m$  increases.

### 6.2 Noisy measurements

In the second experiment, we show the performance of the EM-Sp algorithm when measurements are noisy according to the model proposed in (1) and we compare to the numerical sparsity estimator [20]. In order





**Fig. 2** Experiment 2: MRE (in log-scale) of EM-Sp and Lopes's estimator as a function of the sparsity degree  $k$  and the number of measurements  $m$ , SNR = 10 dB

to have a fair comparison, we perform this test on ternary signals in  $\{-\lambda, 0, \lambda\}^n$  for which sparsity and numerical sparsity coincide. We then consider random sparse signals with non-zero entries uniformly chosen in  $\{\lambda, -\lambda\}$ ,  $\lambda \in \mathbb{R}$ , and  $\text{SNR} = \lambda^2 k / \sigma^2$  (see definition in Section 3). Moreover, we set  $\psi(k)$  constant in order to focus on the effects of the additive noise in the estimation.

We remark that we compare only to [20] because, as illustrated in Section 1.1, the other proposed algorithms for sparsity degree estimation are based on signal reconstruction [14] (requiring a larger number of measurements and increased complexity, which would give

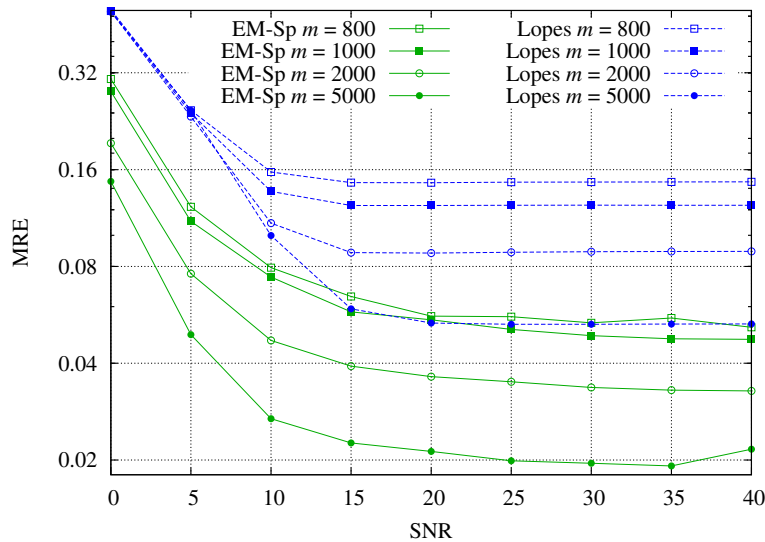
an unfair comparison) or are conceived for very specific applications [13, 22].

In Fig. 2, we show the mean relative error (MRE) defined as

$$\text{MRE} = \mathbb{E} \left[ e(k, \hat{k}) \right] \quad (41)$$

for different values of  $k$  and  $m$  in settings with  $\text{SNR} = 10$  dB and  $\psi(k) = 1/10$ . We appreciate that, in the considered framework, EM-Sp always outperforms the method based on the numerical sparsity estimation.

In Fig. 3, we set  $k = 1000$ ,  $\psi(k) = 1/3$ , and we vary the SNR from 0 to 40 dB, while  $m \in \{800, 1000, 2000, 5000\}$ .



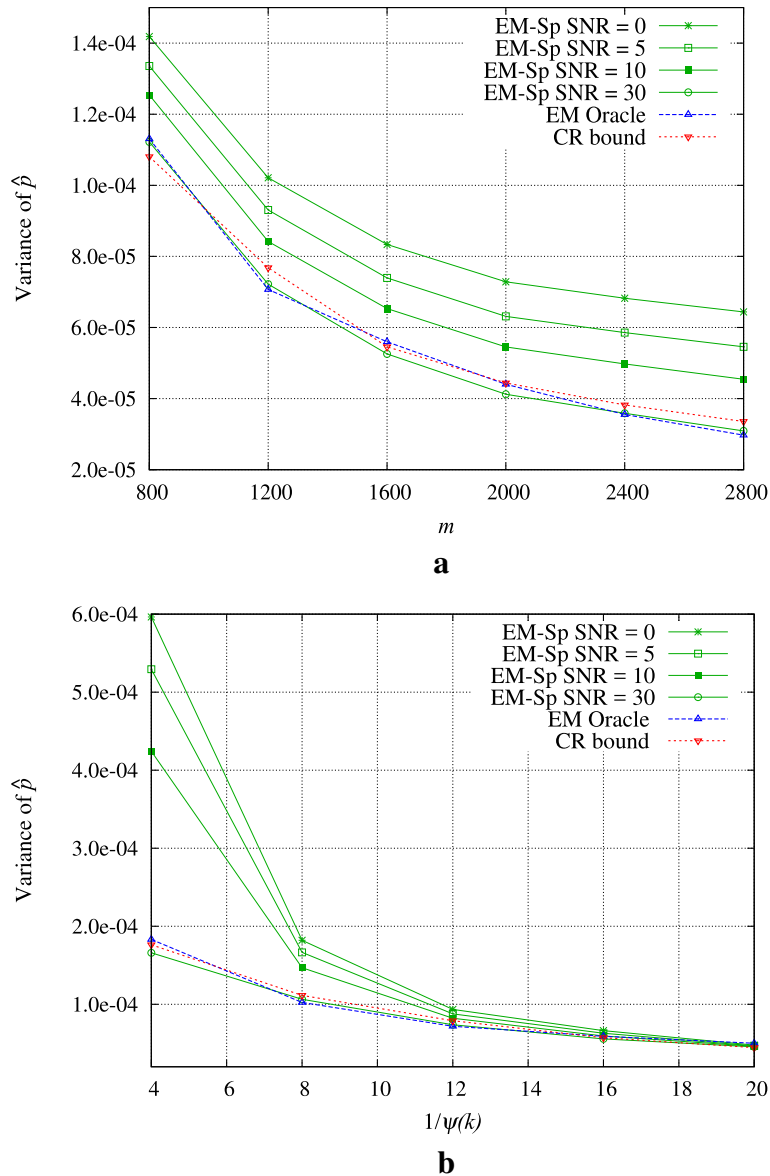
**Fig. 3** Experiment 2: MRE (in log-scale) of EM-Sp and numerical sparsity estimator as a function of the SNR, for different  $m$ 's,  $k = 1000$

Again, we see that EM-Sp outperforms [20]. We specify that a few tens of iterations are sufficient for the convergence of EM-Sp.

Finally, we compare the performance of EM-Sp with an oracle estimator designed as follows: we assume to know exactly the variances  $\alpha$  and  $\beta$  and we generate measurements  $y_i$  distributed according to 2-GMM  $(1-p)\phi(y_i|\alpha) + p\phi(y_i|\beta)$ , for  $i = 1, \dots, m$ ; given the sequence  $y$ ,  $\alpha$ , and  $\beta$ , we then compute the ML estimate of  $p$  via EM. We name this estimator EM oracle. Comparing the estimates of  $p$  of EM-Sp and EM oracle, we can check if our 2-GMM approximation is reliable. We clearly expect that EM oracle performs better, as the measurements are really generated according to a 2-GMM, and also the true  $\alpha$

and  $\beta$  are exploited. However, our results show that the 2-GMM approximation is trustworthy. In Fig. 4, we depict the sample variance of the estimator  $\hat{p}$  of  $p$  (obtained from 1000 runs) of EM-Sp and EM oracle. We show also the CR bound (see Section 5.3), which represents a performance lower bound for the estimation of  $p$ . As explained in Section 5.3, the stochastic mean required in the CR bound for 2-GMM cannot be analytically computed and is here evaluated via Monte Carlo.

In both graphs of Fig. 4, we set  $k = 1000$ ,  $m = k$  in Fig. 4a, and  $\psi(k) = 1/10$  in Fig. 4b. We notice that in the considered regimes, EM oracle and CR bound are very close and not really affected by the SNR. Regarding EM-Sp, we observe that (a) keeping  $k, \gamma$  fixed, EM-Sp gets closer



**Fig. 4** Noisy setting: comparison with oracle EM and Cramér-Rao (CR) bound. **a**  $k = 1000$ ,  $\psi(k) = 1/10$  **b**  $k = m = 1000$

to the optimum as the SNR increases; and (b) keeping  $k, m$  fixed, we can find an optimal  $\gamma$  that allows us to get very close to the optimum.

### 6.3 Compressibility of real signals

In this section, we test our EM-Sp algorithm to evaluate the compressibility of real signals. Specifically, we consider images which are approximately sparse in the discrete cosine transform (DCT) domain, that is, they are well represented by few DCT coefficients. Our aim is to estimate the number  $k$  of significant DCT coefficients. More precisely, we seek the minimum  $k$  such that the best- $k$  approximation  $\hat{x}_k$  has a relative error smaller approximately  $\tau$ , namely  $\|\hat{x}_k - x\|_2^2 \leq \tau \|x\|_2^2$ . Since DCT coefficients of natural images usually have a power-law decay of the form  $x_i \approx c/i$ , the following approximation holds

$$\frac{\|\hat{x}_k - x\|_2^2}{\|x\|_2^2} \approx \frac{\int_k^n x^{-2} dx}{\int_1^n x^{-2} dx} \propto k^{-1} \quad (42)$$

and since according to theoretical derivation  $\gamma \propto 1/k$ , we fix  $\gamma \propto \tau$ . Tuning  $\gamma$  proportionally to  $\tau$  allows to adapt better the sparsity of the sensing matrix to the expected signal sparsity: for larger  $\tau$ 's, we expect smaller  $k$ 's, which call for larger  $\gamma$  to have the sufficient matrix density to pick the non-zero entries.

In the proposed experiments, we fix  $\gamma = c\tau$  with  $c = 5 \cdot 10^{-2}$  and we initialize  $\pi_i(0) = \frac{1}{2}$  for all  $i = 1, \dots, m$ , while we set  $\beta$  and  $\alpha$  respectively as the signal energy and

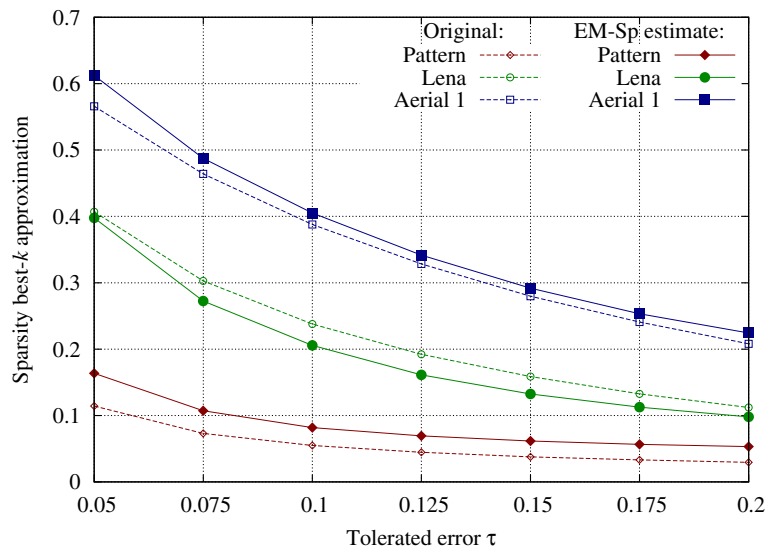
the noise energy (namely, the error of the best- $k$  approximation), evaluated from the measurements:  $\beta = \frac{\|y\|_2^2}{m}$  and  $\alpha = \tau\beta$ .

In Fig. 5, we depict the results on three  $n = 256 \times 256$  images (shown in Fig. 6) represented in the DCT basis (DCT is performed on  $8 \times 8$  blocks). Specifically, we show original and estimated sparsity (the  $y$ -axis represents the ratio  $k/n$ ), averaged on 100 random sensing matrices. The images have been chosen with different compressibilities, to test our method in different settings. We appreciate that for all the images and across different values of  $\tau$ , we are able to estimate  $k$  with a small error. This experiment shows then that EM-Sp can be practically used to estimate the compressibility of real signals.

### 6.4 Sparsity estimation for signal recovery

We have already remarked that the knowledge of the sparsity degree  $k$  is widely used for signal recovery in CS. In this section, we consider CoSaMP [9], an algorithm which can recover a sparse signal (exactly or with a bounded error, in the absence and in the presence of noise, respectively) if a sufficient number of measurements  $m$  is provided, and assuming the knowledge an upper bound  $k_{\max}$  for  $k$ . Our aim is to show that EM-Sp can be used as a pre-processing for CoSaMP when  $k$  is not known; specifically, we estimate  $k$  to design the number of measurements necessary for CoSaMP recovery. Subsequently, we denote this joint procedure as EM-SP/CoSaMP.

We compare CoSaMP with EM-Sp/CoSaMP in the following setting. We consider a family  $\mathcal{S}$  of signals of length



**Fig. 5** Experiment 3: compressibility of real signals, with non-exactly sparse representations. We estimate  $k$  such that  $\frac{\|\hat{x}_k - x\|_2^2}{\|x\|_2^2} < \tau$ , where  $x$  and  $\hat{x}_k$  respectively are the non-exactly sparse representation and its best- $k$  approximation



**Fig. 6** Non-exactly sparse signals: images with approximately sparse DCT representation: pattern, Lena, and aerial

$n = 1600$  and (unknown) sparsity  $k \in \{20, 200\}$  (then,  $k_{\max} = 200$ ). The value of  $k$  and the position of the non-zero coefficients are generated uniformly at random, and the non-zero values are drawn from a standard Gaussian distribution. Since  $k$  is not known, the number of measurements needed by CoSaMP has to be dimensioned on  $k_{\max}$ : assuming  $\text{SNR} = 30$  dB, from the literature, we get that  $m_C = 4k_{\max}$  are sufficient to get a satisfactory recovery using dense Gaussian sensing matrices. In our specific setting, we always observe a mean relative error  $\text{MRE}_{\text{rec}} = \|x - \hat{x}\|_2 / \|x\|_2 < 5.5 \times 10^{-2}$  (for each  $k \in \{20, 200\}$ , 100 random runs have been performed).

We propose now the following procedure.

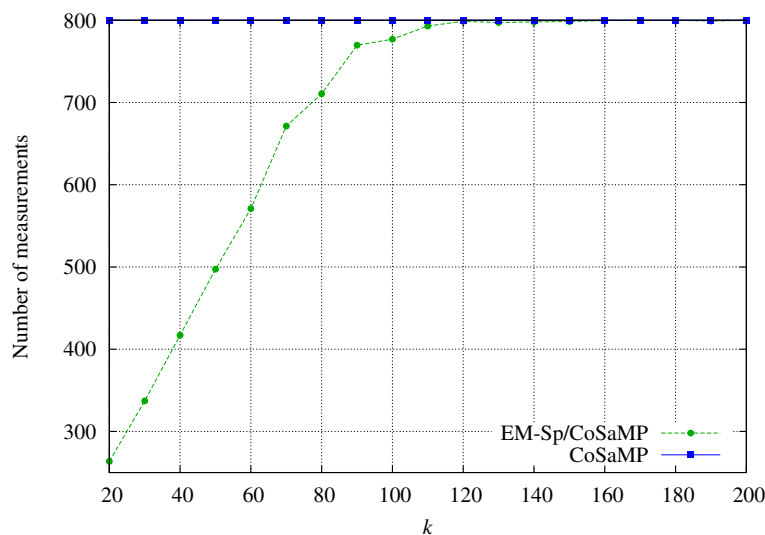
- 1 First sensing stage and sparsity estimation: we take  $m_S \ll m_C$  measurements via  $\gamma$ -sparsified matrix in

(2), and we provide an estimate  $\hat{k}$  of  $k$  using Algorithm 1.

- 2 Second sensing stage and recovery: we add a sufficient number of measurements (dimensioned over  $\hat{k}$ ) and then perform CoSaMP recovery.

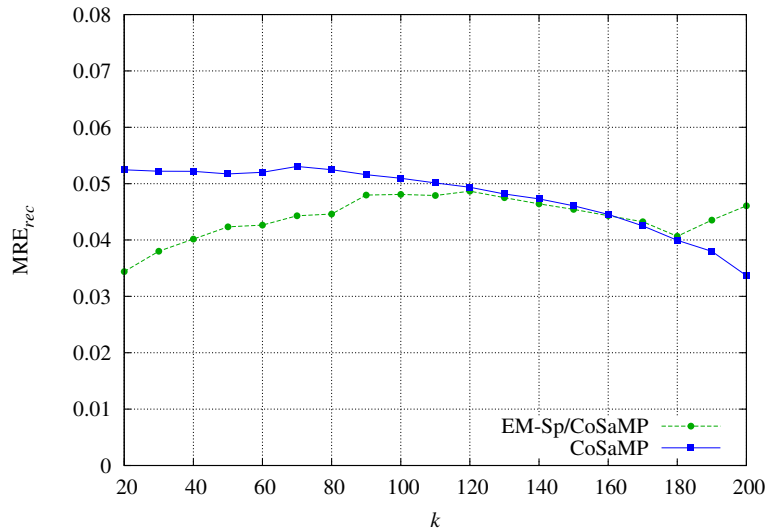
Specifically, the following assessments have been proved to be suitable for our example:

- We estimate  $k$  with EM-Sp from  $m_S = \frac{k_{\max}}{2}$  sparsified measurements, with  $\gamma = 6/k_{\max}$ ;
- Since underestimates of  $k$  are critical for CoSaMP, we consider  $\hat{k}$  equal to 2 times the estimate provided by EM-Sp;
- We add  $m_A = 4\hat{k}$  Gaussian measurements, and we run CoSaMP with the so-obtained sensing matrix with  $m_S + m_A$  rows. When  $m_S + m_A > m_C$ , we reduce the total number of measurements to  $m_C$ .



**Fig. 7** Recovery experiment with unknown  $k$ : EM-Sp/CoSaMP saves a significant number of measurements with respect to pure CoSaMP





**Fig. 8** The gain obtained by EM-Sp/CoSaMP in terms of measurements has no price in terms of accuracy: the recovery mean relative error is close to that of pure CoSaMP

We show the results averaged over 100 random experiments. In Fig. 7, we compare the number of measurements used for recovery, as a function of the sparsity degree  $k$ : a substantial gain is obtained in terms of measurements by EM-Sp/CoSaMP, with no significant accuracy loss. In Fig. 8, we can see that CoSaMP and EM-Sp/CoSaMP algorithms achieve similar  $MRE_{rec}$ .

## 7 Conclusions

In this paper, we have proposed an iterative algorithm for the estimation of the signal sparsity starting from compressive and noisy projections obtained via sparse random matrices. As a first theoretical contribution, we have demonstrated that the estimator is consistent in the noise-free setting and we characterized its asymptotic behavior for different regimes of the involved parameters, namely the sparsity degree  $k$ , the number of measurements  $m$ , and the sensing matrix sparsity parameter  $\gamma$ . Then, we have showed that in the noisy setting, the projections can be approximated using a 2-GMM, for which the EM algorithm provides an asymptotically optimal estimator.

Numerical results confirm that the 2-GMM approach is effective for different signal models and outperforms methods known in the literature, with no substantial increase of complexity. The proposed algorithm can represent a useful tool in several applications, including the estimation of signal sparsity before reconstruction in a sequential acquisition framework, or the estimation of support overlap between correlated signals. An important property of the proposed method is that it does not rely on the knowledge of the actual sensing

matrix, but only on its sparsity parameter  $\gamma$ . This enables applications in which one is not interested in signal reconstruction, but only in embedding the sparsity degree of the underlying signal in a more compact representation.

## Endnote

<sup>1</sup>The code to reproduce the simulations proposed in this section is available at <https://github.com/sophie27/sparsityestimation>

## Appendix

### Proofs of results in Section 4

#### Proof of Proposition 1

It should be noticed that  $\mathbb{P}(\omega_i^* = 0) = (1-\gamma)^{\theta n}$ , where  $\theta \in [0, 1]$  is the parameter to be optimized. Since the rows of the matrix  $A$  are independent, so are  $\omega_i^*$ , and considering that the event  $\omega_i^* = 0$  is equivalent to the event that the support of  $i$ th row of  $A$  is orthogonal to the support of signal  $x$ , the ML estimation computes

$$\hat{\theta}_o = \underset{\theta \in [0,1]}{\operatorname{argmax}} \log f(\omega^* | \theta) = \sum_{i=1}^m f(\omega_i^* | \theta) \quad (43)$$

$$= \underset{\theta \in [0,1]}{\operatorname{argmax}} \sum_{i=1}^m \log \left[ (1 - (1-\gamma)^{\theta n})^{\omega_i^*} (1-\gamma)^{\theta n(1-\omega_i^*)} \right] \quad (44)$$

from which

$$\hat{\theta}_o = \frac{\log \left( 1 - \frac{\|\omega^*\|_0}{m} \right)}{n \log(1-\gamma)}. \quad (45)$$

We conclude that  $\hat{k}_o = \hat{\theta}_o n$ .  $\square$

**Proof of Theorem 2**

Let us consider  $\omega_i^*$  as defined in (15) and let  $\hat{p}_k = \frac{\|\omega_i^*\|_0}{m}$  where the index emphasizes the dependence on the sparsity degree. We thus have:

$$\begin{aligned} \mathbb{P}\left(e(\hat{k}_o, k) > \epsilon\right) &= \mathbb{P}\left(\left|\frac{\hat{k}_o - k}{k}\right| > \epsilon\right) \\ &= \mathbb{P}\left(\frac{|\log(1 - \hat{p}_k) - \log(1 - p_k)|}{|k \log(1 - \gamma)|} > \epsilon\right). \end{aligned} \quad (46)$$

Since  $|k \log(1 - \gamma)| = |\log(1 - \gamma)^k| = -\log(1 - p_k)$ , we obtain

$$\begin{aligned} \mathbb{P}\left(e(\hat{k}_o, k) > \epsilon\right) &= \mathbb{P}\left(\left|\log\left(\frac{1 - \hat{p}_k}{1 - p_k}\right)\right| > -\epsilon \log(1 - p_k)\right) \quad (47) \\ &= \mathbb{P}\left(\left|\log\left(1 + \frac{p_k - \hat{p}_k}{1 - p_k}\right)\right| > -\epsilon \log(1 - p_k)\right) \quad (48) \end{aligned}$$

$$= \mathbb{P}\left(\log\left(1 + \frac{p_k - \hat{p}_k}{1 - p_k}\right) > -\epsilon \log(1 - p_k)\right) \quad (49)$$

$$+ \mathbb{P}\left(\log\left(1 + \frac{p_k - \hat{p}_k}{1 - p_k}\right) < -\epsilon \log(1 - p_k)\right) \quad (50)$$

$$= \mathbb{P}\left(p_k - \hat{p}_k > \left(e^{-\epsilon \log(1 - p_k)} - 1\right)(1 - p_k)\right) \quad (51)$$

$$+ \mathbb{P}\left(p_k - \hat{p}_k < \left(e^{\epsilon \log(1 - p_k)} - 1\right)(1 - p_k)\right) \quad (52)$$

$$\leq \mathbb{P}\left(|p_k - \hat{p}_k| > \xi_k\right) \quad (53)$$

with

$$\xi_k = (1 - p_k) \min \left\{ \left(e^{-\epsilon \log(1 - p_k)} - 1\right), \left(1 - e^{\epsilon \log(1 - p_k)}\right) \right\} \quad (54)$$

$$= (1 - p_k) \left(1 - e^{\epsilon \log(1 - p_k)}\right). \quad (55)$$

It should be noticed that  $p_k = \mathbb{E}[\hat{p}_k]$ , hence applying the Chernoff-Hoeffding theorem [38], the above tail probability is upper bounded as

$$\mathbb{P}\left(e(\hat{k}_o, k) > \epsilon\right) \leq 2e^{-2m\xi_k^2} \quad (56)$$

and we obtain the first part of the statement. Choosing

$$\epsilon = \frac{\log\left(1 - \frac{\sqrt{\rho}}{1 - p_k} \sqrt{\frac{\log m}{m}}\right)}{\log(1 - p_k)} \quad (57)$$

for some  $\rho > 1/2$ , we get

$$\mathbb{P}\left(e(\hat{k}_o, k) > \epsilon\right) \leq 2e^{-2m(1 - p_k)^2 \left(1 - e^{\epsilon \log(1 - p_k)}\right)^2} \quad (58)$$

$$= 2e^{-2m(1 - p_k)^2 \left(1 - e^{\log\left(1 - \frac{\sqrt{\rho}}{1 - p_k} \sqrt{\frac{\log m}{m}}\right)}\right)^2} \quad (59)$$

$$= 2e^{-2\rho \log m} = \frac{2}{m^{2\rho}} \quad (60)$$

and from Borel-Cantelli Lemma [39], we conclude that

$$\mathbb{P}\left(\limsup_{k \rightarrow \infty} \left\{e(\hat{k}_o, k) \geq \epsilon_k\right\}\right) = 0.$$

□

**Proof of Theorem 3**

From Theorem 2, we have

$$\mathbb{P}\left(\left\{e(\hat{k}_o, k) \geq \epsilon_k\right\}\right) \leq \frac{2}{m^{2\rho}},$$

and combining the hypothesis  $m/\log(m) \geq g(k)$ , we get

$$\mathbb{P}\left(\left\{e(\hat{k}_o, k) \geq \epsilon_k\right\}\right) \leq \frac{2}{(g(k))^{2\rho} (\log m)^{2\rho}} \leq \frac{1}{(g(k))^{2\rho}}$$

where the last inequality is obtained noticing that  $\log m \geq 2$  definitely as  $m \rightarrow \infty$ . We distinguish now the different cases

- If  $\psi(k) \rightarrow \infty$  as  $k \rightarrow \infty$ , then the function  $g$  is defined as  $g(k) = \Omega(e^{2\psi(k)})$  from which we get that also  $g(k) \rightarrow \infty$
- If  $\psi(k) = \Theta(1)$  as  $k \rightarrow \infty$ , then the function  $g$  is defined as  $g(k) \rightarrow \infty$  for  $k \rightarrow \infty$  from which we get that also  $g(k) \rightarrow \infty$ ;
- If  $\psi(k) = o(1)$  as  $k \rightarrow \infty$ , then the function  $g$  is defined as  $g(k) = \Omega(\psi(k)^{-2(1+\epsilon)})$ , for any  $\epsilon > 0$  from which we get that also  $g(k) \rightarrow \infty$

Since in all cases (a), (b), and (c), the function  $g(k) \rightarrow \infty$ , we conclude  $\mathbb{P}\left(\left\{e(\hat{k}_o, k) \geq \epsilon_k\right\}\right) \rightarrow 0$  and the Eq. (22) can be deduced.

We now prove that  $\epsilon_k$  tends to zero as  $k \rightarrow \infty$ . Notice that being  $\psi(k) = o(k)$  as  $k \rightarrow \infty$

$$p_k := 1 - (1 - \gamma)^k = 1 - \left(1 - \frac{\psi(k)}{k}\right)^k \sim 1 - e^{-\psi(k)} \quad (61)$$

as  $k \rightarrow \infty$ . We have

$$\epsilon_k = \frac{\log\left(1 - \frac{\sqrt{\rho}}{1 - p_k} \sqrt{\frac{\log m}{m}}\right)}{\log(1 - p_k)} \quad (62)$$

$$\sim \frac{\log\left(1 - \sqrt{\rho} e^{\psi(k)} \sqrt{\frac{\log m}{m}}\right)}{-\psi(k)}. \quad (63)$$

We have

- a) If  $\psi(k) \rightarrow \infty$  then  $\epsilon_k = O(\psi(k)^{-1})$ ;
- b) If  $\psi(k) = \Theta(1)$  then  $\epsilon_k = O(g(k)^{-1/2})$ ;
- c) If  $\psi(k) \rightarrow 0$  then  $\epsilon_k = O(\psi(k)^\epsilon)$ .

We conclude that in all three cases (a), (b), and (c) the threshold  $\epsilon_k \xrightarrow{k \rightarrow \infty} 0$ .

#### Proof of Theorem 4

Let  $\epsilon_k$  be defined as in (23). From Lemma 2, we have, for some  $\rho > 1/2$ ,

$$\mathbb{P}\left(\left\{e(\widehat{k}_o, k) \geq \epsilon_k\right\}\right) \leq \frac{2}{m^{2\rho}} \leq \frac{1}{k^{2\rho}} \quad (64)$$

where the last inequality is obtained noticing that

$$m \geq \max\left\{k, \min\left\{\ell \in \mathbb{N} : \frac{\ell}{\log \ell} \geq g(k)\right\}\right\} \geq k \quad (65)$$

definitely. Since  $2\rho > 1$ , from the Borel-Cantelli lemma, we deduce that

$$\mathbb{P}\left(\limsup_{k \rightarrow \infty} \left\{e(\widehat{k}_o, k) \geq \epsilon_k\right\}\right) = 0. \quad (66)$$

Being  $\log(m)/m \geq g(k)$ , then  $\epsilon_k \rightarrow 0$  as  $k \rightarrow \infty$ , and we conclude that

$$\mathbb{P}\left(\limsup_{k \rightarrow \infty} e(\widehat{k}_o, k) = 0\right) = 1. \quad (67)$$

#### Proof of Theorem 5

In this section, we prove Theorem 5.

**Lemma 1** Let  $A$  be chosen from the  $\gamma$ -sparsified Gaussian ensemble uniformly at random and  $y$  be given in (14),  $p_k = 1 - (1 - \gamma)^k$ , and  $\omega = \mathbb{1}(Ax \neq 0)$  then

$$\mathbb{E}[\text{Var}(y_i)|\omega_i = 1] = \frac{\|x\|^2}{p_k} + \sigma^2$$

$$\text{Var}[\text{Var}(y_i)|\omega_i = 1] = \frac{1}{p_k} \left( \frac{\sum_{\ell=1}^n x_\ell^4}{\gamma} + 2 \sum_{\ell > j} x_\ell^2 x_j^2 - \frac{\|x\|^4}{p_k} \right)$$

*Proof* We recall  $y_i = \sum_{j=1}^n A_{ij}x_j + \eta_i$  with  $\eta_i \sim \mathcal{N}(0, \sigma^2)$ . As already noticed throughout the paper, the measurement  $y_i$  is a mixture of Gaussians with zero mean and variance depending on the overlap between the support of the  $i$ th row of  $A$  and  $\text{supp}(x)$ . Suppose that  $S \subseteq \text{supp}(x)$  is this overlap which happens with probability  $p_S =$

$(1 - \gamma)^{k-|S|} \gamma^{|S|}$ , then the variance of the Gaussian is given by  $\alpha_S = \frac{\|x_S\|^2}{\gamma} + \sigma^2$ . Standard computations lead to

$$\mathbb{E}[\text{Var}(y_i)|\omega_i = 1] = \sum_{S \subseteq \text{supp}(x): S \neq \emptyset} \frac{p_S}{p_k} \left( \frac{\|x_S\|^2}{\gamma} + \sigma^2 \right) \quad (68)$$

$$= \sum_{S \subseteq \text{supp}(x): S \neq \emptyset} \frac{(1 - \gamma)^{k-|S|} \gamma^{|S|}}{p_k} \left( \frac{\|x_S\|^2}{\gamma} \right) + \sigma^2 \quad (69)$$

$$= \frac{1}{p_k} \sum_{S \subseteq \text{supp}(x): S \neq \emptyset} (1 - \gamma)^{k-|S|} \gamma^{|S|-1} \sum_{\ell \in S} x_\ell^2 + \sigma^2. \quad (70)$$

We notice that, fixed a component  $\ell \in \text{supp}(x)$ , we have exactly  $\binom{k-1}{s-1}$  possible sets of cardinality  $s$  containing  $\ell$ , i.e., the number of selections of the remaining  $s - 1$  objects among  $k - 1$  positions. This observation and the fact  $x_\ell = 0, \forall \ell \notin \text{supp}(x)$  leads to

$$\mathbb{E}[\text{Var}(y_i)|\omega_i = 1] = \frac{1}{p_k} \sum_{\ell=1}^n \sum_{s=1}^k (1 - \gamma)^{k-s} \gamma^{s-1} \binom{k-1}{s-1} x_\ell^2 + \sigma^2 \quad (71)$$

$$= \frac{1}{p_k} \sum_{\ell=1}^n \sum_{s=0}^{k-1} (1 - \gamma)^{k-s-1} \gamma^s \binom{k-1}{s} x_\ell^2 + \sigma^2 \quad (72)$$

$$= \frac{1}{p_k} \sum_{\ell=1}^n x_\ell^2 + \sigma^2. \quad (73)$$

We compute now

$$\text{Var}[\text{Var}(y_i)|\omega_i = 1] \quad (74)$$

$$= \sum_{S \subseteq \text{supp}(x)} \frac{p_S}{p_k} \left( \frac{\|x_S\|^2}{\gamma} + \sigma^2 \right)^2 - \left( \frac{\|x\|^2}{p_k} + \sigma^2 \right)^2 \quad (75)$$

$$= \sum_{S \subseteq \text{supp}(x)} \frac{p_S}{p_k} \left( \frac{\|x_S\|^4}{\gamma^2} + \frac{2\|x_S\|^2 \sigma^2}{\gamma} + \sigma^4 \right) \quad (76)$$

$$- \left( \frac{\|x\|^4}{p_k^2} + 2 \frac{\|x\|^2}{p_k} \sigma^2 + \sigma^4 \right) \quad (77)$$

$$= \sum_{S \subseteq \text{supp}(x)} \frac{p_S}{p_k} \frac{\|x_S\|^4}{\gamma^2} - \frac{\|x\|^4}{p_k^2} \quad (78)$$

$$= \frac{1}{p_k} \left( \sum_{S \subseteq \text{supp}(x)} (1 - \gamma)^{k-|S|} \gamma^{|S|} \frac{\|x_S\|^4}{\gamma^2} - \frac{\|x\|^4}{p_k} \right) \quad (79)$$

$$= \sum_{S \subseteq \text{supp}(x)} \frac{(1 - \gamma)^{k-|S|} \gamma^{|S|-2}}{p_k} \sum_{\ell \in S} x_\ell^4 \quad (80)$$

$$+ \sum_{S \subseteq \text{supp}(x)} \frac{(1 - \gamma)^{k-|S|} \gamma^{|S|-2}}{p_k} \sum_{\ell \neq j \in S} x_\ell^2 x_j^2 - \frac{\|x\|^4}{p_k^2} \quad (81)$$

As before we notice that, fixed the component  $\ell$ , we have exactly  $\binom{k-1}{s-1}$  possible sets of cardinality  $s$  containing  $\ell$ . Analogously, the couple  $\ell, j$  with  $\ell \neq j$  is contained in  $\binom{k-2}{s-2}$  possible sets of cardinality  $s$ . We thus have

$$\text{Var}[\text{Var}(y_i)|\omega_i = 1] \quad (82)$$

$$= \sum_{s=1}^k \frac{(1-\gamma)^{k-s} \gamma^{s-2}}{p_k} \binom{k-1}{s-1} \sum_{\ell=1}^n x_\ell^4 \quad (83)$$

$$+ \sum_{s=2}^k \frac{(1-\gamma)^{k-s} \gamma^{s-2}}{p_k} \binom{k-2}{s-2} \sum_{i \neq j} x_i^2 x_j^2 - \frac{\|x\|^4}{p_k^2}. \quad (84)$$

We conclude

$$\text{Var}[\text{Var}(y_i)|\omega_i = 1] \quad (85)$$

$$= \frac{1}{p_k} \left( \frac{\sum_{\ell=1}^n x_\ell^4}{\gamma} + \sum_{i \neq j} x_i^2 x_j^2 - \frac{\|x\|^4}{p_k} \right) \quad (86)$$

□

### Proof of Theorem 5

Let  $\psi(k) = \gamma k$  and consider the sequence of probability density functions

$$f_k(\zeta) = \sum_{S \subseteq \text{supp}(x)} p_S \phi(\zeta | \alpha_S)$$

$$f_k^{2\text{-GMM}}(\zeta) = (1-p_k) \phi(\zeta | \sigma^2) + p_k \phi\left(\zeta \middle| \sigma^2 + \frac{\|x\|^2}{p_k}\right),$$

where  $\alpha_S = \frac{\|x_S\|^2}{\gamma} + \sigma^2$  and  $p_S = (1-\gamma)^{k-|S|} \gamma^{|S|}$ . Let  $\bar{\alpha} = \mathbb{E}[\text{Var}(y_i)|\omega_i = 1] = \sigma^2 + \frac{\|x\|^2}{p_k}$  (see Lemma 1) and denote  $\mathcal{S}$  the set of possible subsets of  $\text{supp}(x)$ , we thus have

$$\begin{aligned} & \|f_k - f_k^{2\text{-GMM}}\|_K \\ &= \sup_{t \in \mathbb{R}} \left| \sum_{S \in \mathcal{S} \setminus \{\emptyset\}} p_S \int_{-\infty}^t \phi(\zeta | \alpha_S) - p_k \int_{-\infty}^t \phi(\zeta | \bar{\alpha}) \right| \\ &= \frac{p_k}{2} \sup_{t \in \mathbb{R}} \left| \sum_{S \in \mathcal{S} \setminus \{\emptyset\}} \frac{p_S}{p_k} \text{erf}\left(\frac{t}{\sqrt{2\alpha_S}}\right) - \text{erf}\left(\frac{t}{\sqrt{2\bar{\alpha}}}\right) \right| \end{aligned} \quad (87)$$

where erf is the Gauss error function. Let  $g : \mathbb{R}^2 \rightarrow \mathbb{R}$  be the function

$$g(t, \alpha) := \text{erf}\left(\frac{t}{\sqrt{2\alpha}}\right) \quad (88)$$

and by the Lagrange Theorem [40], we obtain

$$g(t, \alpha) = g(t, \bar{\alpha}) + \frac{\partial g}{\partial \alpha}(t, \bar{\alpha})(\alpha - \bar{\alpha}) + \frac{1}{2} \frac{\partial^2 g}{\partial \alpha^2}(t, \xi)(\alpha - \bar{\alpha})^2 \quad (89)$$

with  $\xi(\alpha) \in (\min\{\bar{\alpha}, \alpha\}, \max\{\alpha, \bar{\alpha}\})$ . It should be noticed that the first-order term in Taylor's expansion of  $g(t; \alpha)$  vanishes due to conditional mean result from Theorem 2

$$\sum_{S \in \mathcal{S} \setminus \{\emptyset\}} \frac{p_S}{p_k} g(t, \alpha_S) \quad (90)$$

$$= g(t, \bar{\alpha}) + \frac{1}{2} \sum_{S \in \mathcal{S} \setminus \{\emptyset\}} \frac{p_S}{p_k} \left[ \frac{\partial^2 g}{\partial \alpha^2}(t, \xi(\alpha_S))(\alpha_S - \bar{\alpha})^2 \right] \quad (91)$$

with  $\xi(\alpha_S) \in (\min\{\bar{\alpha}, \alpha_S\}, \max\{\alpha_S, \bar{\alpha}\}) \subseteq (\lambda^2/\gamma, \|x\|^2/\gamma)$ . Putting this expression into (87), we obtain

$$\|f_k - f_k^{2\text{-GMM}}\|_K \quad (92)$$

$$\leq \left| \frac{p_k}{4} \sum_{S \in \mathcal{S} \setminus \{\emptyset\}} \frac{p_S}{p_k} \left[ \frac{\partial^2 g}{\partial \alpha^2}(t, \xi(\alpha_S))(\alpha_S - \bar{\alpha})^2 \right] \right| \quad (93)$$

$$\leq \frac{p_k}{4} \sum_{S \in \mathcal{S} \setminus \{\emptyset\}} \frac{p_S}{p_k} \left| \frac{\partial^2 g}{\partial \alpha^2}(t, \xi(\alpha_S)) \right| (\alpha_S - \bar{\alpha})^2 \quad (94)$$

$$\leq \frac{p_k}{4} G_{\max} \sum_{S \in \mathcal{S} \setminus \{\emptyset\}} \frac{p_S}{p_k} (\alpha_S - \bar{\alpha})^2 \quad (95)$$

where

$$G_{\max} = \max_{\xi \in [\lambda^2/\gamma, \|x\|^2/\gamma]} \sup_{t \in \mathbb{R}} \left| \frac{\partial^2 g}{\partial \alpha^2}(t, \xi) \right| \quad (96)$$

and

$$\frac{\partial^2 g}{\partial \alpha^2}(t, \xi) = \frac{1}{2\sqrt{2\pi}} e^{-\frac{t^2}{2\xi}} \frac{t}{\xi^{5/2}} \left( 3 - \frac{t^2}{\xi} \right) \quad (97)$$

Through standard computations, we see that the maximizing value is obtained for  $t = \sqrt{(3-\sqrt{6})\xi}$

$$G_{\max} = \max_{\xi \in [\lambda^2/\gamma, \|x\|^2/\gamma]} \sup_{t \in \mathbb{R}} \left| \frac{1}{2\sqrt{2\pi}} e^{-\frac{t^2}{2\xi}} \frac{t}{\xi^{5/2}} \left( 3 - \frac{t^2}{\xi} \right) \right| \quad (98)$$

$$= C \max_{\xi \in [\lambda^2/\gamma, \|x\|^2/\gamma]} \xi^{-2} = C \frac{\gamma^2}{\lambda^4}. \quad (99)$$

with

$$C = \frac{\sqrt{(3-\sqrt{6})}\sqrt{3}}{2\sqrt{\pi}} e^{-\frac{(3-\sqrt{6})}{2}}. \quad (100)$$

Finally, considering

$$\sum_{S \in \mathcal{S} \setminus \{\emptyset\}} \frac{p_S}{p_k} (\alpha_S - \bar{\alpha})^2 = \text{Var}[\text{Var}(y_i)|\omega_i = 1] \quad (101)$$

and using Lemma 1, we conclude

$$\|f_k - f_k^{2\text{-GMM}}\|_K \quad (102)$$

$$\leq \frac{C' \gamma^2}{\lambda^4} \left( \frac{\sum_{i=1}^n x_i^4}{\gamma} + \sum_{i \neq j} x_i^2 x_j^2 - \frac{\|x\|^4}{p_k} \right) \quad (103)$$

with  $C' = C/4$ .

**Proof of Corollary 1**

From Theorem 5, we have

$$\|f_k - f_k^{2\text{-GMM}}\|_K \leq \frac{C\gamma^2}{\lambda^4} \left( \frac{\sum_{i=1}^n x_i^4}{\gamma} + \sum_{i \neq j} x_i^2 x_j^2 - \frac{\|x\|^4}{p_k} \right) \quad (104)$$

$$\leq C \left( \frac{\lambda_{\max}}{\lambda_{\min}} \right)^4 (\psi(k) + \psi(k)^2). \quad (105)$$

where  $\lambda_{\min} = \min_i |x_i|$  and  $\lambda_{\max} = \max_{\{i: x_i \neq 0\}} |x_i|$ . The assertion is proved with  $C' = C(\lambda_{\max}/\lambda_{\min})^4$ .

**Abbreviations**

2-GMM: Two-component Gaussian mixture model; CS: Compressed sensing; CoSaMP: Compressive sampling matching pursuit; DCT: Discrete cosine transform; EM: Expectation-maximization; ML: Maximum likelihood; MRE: Mean relative error; OMP: Orthogonal matching pursuit; SNR: Signal-to-noise ratio

**Acknowledgments**

The authors thank the European Research Council for the financial support for this research. We express our sincere gratitude to the reviewers for carefully reading the manuscript and for their valuable suggestions.

**Funding**

This work has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 279848.

**Availability of data and materials**

Please contact the corresponding author (chiara.ravazzi@ieiit.cnr.it) for data requests.

**Authors' contributions**

All the authors contributed equally, read, and approved the final manuscript.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>National Research Council of Italy, IEIT-CNR, c/o Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy. <sup>2</sup>Politecnico di Torino, DAUIN, Corso Duca degli Abruzzi 24, 10129 Torino, Italy. <sup>3</sup>Politecnico di Torino, DET, Corso Duca degli Abruzzi 24, 10129 Torino, Italy.

Received: 26 January 2018 Accepted: 19 August 2018

Published online: 10 September 2018

**References**

1. D.L. Donoho, Compressed sensing. *IEEE Trans. Inf. Theory*. **52**(4), 1289–1306 (2006)
2. E. Candes, T. Tao, Near optimal signal recovery from projection: universal encoding strategies?. *IEEE Trans. Inf. Theory*. **52**(12), 5406–5425 (2006)
3. F. Zeng, L. Chen, Z. Tian, Distributed compressive spectrum sensing in cooperative multihop cognitive networks. *IEEE J. Sel. Topics Signal Process.* **5**(1), 37–48 (2011)
4. L. Gan, in *IEEE Int. Conf. DSP*. Block compressed sensing of natural images, (2007), pp. 403–406
5. E. Arias-Castro, E.J. Candes, M.A. Davenport, On the fundamental limits of adaptive sensing. *IEEE Trans. Inf. Theory*. **59**(1), 472–481 (2013)
6. M.A. Figueiredo, R.D. Nowak, S.J. Wright, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J. Sel. Topics Signal Process.* **1**(4), 586–597 (2007)
7. A. Mousavi, A. Maleki, R.G. Baraniuk, Asymptotic analysis of LASSOs solution path with implications for approximate message passing (2013). preprint available at arXiv:1309.5979v1
8. J.A.A. Tropp, C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory*. **53**, 4655–4666 (2007)
9. D. Needell, J.A. Tropp, CoSaMP: iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmonic Anal.* **26**(3), 301–321 (2008)
10. D. Valsesia, S.M. Fosson, C. Ravazzi, T. Bianchi, E. Magli, in *2016 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. Sparsehash: embedding jaccard coefficient between supports of signals, (2016), pp. 1–6
11. R. Ward, Compressed sensing with cross validation. *IEEE Trans. Inf. Theory*. **55**(12), 5773–5782 (2009)
12. Y.C. Eldar, Generalized sure for exponential families: applications to regularization. *IEEE Trans. Signal Process.* **57**(2), 471–481 (2009)
13. SK Sharma, S. Chatzinotas, B. Ottersten, Compressive sparsity order estimation for wideband cognitive radio receiver. *IEEE Trans. Signal Process.* **62**(19), 4984–4996 (2014)
14. Y. Wang, Z. Tian, C. Feng, Sparsity order estimation and its application in compressive spectrum sensing for cognitive radios. *IEEE Trans. Wireless Commun.* **11**(6), 2116–2125 (2012)
15. D.M. Malioutov, S. Sanghavi, A.S. Willsky, in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. Compressed sensing with sequential observations, (2008), pp. 3357–3360
16. M.S. Asif, J. Romberg, in *Signals, Systems and Computers, 2008 42nd Asilomar Conference On*. Streaming measurements in compressive sensing:  $\ell_1$  filtering, (2008), pp. 1051–1058
17. P.J. Garrigues, L. El Ghaoui, in *Neural Information Processing Systems (NIPS)*. An homotopy algorithm for the Lasso with online observations, vol. 21, (2008)
18. Y. You, J. Jin, W. Duan, N. Liu, Y. Gu, J. Yang, Zero-point attracting projection algorithm for sequential compressive sensing. *IEICE Electron. Express*. **9**(4), 314–319 (2012)
19. T.V. Nguyen, T.Q.S. Quek, H. Shin, Joint channel identification and estimation in wireless network: sparsity and optimization. *IEEE Trans. Wireless Commun.* **17**(5), 3141–3153 (2018)
20. M.E. Lopes, in *Proc. 30th Int. Conf. Machine Learning*. Estimating unknown sparsity in compressed sensing, vol. 28 (Proceedings of Machine Learning Research, Atlanta, 2013)
21. M.E. Lopes, Unknown sparsity in compressed sensing: denoising and inference. *IEEE Trans. Inf. Theory*. **62**(9), 5145–5166 (2016)
22. S. Lan, Q. Zhang, X. Zhang, Z. Guo, in *IEEE Int. Symp. on Circuits and Systems (ISCAS)*. Sparsity estimation in image compressive sensing, (2012)
23. A. Agarwal, L. Flodin, A. Mazumdar, in *IEEE International Symposium on Information Theory (ISIT)*. Estimation of sparsity via simple measurements, (2017), pp. 456–460
24. Z. Bar-Yossef, T.S. Jayram, R. Kumar, D. Sivakumar, L. Trevisan, ed. by J.D.P. Rolim, S. Vadhan. Randomization and Approximation Techniques in Computer Science (RANDOM) (Springer, Berlin, 2002), pp. 1–10
25. P.B. Gibbons, ed. by M. Garofalakis, J. Gehrke, and R. Rastogi. Data Stream Management: Processing High-Speed Data Streams (Springer, Berlin, 2016), pp. 121–147
26. V. Bioglio, T. Bianchi, E. Magli, in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. On the fly estimation of the sparsity degree in compressed sensing using sparse sensing matrices, (2015), pp. 3801–3805
27. W. Wang, M.J. Wainwright, K. Ramchandran, Information-theoretic limits on sparse signal recovery: dense versus sparse measurement matrices. *IEEE Trans. Inf. Theory*. **56**(6), 2967–2979 (2010)
28. D. Omidiran, M.J. Wainwright, High-dimensional variable selection with sparse random projections: measurement sparsity and statistical efficiency. *JMLR*. **11**, 2361–2386 (2010)
29. C. Ravazzi, S.M. Fosson, T. Bianchi, E. Magli, in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Signal sparsity estimation from compressive noisy projections via  $\gamma$ -sparsified random matrices, (2016), pp. 4029–4033
30. E.J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pur. Appl. Math.* **59**(8), 1207–1223 (2006)



31. D. Omidiran, M.J. Wainwright, High-dimensional variable selection with sparse random projections: measurement sparsity and statistical efficiency. *J. Mach. Learn. Res.* **11**, 2361–2386 (2010)
32. R. Berinde, P. Indyk, M. Ruzic, in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference On*. Practical near-optimal sparse recovery in the  $\ell_1$  norm, (2008), pp. 198–205
33. A. Gilbert, P. Indyk, Sparse recovery using sparse matrices. *Proc. IEEE*. **98**(6), 937–947 (2010)
34. P. Li, C-H Zhang, in *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS) 2015*. Compressed sensing with very sparse gaussian random projections, (2006), pp. 617–625
35. E.J. Candès, *The restricted isometry property and its implications for compressed sensing*. (Compte Rendus de l'Academie des Sciences, Paris, France). ser. I **346**, 589–592 (2008)
36. M.A.T. Figueiredo, A.K. Jain, Unsupervised learning of finite mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 381–396 (2000)
37. C.F.J. Wu, On the convergence properties of the em algorithm. *Ann. Statist.* **11**(1), 95–103 (1983)
38. W. Hoeffding, Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**(301), 13–30 (1963)
39. W. Feller, *An introduction to probability theory and its applications*, vol. 1. Wiley, New York, 1968)
40. T.M. Apostol, *Calculus*, vol. 2: multi-variable calculus and linear algebra with applications to differential equations and probability. (Wiley, New York, 1967)

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)