



**HAL**  
open science

# Compressed Sensing with Basis Mismatch: Performance Bounds and Sparse-Based Estimator

Stéphanie Bernhardt, Remy Boyer, Sylvie Marcos, Pascal Larzabal

► **To cite this version:**

Stéphanie Bernhardt, Remy Boyer, Sylvie Marcos, Pascal Larzabal. Compressed Sensing with Basis Mismatch: Performance Bounds and Sparse-Based Estimator. IEEE Transactions on Signal Processing, 2016, 10.1109/TSP.2016.2544742 . hal-01313459v2

**HAL Id: hal-01313459**

**<https://hal.science/hal-01313459v2>**

Submitted on 4 Mar 2020

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

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

# Compressed Sensing with Basis Mismatch : Performance Bounds and Sparse-Based Estimator

Stéphanie Bernhardt, Rémy Boyer, Sylvie Marcos and Pascal Larzabal

**Abstract**—Compressed sensing (CS) is a promising emerging domain which outperforms the classical limit of the Shannon sampling theory if the measurement vector can be approximated as the linear combination of few basis vectors extracted from a redundant dictionary matrix. Unfortunately, in realistic scenario, the knowledge of this basis or equivalently of the entire dictionary is often uncertain, *i.e.* corrupted by a Basis Mismatch (BM) error. The consequence of the BM problem is that the estimation accuracy in terms of Bayesian Mean Square Error (BMSE) of popular sparse-based estimators collapses even if the support is perfectly estimated and in the high Signal to Noise Ratio (SNR) regime. This saturation effect considerably limits the effective viability of these estimation schemes. In the first part of this work, the Bayesian Cramér-Rao Bound (BCRB) is derived for CS model with unstructured BM. We show that the BCRB foresees the saturation effect of the estimation accuracy of standard sparse-based estimators as for instance the OMP, Cosamp or the BP. In addition, we provide an approximation of this BMSE threshold. In the second part and in the context of the structured BM model, a new estimation scheme called Bias-Correction Estimator (BiCE) is proposed and its statistical properties are studied. The BiCE acts as a post-processing estimation layer for any sparse-based estimator and mitigates considerably the BM degradation. Finally, the BiCE (*i*) is a blind algorithm, *i.e.*, is unaware of the uncorrupted dictionary matrix, (*ii*) is generic since it can be associated to any sparse-based estimator, (*iii*) is fast, *i.e.*, the additional computational cost remains low and (*iv*) has good statistical properties. To illustrate our results and propositions, the BiCE is applied in the challenging context of the compressive sampling of non-bandlimited impulsive signals.

## I. INTRODUCTION

Compressed Sensing (CS) [1]–[3] is a challenging domain that has driven a lot of research interests in a wide panel of applications as for instance Ultra-Sound Imaging [4], MRI imaging [5], channel estimation [6], array processing [7,8], RADAR processing [9], *etc.* CS potentially outperforms the limit sampling rate predicted by the classical sampling theory [10]. This is done by exploiting the a priori knowledge that many natural measurement signals admit a sparse representation, *i.e.*, as a linear combination of few vectors extracted from a redundant dictionary. In today's world that has to face more and more data, this technique has opened many perspectives on

a wide panel of application domains. The CS framework has thus driven the design of a plethora of sparse-based estimation algorithms as for instance the OMP [11], Cosamp [12], BP [13], LASSO [14].

In the CS framework, we aim to recover the measurement signal with a sparse representation from a small number of linear measurements extracted from a known redundant dictionary [1]–[3]. Note that a similar philosophy can be found in the frame based theory [15]. The major difference is in the design of the dictionary. While in the frame based theory, the dictionary is viewed as a collection of deterministic atoms, the dictionary involved in the CS framework is the product of a random overcomplete measurement/sensing matrix and a deterministic unitary/orthonormal basis matrix [2]. Given a measurement vector, the inverse problem does not admit a unique solution and the prominent idea is to select the sparsest vector as the solution. On the contrary, the unitary basis in CS and the redundant dictionary in the frame based theory are based on a similar sampling process of a given continuous kernel.

Unfortunately, in operational context, the basis matrix is often known with uncertainty, *i.e.*, there exists a Basis Mismatch (BM) effect. As noted in the literature, standard sparse-based estimators become inefficient since their MSEs are saturated even if the SNR is high and if the support set has been perfectly estimated. This situation is a severe drawback in real-world operational context. This non-ideal scenario has been often observed in practice [16]–[19] and the exact sparse recovery guarantees for the estimation of a signal corrupted by a BM has been studied in [20]. In particular, in a Tx (Transmit)-Rx (Receipt) context, the Rx is usually unaware (partially or not) of the true dictionary used at the Tx. As a consequence, a BM degradation may occur at the Rx-side as for instance a grid mismatching. A common but not satisfactory strategy to mitigate this problem is to transmit the entire dictionary as a side information. This is the case for example in digital data hiding application [21,22] and compressed wireless communication [23,24], where the Rx needs to know perfectly the dictionary to estimate the secret message at the Rx-side with a minimal error probability. In the context of robust data hiding [25], the dictionary matrix is thus periodically regenerated and transmitted to the Rx as a side information acting as an encryption key. This is of course done at the price of a heavy extra transmission cost. As a partial conclusion, there is a need to find a new strategy to mitigate the BM problem.

There exist two main approaches of the BM problem differing by the assumptions on the perturbation matrix. For instance

Stéphanie Bernhardt, Rémy Boyer and Sylvie Marcos are with Université Paris-Sud/CNRS in Laboratoire des Signaux et Systèmes (L2S) and Pascal Larzabal is with Université Paris-Sud/SATIE (ENS-Cachan).

Rémy Boyer and Pascal Larzabal are involved in the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM# (Grant agreement no. 318306) and in the following projects: MAGELLAN (ANR-14-CE23-0004-01), MI-CNRS TITAN and Icode projects.

Some results of this work have been partially exposed during the workshop CAMSAP'15

in [26], the Gaussian Error in Variables (EIV) model is adopted meaning that the perturbation matrix is assumed to follow a matrix Gaussian distribution. This model is weakly structured or unstructured and thus is rather generic. In our view, this context is well adapted to the derivation of fundamental and theoretical performance limits. The second BM modelization is based on the assumption that the perturbation matrix is sufficiently small to approximate the corrupted dictionary by a first-order Taylor expansion [27]–[29]. This modelization is highly structured but also more delicate to use. But, this underlying structure is well adapted to derive practical estimators. In the first part of this work, we propose the derivation and the analysis of the Bayesian lower bound in the context of unstructured BM. In the CS framework (without BM problem), the utility of the deterministic CRB in the Genie-Aided or oracle context, *i.e.*, when the support set is perfectly known, has been investigated and demonstrated in several publications [30]–[32]. Unlike these contributions, we assume that our CS model is corrupted by a BM degradation. In [33], the Bayesian lower bound in the specific case of direction of arrival estimation is derived in case of structured BM. Our Bayesian lower bound, taking into account data-dependence on the noise, provides new insights into the MSE saturation. The second part is dedicated to the proposition of a new estimator, called Bias-Correction Estimator (BiCE), which improves the robustness to BM of any sparse-based estimators in the structured BM context. In addition, the BiCE is applied in the important context of the non-bandlimited signal compressive sampling.

The paper is structured as the following. The section II presents an introduction of the CS theory and the BM in case of structured and unstructured models. Section III derives the Bayesian CRB (BCRB) for unstructured BM and gives an analytical BCRB for the BMSE saturation. Section IV introduces the BiCE and presents a complete statistical performance analysis. Section V applies our results in the context of compressive sampling of non-bandlimited signals and the last section concludes the paper.

The notation used through this paper is the following: scalars, vectors, and matrices are represented by italic lower-case, boldface lower-case and boldface upper-case, respectively. Sets are denoted by calligraphic upper-case symbols, *e.g.*,  $\mathcal{X}$ . The symbols  $(\cdot)^T$ ,  $(\cdot)^{-1}$ ,  $(\cdot)^\dagger$  and  $\text{Tr}(\cdot)$  denote the transpose, the inverse, the pseudo-inverse and the trace operator, respectively. Define  $|\mathcal{X}|$  as the cardinality of set  $\mathcal{X}$ , then vector  $\mathbf{x}_{\mathcal{X}}$  is the  $|\mathcal{X}| \times 1$  vector with  $\{x_i, i \in \mathcal{X}\}$ . In the same spirit, matrix  $\mathbf{X}_{\mathcal{S}}$  is the matrix  $\mathbf{X}$  composed by the columns indexed by  $i \in \mathcal{X}$ .  $\|\cdot\|_0$  stands for the pseudo-norm  $l_0$ , without indication the 2-norm is considered. Distribution  $\mathbf{Q} \sim \mathcal{MG}(\mathbf{M}, \mathbf{U}, \mathbf{V})$  stands for a Matrix Gaussian distribution where  $\mathbf{M}$  is the location matrix,  $\mathbf{U}$  and  $\mathbf{V}$  are the scale positive definite matrices. The vector computed by the vectorization of matrix is  $\text{vec}\mathbf{Q} \sim \mathcal{N}(\text{vec}\mathbf{M}, \mathbf{V} \otimes \mathbf{U})$  where  $\otimes$  is the the Kronecker product and  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  is the (vector) Gaussian distribution of mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . The Generalized Gaussian distribution is denoted by  $\mathcal{GN}(\mu, \alpha, \beta)$  where  $\mu$ ,  $\alpha$  and  $\beta$  is location, scale and shape parameters, respectively. The distribution  $\text{Inv}-\chi_n^2$  stands for the central

inverse chi-squared distribution with  $n$  degrees of freedom.  $\Gamma(\cdot)$  denotes the gamma function. The symbol  $\partial$  denotes the partial derivative operation and we note the first-order (*resp.* second-order) derivative of matrix  $\mathbf{M}$  by  $\dot{\mathbf{M}}$  (*resp.* by  $\ddot{\mathbf{M}}$ ). The symbol  $\text{diag}(\mathbf{x})$  denotes a diagonal matrix, where the elements of the vector  $\mathbf{x}$  specify its diagonal elements. The big-O notation is given by  $O(\cdot)$ .  $\text{Var}(x)$  is the variance of  $x$ . Furthermore,  $\langle \mathbf{X} \rangle$  is the space spanned by  $\mathbf{X}$  and  $\langle \mathbf{X} \rangle^\perp$  is its orthogonal space. Finally, the range and the null spaces of  $\mathbf{X}$  are defined by  $\text{R}(\mathbf{X})$  and  $\text{N}(\mathbf{X})$ , respectively.

## II. COMPRESSED SENSING (CS) WITH BM

### A. The CS framework

Let  $\mathbf{y}$  be a  $N \times 1$  noisy measurement vector that follows the Compressed Sensing (CS) model [2,3] according to

$$\mathbf{y} = \boldsymbol{\Psi}\mathbf{s} + \mathbf{n} \quad (1)$$

where  $\mathbf{n}$  is a centered circular Gaussian white noise of unknown covariance matrix  $\sigma^2\mathbf{I}$  and  $\boldsymbol{\Psi}$  is the  $N \times K$  sensing/measurement matrix. Let  $\mathbf{s}$  be defined as  $\mathbf{s} \stackrel{\text{def.}}{=} \boldsymbol{\Phi}\boldsymbol{\theta}$  where matrix  $\boldsymbol{\Phi}$  is a  $K \times K$  orthonormal basis and  $\boldsymbol{\theta}$  is a  $L$ -sparse amplitude vector of size  $K \times 1$ . The CS model (1) is given by

$$\mathbf{y} = \boldsymbol{\Psi}\boldsymbol{\Phi}\boldsymbol{\theta} + \mathbf{n} \stackrel{\text{def.}}{=} \mathbf{H}\boldsymbol{\theta} + \mathbf{n} \quad (2)$$

where  $\mathbf{H}$  is usually called the  $N \times K$  dictionary matrix.

In classical sampling theory, the number of measurement  $N$  needed to ensure that there is no loss of information should be at least equal to  $K$ , where  $K$  is given by the Nyquist rate. In contrast, CS theory reaches this goal for  $N \ll K$  as long as the  $K \times 1$  amplitude vector  $\boldsymbol{\theta}_{\mathcal{S}}$  is sparse in a given dictionary  $\boldsymbol{\Phi}$  [1] (as for instance, the canonical dictionary of  $\mathbb{R}^K$ , Fourier dictionary). CS theory allows to solve the ill-posed problem where the dictionary  $\mathbf{H}$  is an overcomplete/redundant matrix. A fundamental question in CS is how many measurements  $N$ , are required to enable the recovery of the amplitude vector  $\boldsymbol{\theta}$ . To answer to this question, the properties of the dictionary matrix  $\mathbf{H}$  are fundamental to guarantee satisfactory performance of CS based estimator. In [2,3,34], the authors introduced the Restricted Isometry Property (RIP) of a matrix and established its important role in CS. First define the set of the  $L$ -sparse vectors by  $\mathcal{W}_L = \{\boldsymbol{\theta} \in \mathbb{R}^K, \|\boldsymbol{\theta}\|_0 = |\mathcal{S}| \leq L\}$  where  $\mathcal{S}$  is the support set *the location of the non-zero entries of vector*  $\boldsymbol{\theta}$ . We say that matrix  $\mathbf{H}$  verifies the RIP of order  $L$  if there exists a constant  $\epsilon_L \in (0, 1)$  such as

$$\|\boldsymbol{\theta}\|^2 (1 - \epsilon_L) \leq \|\mathbf{H}\boldsymbol{\theta}\|^2 \leq \|\boldsymbol{\theta}\|^2 (1 + \epsilon_L) \quad (3)$$

for any vector  $\boldsymbol{\theta} \in \mathcal{W}_L$ . The RIP characterizes nearly isometric matrix  $\mathbf{H}$  for the  $L$ -sparse vector  $\boldsymbol{\theta}$ . In addition, the RIP ensures that practical algorithms using  $l_1$  norm optimization can successfully recover any compressible amplitude vector from noisy measurements.

First, note that direct derivation of a matrix  $\mathbf{H}$  verifying the RIP is a combinatorial problem since all the possible combinations of  $L$  nonzero entries in the vector  $\boldsymbol{\theta}$  have to be tested. At first glance, it seems hard to find matrix  $\mathbf{H}$ . Fortunately, simple and efficient strategies have been proposed (see [1,34] for instance) to circumvent this problem. A convenient strategy

to satisfy the RIP is to consider the entries of the measurement matrix  $\Psi$  as a sub-Gaussian independent and identically distributed (i.i.d.), with zero mean and finite variance. This choice ensures that the entries of matrix  $\mathbf{H}$  will be sub-Gaussian, i.i.d., with zero mean and finite variance for an orthogonal basis  $\Phi$ . The class of sub-Gaussian p.d.f. satisfies the RIP with high probability since the random quantity  $\|\mathbf{H}\theta\|^2$  is concentrated around  $\|\theta\|^2$  with high probability [35].

Finally, the CS model in (2) restricted to the support takes the following expression:

$$\mathbf{y} = \mathbf{H}_S \theta_S + \mathbf{n} \quad \text{with} \quad \mathbf{H}_S \stackrel{\text{def.}}{=} \Psi \Phi_S \quad (4)$$

where  $\Phi_S$  is a  $K \times L$  partial basis matrix. The rank- $L$  dictionary  $\mathbf{H}_S$  is viewed as a single realization of a given p.d.f. and thus will be considered in the sequel as a deterministic matrix.

### B. The Basis Mismatch (BM) problem

In the literature, we can find two types of modelization adapted to the BM problem. So, in the two next sections, we define these two approaches.

1) *Unstructured BM*: The BM problem arises when we have an uncertain knowledge of the true dictionary  $\mathbf{H}$ . More precisely, in the BM problem, we have the knowledge of a corrupted/erroneous dictionary  $\tilde{\mathbf{H}}$  while the  $L$ -sparse signal admits a decomposition into few vectors belonging to the unknown deterministic dictionary  $\mathbf{H}$ . This error can occur during the noisy transmission of the dictionary between a transmitter and a receiver, or if the construction of dictionary  $\tilde{\mathbf{H}}$  is based on a different discretization of the parameter set. As illustrated in section V, sparse-based estimators (OMP, LASSO, BP, Cosamp, ...) exhibit a saturated estimation accuracy even if the noise variance is low and even worse if the support has been perfectly estimated.

A generic error modelization is the Gaussian EIV model [26] described by

$$\tilde{\mathbf{H}} = \mathbf{H} + \mathbf{E} \quad \sim \quad \mathcal{MG}(\mathbf{H}, \sigma_e^2 \mathbf{I}_N, \mathbf{I}_K). \quad (5)$$

Note that the BM problem described on Fig. 1, is slightly different that the well-known off-grid (OG) problem and we can find in [36] a comparison of these two types of uncertainty. We can describe the two problems in the following manner:

- BM problem :  $\mathbf{y} = \mathbf{H}\theta + \mathbf{n} = \tilde{\mathbf{H}}\theta - \mathbf{E}\theta + \mathbf{n}$ ,
- $\mathbf{H}$  is unknown,
  - $\tilde{\mathbf{H}}$  is known and stochastic (due to  $\mathbf{E}$ ).
- OG problem :  $\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\theta + \mathbf{n} = \mathbf{H}\theta + \mathbf{E}\theta + \mathbf{n}$ ,
- $\mathbf{H}$  is known and non-stochastic,
  - $\tilde{\mathbf{H}}$  is unknown.

So, in the OG problem, it is assumed that we know dictionary  $\mathbf{H}$  and the  $L$ -sparse signal admits a decomposition into few vectors belonging to an unknown corrupted dictionary  $\tilde{\mathbf{H}}$ .

At this point we can formulate two remarks. Firstly, note that

in the two above models, the noise term  $\mathbf{E}\theta + \mathbf{n}$  is data-dependent due to the BM  $\mathbf{E}$  and secondly, the BM context changes drastically the underlying estimation problem. Indeed, recall that dictionary  $\mathbf{H}$  is a fixed and unknown matrix in the point of view of the considered estimation problem. However, assuming the mismatched context implies that the known corrupted dictionary,  $\tilde{\mathbf{H}}$  turns to be now a random matrix due to the stochastic nature of its entries.

2) *Structured BM*: A well-known family of sparse promoting estimators are based on a basis matrix  $\Phi$  which is designed by the regular/uniform discretization at rate  $T_S$  of a known continuous kernel  $\phi(t)$ . This kernel is dependent on the considered framework.

In the BM framework, the true/uncorrupted discretization of the parameter set denoted by  $\{\tau_1, \dots, \tau_K\}$  is unavailable to our knowledge. Equivalently, this also means that the uncorrupted dictionary  $\mathbf{H}$  based on the true basis matrix, defined by  $[\Phi]_{kk'} = \phi(t)|_{t=\tau_{k'}-T_S k}$  where  $1 \leq k, k' \leq K$ , is also unknown to our knowledge. Now assume that the mismatch on the basis matrix, denoted by  $\tilde{\Phi}$ , comes from an uncertain but known discretization of the kernel  $\phi(t)$ , defined by  $\{\tilde{\tau}_1, \dots, \tilde{\tau}_K\}$ . Specifically, we observe the corrupted parameter  $\tilde{\tau}_k = \tau_k + e_k$  while  $e_k$  is an unknown error term corrupting the unknown parameter  $\tau_k$ .

A Taylor expansion of the known corrupted basis  $\tilde{\Phi}$  is given by

$$\begin{aligned} [\tilde{\Phi}]_{kk'} &= \phi(t)|_{t=\tilde{\tau}_{k'}-T_S k} \\ &= \phi(t)|_{t=\tau_{k'}-T_S k} + \left. \frac{\partial \phi(t)}{\partial t} \right|_{t=\tau_{k'}-T_S k} e_{k'} + O(e_{k'}^2). \end{aligned}$$

So, for a sufficiently small BM, we assume that the known corrupted basis  $\tilde{\Phi}$  can be approximated by a first-order Taylor expansion given by

$$\tilde{\Phi}_S \approx \Phi_S + \dot{\Phi}_S \text{diag}(\mathbf{e}_S)$$

where  $[\dot{\Phi}_S]_{kk'} = \left. \frac{\partial \phi(t)}{\partial t} \right|_{t=\tau_{k'}-T_S k}$  and  $\mathbf{e}_S = [\dots e_{k'}, \dots]^T$  where  $k' \in \mathcal{S}$ . In term of the dictionary, we have the following relations:

$$\tilde{\mathbf{H}}_S = \Psi \tilde{\Phi}_S \approx \Psi \Phi_S + \Psi \dot{\Phi}_S \text{diag}(\mathbf{e}_S) = \mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S). \quad (6)$$

This structured BM modelization has been largely adopted in the literature (see [26] for instance and the references therein).

## III. BAYESIAN LOWER BOUNDS FOR UNSTRUCTURED BM

In this section, we present the VanTrees' Bayesian lower bound for the estimation of a sparse random vector under BM. In addition, we propose an analytical expression of this bound that well predicts the saturated BMSE value when the noise is dominated by the BM effect.

### A. The VanTrees' Bayesian bound

In this section, we derive the VanTrees' Bayesian bound [37, 38] as a benchmark against which any Bayesian estimator can be compared. The bound informs us about the smallest BMSE

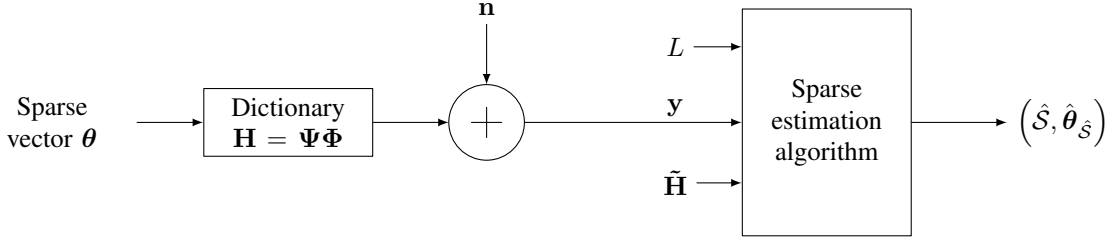


Fig. 1. Sparse estimation with BM

achievable. Note that the proposed derivation of the Fisher Information Matrix (FIM) is similar to the one proposed in [39]. But our main concern here is to generalize the analysis in the Bayesian context and to analyze this bound in the context of the CS framework. Let  $\hat{\boldsymbol{\theta}}$  be an estimate of  $\boldsymbol{\theta}_S$ , for which the BMSE is defined by :

$$\text{BMSE}_S = \mathbb{E}_{\mathbf{y}, \tilde{\mathbf{H}}_S, \boldsymbol{\theta}_S} \left\| \boldsymbol{\theta}_S - \hat{\boldsymbol{\theta}}_S(\mathbf{y}, \tilde{\mathbf{H}}_S) \right\|^2 \geq \text{Tr}[\mathbf{B}_S] \quad (7)$$

where for  $1 \leq i, j \leq L$ ,

$$[\mathbf{B}_S^{-1}]_{ij} = \left[ \text{Var} \left( \frac{\partial \log p(\mathbf{y}, \tilde{\mathbf{H}}_S, \boldsymbol{\theta}_S)}{\partial \boldsymbol{\theta}_S} \right) \right]_{ij}. \quad (8)$$

The log-joint pdf is given by  $\log p(\mathbf{y}, \tilde{\mathbf{H}}_S, \boldsymbol{\theta}_S) = \log p(\mathbf{y} | \tilde{\mathbf{H}}_S, \boldsymbol{\theta}_S) + \log p(\tilde{\mathbf{H}}_S) + \log p(\boldsymbol{\theta}_S)$ .

From model (4) and (5), we can write that

$$\mathbf{y} = \tilde{\mathbf{H}}_S \boldsymbol{\theta}_S - \mathbf{E}_S \boldsymbol{\theta}_S + \mathbf{n}. \quad (9)$$

Recalling that  $\mathbf{E}_S$  follows a matrix Gaussian distribution of variance  $\sigma_e^2 \mathbf{I}$ , and that the noise is also Gaussian of variance  $\sigma^2$ , the conditional observation distribution is given by

$$\mathbf{y} | \tilde{\mathbf{H}}_S, \boldsymbol{\theta}_S \sim \mathcal{N} \left( \tilde{\mathbf{H}}_S \boldsymbol{\theta}_S, (\sigma_e^2 \|\boldsymbol{\theta}_S\|^2 + \sigma^2) \mathbf{I} \right). \quad (10)$$

Furthermore  $\partial \log p(\tilde{\mathbf{H}}_S) / \partial \boldsymbol{\theta}_S$  vanishes [39] and thus the trace of the BCRB matrix is given by

$$\text{BCRB}_{\text{BM}} = \text{Tr}[\mathbf{B}_S] = \text{Tr} \left[ \left( \mathbb{E}_{\boldsymbol{\theta}_S, \tilde{\mathbf{H}}_S} (\mathbf{F}_S) + \mathbf{P}_{\boldsymbol{\theta}_S} \right)^{-1} \right] \quad (11)$$

where  $\mathbf{P}_{\boldsymbol{\theta}_S} = \text{Var} \left( \frac{\partial \log p(\boldsymbol{\theta}_S)}{\partial \boldsymbol{\theta}_S} \right)$  and the FIM  $\mathbf{F}_S$  is given by the Slepian-Bang formula according to

$$\mathbf{F}_S = \frac{\tilde{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S}{\sigma_e^2 \|\boldsymbol{\theta}_S\|^2 + \sigma^2} + \frac{2\sigma_e^4 N \boldsymbol{\theta}_S \boldsymbol{\theta}_S^T}{(\sigma_e^2 \|\boldsymbol{\theta}_S\|^2 + \sigma^2)^2}. \quad (12)$$

Noting that  $\boldsymbol{\theta}_S$  and  $\tilde{\mathbf{H}}_S$  are two multidimensional independent processes, we have

$$\mathbb{E}_{\boldsymbol{\theta}_S, \tilde{\mathbf{H}}_S} (\mathbf{F}_S) = \mathbb{E}_{\boldsymbol{\theta}_S} \left( \frac{\mathbb{E}_{\tilde{\mathbf{H}}_S} (\tilde{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S)}{\sigma_e^2 \|\boldsymbol{\theta}_S\|^2 + \sigma^2} \right) \quad (13)$$

$$+ \mathbb{E}_{\boldsymbol{\theta}_S} \left( \frac{2\sigma_e^4 N \boldsymbol{\theta}_S \boldsymbol{\theta}_S^T}{(\sigma_e^2 \|\boldsymbol{\theta}_S\|^2 + \sigma^2)^2} \right) \quad (14)$$

where  $\mathbb{E}_{\tilde{\mathbf{H}}_S} (\tilde{\mathbf{H}}_S^T \tilde{\mathbf{H}}_S) = \mathbf{H}_S^T \mathbf{H}_S + \sigma_e^2 \mathbf{I}$ . Finally, we obtain

$$\text{BCRB}_{\text{BM}} = \text{Tr} \left[ \left( \left( \mathbf{H}_S^T \mathbf{H}_S + \sigma_e^2 \mathbf{I} \right) \mathcal{P}(\sigma_e^2, \sigma^2) + \boldsymbol{\Sigma}(\sigma_e^2, \sigma^2) + \mathbf{P}_{\boldsymbol{\theta}_S} \right)^{-1} \right] \quad (15)$$

where

$$\mathcal{P}(\sigma_e^2, \sigma^2) = \mathbb{E}_{\boldsymbol{\theta}_S} \frac{1}{\sigma_e^2 \|\boldsymbol{\theta}_S\|^2 + \sigma^2}, \quad (16)$$

$$\boldsymbol{\Sigma}(\sigma_e^2, \sigma^2) = \mathbb{E}_{\boldsymbol{\theta}_S} \frac{2\sigma_e^4 N \boldsymbol{\theta}_S \boldsymbol{\theta}_S^T}{(\sigma_e^2 \|\boldsymbol{\theta}_S\|^2 + \sigma^2)^2}. \quad (17)$$

*B. Closed-form expressions of the saturated BMSE for i.i.d. Gaussian amplitudes*

We now make the following assumptions to derive a closed-form expression of the saturated BCRB, denoted by  $\text{BCRB}_{\text{BM}}^{\text{sat}}$ , based on the BCRB given in (15):

$\mathcal{A}_1$ . We consider the low noise variance regime since it is well known that the BM problem appears only in the regime where the noise variance is sufficiently low to be dominated by  $\sigma_e^2$ . This fact directly implies that the BMSE of any sparse-based estimator in this context is lower bounded by  $\text{BCRB}_{\text{BM}}^{\text{sat}}$  and thus cannot be statistically efficient. As this property is a highly desired feature in the context of the estimation theory, the reader can measure the importance of the analysis of this context. At contrary, when the noise variance is high with respect to  $\sigma_e^2$ , the uncertainty context is dominated by the error due to the noise and cannot be measured and thus can be ignored.

$\mathcal{A}_2$ . The last assumption is to assume that the  $L$  amplitudes belonging to the support, *i.e.* taking non-zeros values, follow an i.i.d. centered Gaussian distribution with variance  $\sigma_\theta^2/L$ . So, the sparse amplitude vector, denoted by  $\boldsymbol{\theta}$ , is composed by  $K-L$  zeros-values and  $L$  random non-zero amplitudes. In this case, the prior matrix of the BIM is given by  $\mathbf{P}_{\boldsymbol{\theta}_S} = \frac{L}{\sigma_\theta^2} \mathbf{I}_L$ .

*Lemma 3.1:* In the low noise variance regime, we have

$$\lim_{\sigma^2 \rightarrow 0} \mathcal{P}(\sigma_e^2, \sigma^2) = \frac{L}{(L-2)\sigma_e^2 \sigma_\theta^2}, \quad (18)$$

$$\lim_{\sigma^2 \rightarrow 0} \boldsymbol{\Sigma}(\sigma_e^2, \sigma^2) = \frac{2N}{(L-2)\sigma_\theta^2} \mathbf{I}_L. \quad (19)$$

**Proof** We first prove (18). For centered i.i.d. Gaussian amplitudes of variance  $\sigma_\theta^2$ , we have  $1/\|\boldsymbol{\theta}_S\|^2 \sim \text{Inv} - \chi_L^2$ . Thus, for  $L > 2$ ,  $\mathbb{E}_{\boldsymbol{\theta}_S} \frac{1}{\|\boldsymbol{\theta}_S\|^2} = \frac{L}{\sigma_\theta^2(L-2)}$ . Using the fact that  $\lim_{\sigma^2 \rightarrow 0} \mathcal{P}(\sigma_e^2, \sigma^2) = \frac{1}{\sigma_e^2} \mathbb{E}_{\boldsymbol{\theta}_S} \frac{1}{\|\boldsymbol{\theta}_S\|^2}$ , we obtain expression (18).

To prove (19), it is easy to see that  $\mathbb{E}_{\boldsymbol{\theta}_S} \left( \frac{\boldsymbol{\theta}_S \boldsymbol{\theta}_S^T}{\|\boldsymbol{\theta}_S\|^4} \right)$  is a diagonal matrix proportional to the identity matrix, *i.e.*  $\mathbb{E}_{\boldsymbol{\theta}_S} \left( \frac{\boldsymbol{\theta}_S \boldsymbol{\theta}_S^T}{\|\boldsymbol{\theta}_S\|^4} \right) = t \mathbf{I}_L$ . Note that  $\text{Tr} \left[ \mathbb{E}_{\boldsymbol{\theta}_S} \left( \frac{\boldsymbol{\theta}_S \boldsymbol{\theta}_S^T}{\|\boldsymbol{\theta}_S\|^4} \right) \right] =$

$\mathbb{E}_{\theta_S} \left( \frac{1}{\|\theta_S\|^2} \right) = \frac{L}{(L-2)\sigma_\theta^2}$ . Thus, we obtain  $t = \frac{1}{(L-2)\sigma_\theta^2}$ . Using the fact that  $\lim_{\sigma_e^2 \rightarrow 0} \Sigma(\sigma_e^2, \sigma^2) = 2N\mathbb{E}_{\theta_S} \left( \frac{\theta_S \theta_S^T}{\|\theta_S\|^4} \right)$ , we obtain expression (19).  $\square$

An approximated lower bound for the BMSE saturated is given in the following result.

*Result 3.2:* For i.i.d. Gaussian amplitude vector, for large  $N$  and  $L$  of same order of magnitude and for  $\sigma_e^2$  small with respect to  $O(1/L)$ , the BMSE of any sparse based estimators is lower bounded by

$$\text{BCRB}_{\text{BM}}^{\text{sat}} \approx \mathcal{H}_{N,L} \sigma_\theta^2 \sigma_e^2. \quad (20)$$

where  $\mathcal{H}_{N,L} = \frac{NL}{N-L}$ .

**Proof** Using the above Lemma, we obtain

$$\begin{aligned} \text{BCRB}_{\text{BM}}^{\text{sat}} &\approx \frac{\sigma_\theta^2 \sigma_e^2 (L-2)}{L} \\ &\times \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{H}_S + \sigma_e^2 \left( 2\frac{N}{L} - 1 + L \right) \mathbf{I} \right)^{-1} \right]. \end{aligned} \quad (21)$$

For large  $L$  and  $N$  of same order of magnitude,  $2\frac{N}{L} - 1 + L \approx O(L)$ . So, for  $\sigma_e^2$  small with respect to  $O(1/L)$ , a Taylor approximation is

$$\text{BCRB}_{\text{BM}}^{\text{sat}} \approx \sigma_\theta^2 \sigma_e^2 \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{H}_S \right)^{-1} \right] \quad (22)$$

Random Matrix Theory [40,41] provides some powerful (doubly) asymptotic, *i.e.* for  $N, L \rightarrow \infty$  with  $N/L \rightarrow c$  (meaning that  $L$  and  $N$  are of same order of magnitude), closed-forms for the empirical moments. More precisely, we have  $\lim_{N,L \rightarrow \infty} \frac{1}{N} \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{H}_S \right)^{-1} \right] \rightarrow \frac{1}{c-1}$  in almost sure convergence. In this work, this (doubly) asymptotic context is not assumed. But, interestingly, the above limit for finites  $N$  and  $L$  of same order of magnitude remains an accurate numerical closed-form approximations given by  $\text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{H}_S \right)^{-1} \right] \approx \frac{NL}{N-L} = \mathcal{H}_{N,L}$ . This property is illustrated on Fig. 2. Consequently, using it with (22), we obtain (20).  $\square$

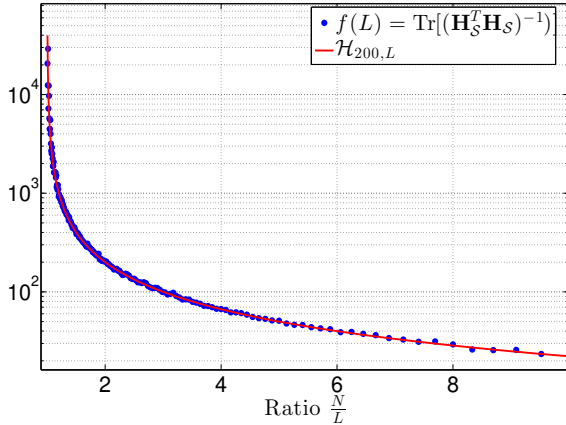


Fig. 2. Approximation of  $f(L)$  by  $\mathcal{H}_{N,L}$  (involved in the proof of Result 3.2) for large  $N$ , Vs. ratio  $N/L$  with  $\sigma^2 = 10^{-10}$ ,  $K = 300$  and  $20 \leq L \leq N-1$  where  $N = 200$ .

## IV. BiCE FOR A STRUCTURED BM

In this section, the BiCE is introduced as an universal solution to improve any sparse-based estimator under BM. A complete statistical analysis of the BiCE is also provided.

### A. Principles and description of the BiCE

The BiCE is designed as an efficient and computational attractive post-processing estimator which is able to enhance any sparse-based recovery estimator, denoted by  $\hat{\theta}(\mathbf{y})$ , in case of a structured BM.

In the BM problem (see section II.B), the corrupted dictionary  $\tilde{\mathbf{H}} = \Psi \tilde{\Phi}$  is available to our knowledge and is designed based on a known but erroneous discretization parametrized by  $\{\tilde{\tau}_1, \dots, \tilde{\tau}_K\}$  of a continuous kernel  $\phi(t)$ . As a consequence, matrix  $[\tilde{\Phi}]_{kk'} = \phi(\tilde{\tau}_{k'} - T_S k)$  is known but corrupted by an BM error with respect to the true but unknown discretization  $\{\tau_1, \dots, \tau_K\}$ . The initial idea of the BiCE is to remark that as  $\tilde{\Phi}$  is available to us, its first-order derivative, given by  $[\dot{\tilde{\Phi}}]_{kk'} = \frac{\partial \phi(t)}{\partial t} \Big|_{t=\tilde{\tau}_{k'} - T_S k}$ , is also known. So, the first-order

derivative of the corrupted dictionary given by  $\dot{\tilde{\mathbf{H}}} = \Psi \dot{\tilde{\Phi}}$  can be exploited jointly with the corrupted dictionary  $\tilde{\mathbf{H}}$ . For a given support, the subspaces  $\langle \tilde{\mathbf{H}}_S \rangle$  and  $\langle \dot{\tilde{\mathbf{H}}}_S \rangle$  are available to us and completely characterize the following range and null subspaces:

$$\mathbf{R} \left( \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\tilde{\mathbf{H}}}_S} \right) = \langle \tilde{\mathbf{H}}_S \rangle \quad \text{and} \quad \mathbf{N} \left( \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\tilde{\mathbf{H}}}_S} \right) = \langle \dot{\tilde{\mathbf{H}}}_S \rangle \quad (23)$$

where  $\mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\tilde{\mathbf{H}}}_S}$  stands for of an oblique projector [42] defined by

$$\mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\tilde{\mathbf{H}}}_S} = \tilde{\mathbf{H}}_S \left( \tilde{\mathbf{H}}_S^T \mathbf{P}_{\dot{\tilde{\mathbf{H}}}_S}^\perp \tilde{\mathbf{H}}_S \right)^{-1} \tilde{\mathbf{H}}_S^T \mathbf{P}_{\dot{\tilde{\mathbf{H}}}_S}^\perp \quad (24)$$

in which  $\mathbf{P}_{\dot{\tilde{\mathbf{H}}}_S}^\perp$  is the orthogonal projector on subspace  $\langle \dot{\tilde{\mathbf{H}}}_S \rangle^\perp$ . The BiCE is based on the two following steps.

- 1) Oblique projection of the measurement vector according to

$$\mathbf{y}_S = \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\tilde{\mathbf{H}}}_S} \mathbf{y}. \quad (25)$$

- 2) Minimization over  $\theta_S$  of the following standard least squares criterion:

$$\Gamma(\theta_S) = \|\mathbf{y}_S - \tilde{\mathbf{H}}_S \theta_S\|^2 \quad (26)$$

$$= \|\mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\tilde{\mathbf{H}}}_S} \mathbf{y} - \tilde{\mathbf{H}}_S \theta_S\|^2. \quad (27)$$

We will show in the sequel that the BM effect is mitigated thanks to the considered oblique projection. Specifically, the BiCE associated with a sparse-based estimator is described on Fig. 3 and in Algorithm 1.

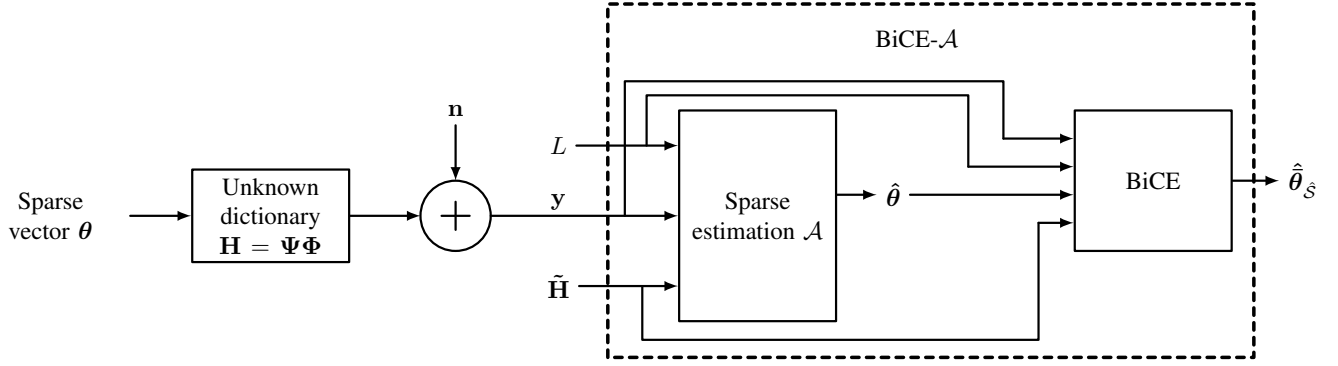


Fig. 3. Description of the BiCE-A. Note that algorithm  $\mathcal{A}(\cdot)$  stands for any sparse-based estimators.

---

**Algorithm 1** BiCE- $\mathcal{A}(\mathbf{y}, \tilde{\mathbf{H}}, L, \phi(t))$ 


---

**Require:**  $\mathbf{y}, \tilde{\mathbf{H}}, L, \phi(t), \mathcal{A}(\cdot)$

**Ensure:**  $\hat{\theta}_{\mathcal{S}}$

- 1: Run the sparse-based estimator *i.e.*,

$$\mathcal{A}(\mathbf{y}, \tilde{\mathbf{H}}, L, \phi(t)) \rightarrow \hat{\theta}. \quad (28)$$

- 2: Compute  $\hat{\mathbf{H}} = \Psi \hat{\Phi}$

- 3: From  $\hat{\theta}$  compute  $\hat{\mathcal{S}}$

- 4: Using  $\hat{\mathcal{S}}$ , compute  $\tilde{\mathbf{H}}_{\mathcal{S}}$  and  $\hat{\mathbf{H}}_{\mathcal{S}}$

- 5: Compute  $\mathbf{E}_{\tilde{\mathbf{H}}_{\mathcal{S}} \hat{\mathbf{H}}_{\mathcal{S}}}$  defined in (24)

- 6: Minimize  $\Gamma(\hat{\theta}_{\mathcal{S}})$  defined by (27) according to

$$\hat{\theta}_{\mathcal{S}} = \tilde{\mathbf{H}}_{\mathcal{S}}^{\dagger} \mathbf{E}_{\tilde{\mathbf{H}}_{\mathcal{S}} \hat{\mathbf{H}}_{\mathcal{S}}} \mathbf{y}. \quad (29)$$


---

### B. Analytic expressions of the bias vector and the BMSE

*Result 4.1:* For a perfect support estimation and for sufficiently small  $\sigma_e^2 \sigma_e^2$ , the conditional bias vector and the BMSE can be approximated according to

$$\mathbf{b}(\theta_{\mathcal{S}}) \approx \sigma_e^2 \mathbf{H}_{\mathcal{S}}^{\dagger} \mathbf{E}_{\tilde{\mathbf{H}}_{\mathcal{S}} \hat{\mathbf{H}}_{\mathcal{S}}} \ddot{\mathbf{H}}_{\mathcal{S}} \theta_{\mathcal{S}}, \quad (30)$$

$$\text{BMSE}_{\text{BiCE}} \approx \sigma^2 F_{\mathcal{S}} + \sigma_e^2 m_4^e \bar{F}_{\mathcal{S}} \quad (31)$$

where  $m_4^e = \mathbb{E}(e_i^4)$  and

$$F_{\mathcal{S}} = \text{Tr} \left[ \left( \mathbf{H}_{\mathcal{S}}^T \mathbf{P}_{\tilde{\mathbf{H}}_{\mathcal{S}}}^{\perp} \mathbf{H}_{\mathcal{S}} \right)^{-1} \right], \quad (32)$$

$$\bar{F}_{\mathcal{S}} = \text{Tr} \left[ \ddot{\mathbf{H}}_{\mathcal{S}}^T \mathbf{P}_{\tilde{\mathbf{H}}_{\mathcal{S}}}^{\perp} \mathbf{H}_{\mathcal{S}} \left( \mathbf{H}_{\mathcal{S}}^T \mathbf{P}_{\tilde{\mathbf{H}}_{\mathcal{S}}}^{\perp} \mathbf{H}_{\mathcal{S}} \right)^{-2} \mathbf{H}_{\mathcal{S}}^T \mathbf{P}_{\tilde{\mathbf{H}}_{\mathcal{S}}}^{\perp} \ddot{\mathbf{H}}_{\mathcal{S}} \right] \quad (33)$$

where  $\ddot{\mathbf{H}}$  is the second-order derivative of the dictionary matrix.

**Proof** See the appendix VII-A.  $\square$

Hereafter, the robustness of the BiCE is examined in the BM scenario, *i.e.*, where  $\sigma_e^2$  is small but dominates the noise variance  $\sigma^2$ .

*Result 4.2:* In the BM context, the BMSE of the BiCE is given by

$$\text{BMSE}_{\text{BiCE}} \approx \sigma_e^2 m_4^e \bar{F}_{\mathcal{S}} = O(m_4^e). \quad (34)$$

**Proof** The proof is straightforward.  $\square$

### C. Statistical efficiency of the BiCE

1) *Without BM* : Here, we show that the accuracy of a sparse-based estimation scheme is not degraded by the BiCE post-processing where there is no BM. We begin by the derivation of the biased Interfering-ECRB (I-ECRB) for the projected measurement vector  $\mathbf{y}$  onto the subspace  $\langle \hat{\mathbf{H}}_{\mathcal{S}} \rangle^{\perp}$ .

*Lemma 4.3:* The biased Interfering-ECRB (I-ECRB) [43] with respect to the projected measurements  $\dot{\mathbf{U}}_{\mathcal{S}}^T \mathbf{y}$  according to the orthogonal decomposition  $\mathbf{P}_{\hat{\mathbf{H}}_{\mathcal{S}}}^{\perp} = \dot{\mathbf{U}}_{\mathcal{S}} \dot{\mathbf{U}}_{\mathcal{S}}^T$  verifies the following inequality:

$$\mathbb{E} \left( \|\theta_{\mathcal{S}} - \hat{\theta}_{\mathcal{S}} \left( \dot{\mathbf{U}}_{\mathcal{S}}^T \mathbf{y} \right)\|^2 \right) \geq \mathcal{C}_{\mathcal{S} | \dot{\mathbf{U}}_{\mathcal{S}}^T \mathbf{y}} \quad (35)$$

where

$$\mathcal{C}_{\mathcal{S} | \dot{\mathbf{U}}_{\mathcal{S}}^T \mathbf{y}} \approx \sigma^2 F_{\mathcal{S}} + \sigma_e^2 \sigma_e^4 \bar{F}_{\mathcal{S}} \quad (36)$$

where quantities  $F_{\mathcal{S}}$  and  $\bar{F}_{\mathcal{S}}$  are given by expressions (32) and (33), respectively.

**Proof** See the appendix VII-B.  $\square$

Thanks to Lemma 4.3 and Result 4.1, we can formulate the following remark.

*Remark 4.4:* For the scenario where there is no BM ( $\sigma_e^2 \rightarrow 0$ ), we have:

- The BiCE is conditionally unbiased.
- The BMSE of the BiCE converges toward  $\mathcal{C}_{\mathcal{S} | \dot{\mathbf{U}}_{\mathcal{S}}^T \mathbf{y}}$  given in (36).

This remark is important because it means that the BiCE when there is no BM remains statistically efficient with respect to the projected measurement vector. This is a form of statistical efficiency.

But this property means that the BiCE efficiency may be sub-optimal. Indeed, the most favorable or ideal scenario is described by the unbiased ECRB, *i.e.* the ECRB for unprojected measurement vector and for an unbiased estimator. This lower bound is

$$\mathbb{E} \left( \|\theta_{\mathcal{S}} - \hat{\theta}_{\mathcal{S}}(\mathbf{y})\|^2 \right) \geq \mathcal{C}_{\mathcal{S} | \mathbf{y}} = \sigma^2 \text{Tr} \left[ \left( \mathbf{H}_{\mathcal{S}}^T \mathbf{H}_{\mathcal{S}} \right)^{-1} \right]. \quad (37)$$

It is easy to show that  $\mathcal{C}_{\mathcal{S} | \dot{\mathbf{U}}_{\mathcal{S}}^T \mathbf{y}} > \mathcal{C}_{\mathcal{S} | \mathbf{y}}$ , thus theoretically, the BiCE efficiency is sub-optimal. But in practice, as we will show in the simulation section, the biased I-ECRB is numerically very close to the ideal bound  $\mathcal{C}_{\mathcal{S} | \mathbf{y}}$ .

2) *With BM:*

a) *General expression:*

*Result 4.5:* In the case of BM, the I-ECRB,  $\mathcal{C}_{S|\dot{\mathbf{U}}_S^T \mathbf{y}}$ , is a lower bound of the BMSE of the BiCE.

**Proof** In the case of BM, the statistical efficiency of the BiCE is given by the following approximation:

$$\text{BMSE}_{\text{BiCE}} - \mathcal{C}_{S|\dot{\mathbf{U}}_S^T \mathbf{y}} \approx \sigma_\theta^2 (m_4^e - \sigma_e^4) \bar{F}_S. \quad (38)$$

Using the Jensen's inequality for a quadratic (convex) function, we have  $m_4^e \geq \sigma_e^4$  and thus  $\text{BMSE}_{\text{BiCE}} \geq \mathcal{C}_{S|\dot{\mathbf{U}}_S^T \mathbf{y}}$ .

b) *Illustration for a Generalized Gaussian distribution of the error:* To illustrate the BMSE of the BiCE and its statistical efficiency the error vector is assumed to follow a centered Generalized Gaussian distribution [44] such as  $e_\ell \sim \mathcal{GN}(0, \sigma_e^2 \frac{\Gamma(1/s)}{\Gamma(3/s)}, s)$ . The Generalized Gaussian distribution is a general distribution which encompasses the Laplacian ( $s = 1$ ), Gaussian ( $s = 2$ ) and uniform ( $s \rightarrow \infty$ ) distributions. The  $k$ -th moment of a centered Generalized Gaussian distribution is given by  $m_k^e = \sigma_e^k \sqrt{\frac{\Gamma(1/s)^k}{\Gamma(3/s)}} \frac{\Gamma((k+1)/s)}{\Gamma(1/s)}$ . Thus  $m_2^e = \sigma_e^2$  and  $m_4^e = \sigma_e^4 k_s$  where  $k_s = \frac{\Gamma(1/s)\Gamma(5/s)}{\Gamma(3/s)^2}$ . Let  $\sigma_e = d/6$  to ensure that for  $s \rightarrow \infty$  the support of the Uniform distribution matches with the grid spacing  $d$ . In this case, the BM variance is  $\sigma_e^2 = d^2/36$  and the 4-th order moment is given by  $m_4^e = k_s d^4/(36)^2$ . Note that  $k_s > 1$  which shows regarding equation (34) that the BiCE improves drastically the estimation robustness in the BM scenario when the grid spacing  $d < 1$  for a reasonably small additional computational cost. The statistical efficiency of the BiCE is

$$\text{BMSE}_{\text{BiCE}} - \mathcal{C}_{S|\dot{\mathbf{U}}_S^T \mathbf{y}} \approx \sigma_\theta^2 \sigma_e^4 \bar{F}_S (k_s - 1) \quad (39)$$

which is a decreasing function with respect to  $s$ . Thus, the statistical efficiency of the BiCE is in  $O(\sigma_e^4)$  and the higher  $s$  the better the efficiency of BiCE. Thus, the sharpest prior leads to the lowest statistical efficiency as illustrated in Fig. 8.

#### D. Evaluation of the dominating computational cost in Flops

In this section, the dominating computational cost of the BiCE is given and compared to the OMP computational cost. The OMP complexity is in  $O(LNK)$  for large  $K$  [45]. The OMP is known to have a relatively low computational cost with respect to other sparse-based estimators.

The BiCE involves as a first step, the computation of a  $(N - L)$ -rank orthogonal projector,  $\mathbf{P}_{\hat{\mathbf{H}}_S}^\perp$  which can be evaluated at the cost  $O((N - L)NL)$  which is the cost of a reduced-rank SVD. The second step is the projection  $\mathbf{Z} = \mathbf{P}_{\hat{\mathbf{H}}_S}^\perp \hat{\mathbf{H}}_S$  which can be evaluated in  $O(N^2L)$  and finally, the last step is the resolution of an ordinary LS criterion given by  $\mathbf{Z}^T \mathbf{Z} \hat{\boldsymbol{\theta}}_S = \mathbf{Z}^T \mathbf{y}$  which is evaluated in  $O(NL^2)$ . So, we can conclude that the dominating cost of the BiCE is  $O(N^2L)$ . As  $N \ll K$  by assumption, we can conclude that the BiCE dominating computational cost is much lower than the dominating cost of the OMP.

## V. APPLICATION TO THE COMPRESSIVE SAMPLING OF NON-BANDLIMITED SIGNALS

To illustrate our results, we consider the challenging problem of the sampling of non-bandlimited stream of pulses. A typical non-bandlimited continuous-time signal [46,47] which violates the Shannon's sampling theory [10] is given by

$$x(t) = \sum_{\ell \in \mathcal{S}} \theta_\ell \delta(t - \tau_\ell) \quad (40)$$

where  $\delta(t)$  is the Dirac delta symbol and signal  $x(t)$  is parametrized by  $L$  time-delays and amplitudes, denoted by  $\{\tau_\ell, \theta_\ell\}_{\ell \in \mathcal{S}}$ , respectively. Note that this model is widely used in signal-based applications and exhibits sparsity in the time domain.

Let  $g(t)$  be a sampling kernel<sup>1</sup> which verifies the Strang-Fix conditions [51]. Given the available vector  $\mathbf{s} = [s_1 \dots s_K]^T$  constituted by the collection of the sampled coefficients obtained by the noisy uniform sampling of  $x(t)$  according to

$$s_k = \int g(t - (k - 1)T_S)x(t)dt + n_k \quad (41)$$

in which  $T_S = 1$ ,  $n_k$  is a real discrete zero-mean white Gaussian noise process of variance  $\sigma^2$  and

$$\int g(t - (k - 1)T_S)x(t)dt = \sum_{\ell \in \mathcal{S}} \theta_\ell g(\tau_\ell - (k - 1)T_S), \quad (42)$$

the problem of interest can be described as the estimation of the  $2L$  unknown parameters of interest. We can find several efficient estimators in the literature [46,47,51].

Unlike the traditional approaches, the considered sampling problem is formulated, here, in the CS framework. To do so, the following basis matrix is defined as  $[\Phi]_{kk'} = g\left(\frac{\tau_{k'}}{T_S} - k\right)$  where we have considered a regular  $K$ -length partition of the time-delay range and a sampling kernel given by  $g(t) = \frac{1}{T_S} \text{sinc}(t)$  [46,47]. In this section, the time-delays are generated as  $L < N$  integer values extracted from a random permutation of the set  $\{1, \dots, K\}$ . As a consequence, matrix  $\Phi$  can be approximated as the  $K$ -dimensional "delta spikes" basis and can be considered as orthonormal. The entries of the measurement matrix  $\Psi$  are generated as an i.i.d. realization of the distribution  $\mathcal{N}(0, 1/N)$ . Consequently, as the measurement matrix follows the requirements of section II.A, it follows that the dictionary matrix  $\mathbf{H} = \Psi\Phi$  is incoherent. This means that the RIP holds with high probability if the compressive measurement vector given by  $\mathbf{y} = \Psi\mathbf{s}$  is a  $N$ -length vector with  $N = O(L \log(K/L))$  [1,34,52] which is smaller than  $K$ . So, the compressive sampling model is formally similar to model (4) and thus can be formulated in the CS framework.

#### A. Saturated BMSE and BCRB

On Fig. 4, we have drawn the BMSE of the OMP with and without BM and the corresponding BCRB as a function of the SNR =  $\mathbb{E}(\|\boldsymbol{\theta}_S\|^2)/\mathbb{E}(\|\mathbf{n}\|^2)$  in dB. As an initial observation, we can note that the BM effect is highly problematic in the

<sup>1</sup>Note that the best choice of this kernel is under some aspects always an open problem [48]–[50].



regime where the SNR is sufficiently high, *i.e.*, when the noise variance is dominated by  $\sigma_e^2$ . In our simulation set-up, this SNR limit is around 10 dB. In the case of no BM, we verify as expected that the OMP BMSE reaches the BCRB without BM for a sufficiently large SNR. On the other hand, the BMSE of the OMP with BM suffers from the well-known BMSE saturation effect even in the high SNR regime and for a perfect estimation of the support  $\mathcal{S}$  [16]–[20]. In addition, this figure shows that the proposed lower bounds under BM, namely  $\text{BCRB}_{\text{BM}}$  and  $\text{BCRB}_{\text{BM}}^{\text{sat}}$ , accurately predict this behavior.

On Fig. 5, the saturated value of the bound,  $\text{BCRB}_{\text{BM}}^{\text{sat}}$ , given in (20) is drawn as a function of the ratio  $N/L$  and for different  $\sigma_e^2$ . Even if the doubly asymptotic framework of the RMT is not fulfilled here, we can note the good agreement of the approximation  $\text{BCRB}_{\text{BM}}^{\text{sat}}$  with the exact BCRB for different values of  $\sigma_e^2$  and for a wide range of ratio  $N/L$ , but not too close to the limit case of  $N/L \rightarrow 1$ . So,  $\text{BCRB}_{\text{BM}}^{\text{sat}}$  is an accurate prediction of the saturated BMSE effect due to the BM.

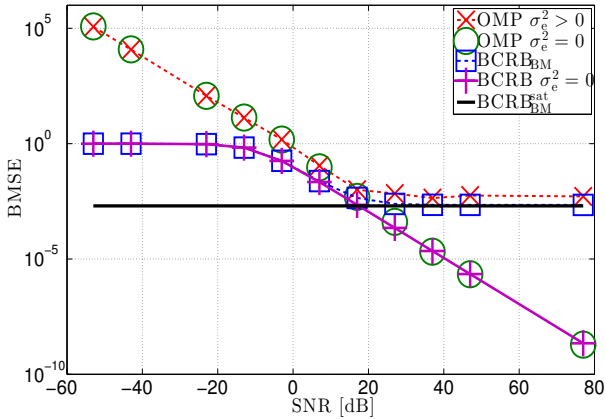


Fig. 4. BMSEs Vs. SNR in dB. We consider the OMP estimator with and without BM and the corresponding BCRB. The approximation  $\text{BCRB}_{\text{BM}}^{\text{sat}}$  is given in (20). We fix  $L = 50$  i.i.d. Gaussian amplitudes,  $N = 200$ ,  $K = 300$  and  $\sigma_e^2 = 10^{-4}$ .

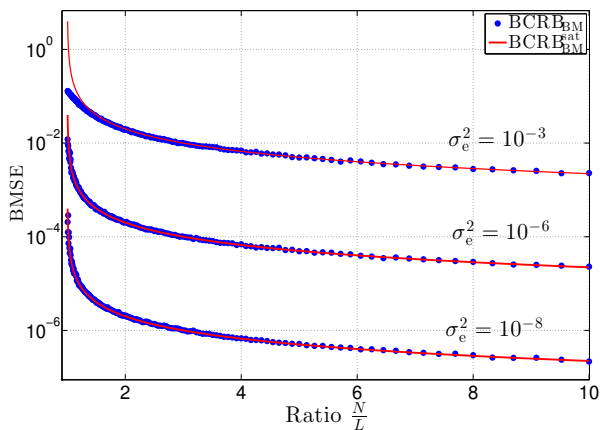


Fig. 5.  $\text{BCRB}_{\text{BM}}$  and  $\text{BCRB}_{\text{BM}}^{\text{sat}}$ , Vs. ratio  $N/L$  with  $\sigma_e^2 = 10^{-10}$ ,  $K = 300$ ,  $20 \leq L \leq N - 1$  where  $N = 200$  and for different  $\sigma_e^2$ .

## B. Numerical accuracy of the BiCE

Before showing the BMSE measurements of the BiCE, we illustrate on Fig. 6, the theoretical BMSE of the proposed estimator given in (31), the approximated biased I-ECRB and the ECRB are computed. At low SNR when the noise dominates the BM effect, the theoretical approximated BMSE, and the two lower bounds are merged, as expected since it is well-known that the BM effect appears at high SNR. At contrary, for a sufficiently high SNR (around 20 dB on Fig. 6), the BM becomes the main contribution and dominates for the noise. In this case, we can see that the theoretical approximated BMSE given in (31) and the biased I-ECRB well predict the saturated effect due to the BM. This last observation has to be understood relatively to the ECRB corresponding to the most favorable scenario since this bound has been derived for an unbiased estimator and without BM. Fig. 7 considers the same set-up as in Fig. 6 but is drawn with respect to  $\sigma_e^2$ . We first note that even if  $\sigma_e^2$  is small the biased I-ECRB cannot meet the ideal bound given by the ECRB. But, the two bounds remains relatively close for  $\sigma_e^2 < 10^{-3}$ . A more important observation is that for small  $\sigma_e^2$  the theoretical BMSE of the BiCE coincides with the biased I-ECRB. This observation illustrates the intimate dependency between of the BiCE and the presented biased I-ECRB. For higher  $\sigma_e^2$ , the BMSE of the BiCE and the I-ECRB are not fully in line. This can be explained by the fact that for high  $\sigma_e^2$ , the first-order Taylor expansion assumed in the BiCE becomes a rough approximation. But, the biased I-ECRB remains a lower bound that well predicts the behavior of the BiCE estimation accuracy. To complete this analysis, expression (35) is computed on Fig. 8 with respect to the different distributions (parametrized by  $s$ ) of the BM error of several variances. This relation gives a closed-form expression of the difference between the BMSE of the BiCE and the biased I-ECRB. We can see that the choice of parameter  $s$  has a low impact on the difference between the BMSE of the BiCE and the biased I-ECRB. The distributions highly concentrated around zero as for instance the Laplacian distribution are less favorable than the "flat" distribution.

Toward the estimation of the support (*i.e.* the time-delays) and of the amplitudes for the considered application, we propose a complete estimation scheme based on the architecture described on Fig. 3. To do so, we consider three popular sparse-based estimators: the OMP [11], the Cosamp [12] and the BP [13] based on the SPGL1 MatLab ToolBox [53]. Unlike, the OMP and the Cosamp, the BP does not know the cardinality of the support, so the  $L$  highest amplitudes are selected in the output estimated vector. Note that the BP assumes an *a priori* knowledge of the noise variance and of the amplitude prior. This is not the case for the OMP and the Cosamp. On Fig. 9, we plot the BMSE measurements for the considered sparse-based estimators and their improved versions denoted by BiCE-OMP, BiCE-Cosamp and BiCE-BP. As expected, we first note that even for a small BM, the BMSEs of the OMP, the Cosamp and the BP are saturated. At contrary, with a small additional computational cost, our proposed schemes, namely the BiCE-OMP, the BiCE-Cosamp

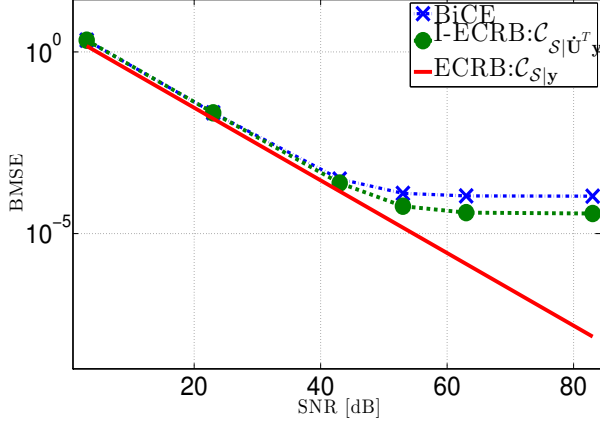


Fig. 6. BMSE and ECRBs Vs. SNR in dB. The approximated BMSEs of the BiCE are given in expression (45). In addition, the (biased) I-ECRB and the (unbiased) ECRB are in expressions (36) and (37), respectively. We fix  $N = 50$ ,  $K = 200$ ,  $L = 10$ ,  $\sigma_e^2 = 10^{-3}$  and  $s = 2$  (Gaussian BM error).

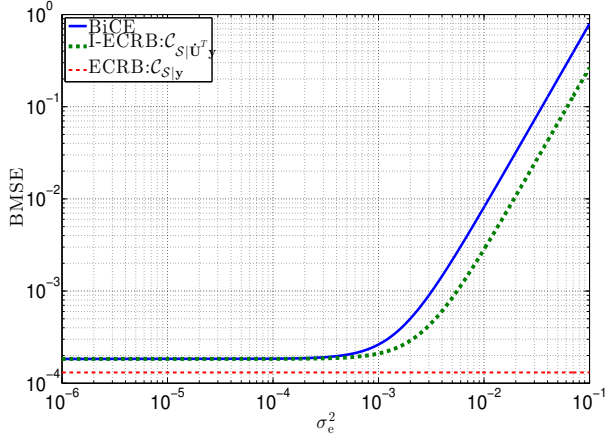


Fig. 7. BMSEs Vs.  $\sigma_e^2$ . The approximated BMSEs of the BiCE are given in expression (45). In addition, the (biased) I-ECRB and the (unbiased) ECRB are in expressions (36) and (37), respectively. We fix  $N = 50$ ,  $K = 200$ ,  $L = 10$ ,  $\sigma^2 = 10^{-5}$  and  $s = 2$  (Gaussian BM error).

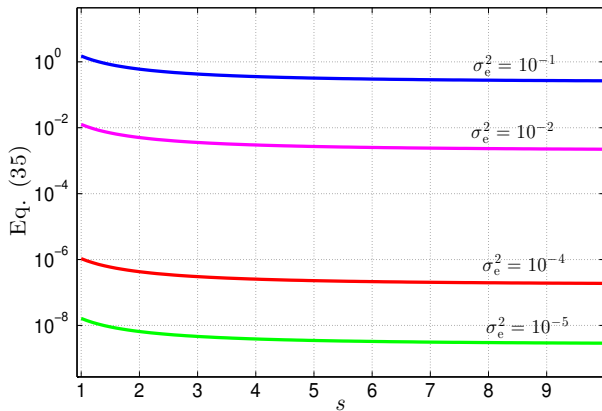


Fig. 8. Difference between the BMSE of the BiCE and the biased I-ECRB given by (35) Vs. the choice of the distribution of the BM error (parametrized by  $s$ ) and for several values of  $\sigma_e^2$ .

and the BiCE-BP exhibit considerably lower BMSEs. So, the association of a standard sparse-based estimator and the BiCE allows to improve efficiently the robustness to BM.

In addition, the BMSEs of the proposed schemes in presence of BM are very close for a wide range of SNR to the ECRB, given in (37), corresponding to the most favorable scenario. As a conclusion, our proposed scheme is able to drastically robustify against BM any sparse-based estimators with a low additional computational cost and without the knowledge of the true dictionary.

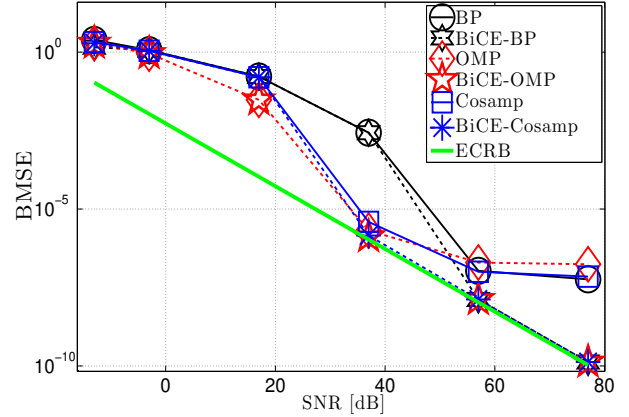


Fig. 9. BMSEs Vs. SNR in dB. Standard sparse-based estimators and their improved versions thanks to the BiCE. We fix  $N = 200$ ,  $K = 300$ ,  $L = 10$ ,  $\sigma_e^2 = 10^{-6}$ ,  $\sigma_\theta^2 = 10^{-1}$ .

## VI. CONCLUSIONS

Compressed Sensing (CS) has a wide range of challenging signal-based applications. The principle of CS or sparse-based estimation techniques is to look for a decomposition (*i.e.*, an amplitude vector) of the available measurement vector into few basis vectors. But, in operational context, the basis matrix may not to be perfectly known. In this work, we assume that the basis matrix suffers from a mismatch error and only the Basis Mismatch (BM) matrix is available to us. Unfortunately, the plethora of existing sparse-based estimators are not robust to the BM degradation. Specifically, even for a small mismatch error which dominates the noise variance and for a perfect estimation of the support, *i.e.*, indexes of the non-zero amplitudes, the estimation accuracy in terms of Bayesian Mean Square Error (BMSE), of any sparse-based estimators is saturated. This means that these estimators are not statistically efficient and this is a serious drawback in the concrete exploitation of CS or sparse-based methods. In this context, we first derive the BCRB under BM and we provide an approximated lower bound for the saturated BMSE. In the second part, an estimator, called BiCE for Bias-Correction Estimator, is proposed for considerably mitigate the BM degradation. The BiCE acts as a post-processing layer with respect to any sparse-based estimator and improves the amplitude vector estimation. The theoretical statistical analysis of the bias and the BMSE of this estimator is derived. To illustrate the interest of our approach, the BiCE is applied in the important

context of the non-bandlimited signal compressive sampling. Finally, we can say that the proposed estimator has several advantages since the BiCE (i) is a blind algorithm, *i.e.*, is unaware of the uncorrupted dictionary matrix, (ii) is generic since it can be associated to any sparse-based estimator, (iii) is fast, *i.e.*, the additional computational cost remains low and (iv) has good statistical properties.

## VII. APPENDIX

### A. Proof of Result 4.1

The measurement vector admits the following expressions:

$$\mathbf{y} = \mathbf{H}_S \boldsymbol{\theta}_S + \mathbf{n} \approx \tilde{\mathbf{H}}_S \boldsymbol{\theta}_S - \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S) \boldsymbol{\theta}_S + \mathbf{n}. \quad (43)$$

First remark that the structure of the BM given in expression (6), leads to the following key relation:

$$\dot{\mathbf{H}}_S = \dot{\mathbf{H}}_S + \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S). \quad (44)$$

Using the above relation with (43) leads to the following expression:

$$\mathbf{y} \approx \left( \tilde{\mathbf{H}}_S - \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S) \right) \boldsymbol{\theta}_S + \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S)^2 \boldsymbol{\theta}_S + \mathbf{n}$$

where  $\mathbf{e}_S$  is vector  $\mathbf{e}$  restricted to the errors corresponding to the support  $\mathcal{S}$ . Now apply the BiCE in the case where the support set has been estimated without error, *i.e.*  $\hat{\mathcal{S}} = \mathcal{S}$ , we obtain

$$\begin{aligned} \hat{\boldsymbol{\theta}}_S &\approx \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S} \left[ \left( \tilde{\mathbf{H}}_S - \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S) \right) \boldsymbol{\theta}_S \right. \\ &\quad \left. + \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S)^2 \boldsymbol{\theta}_S + \mathbf{n} \right] \\ &\approx \boldsymbol{\theta}_S + \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S} \left( \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S)^2 \boldsymbol{\theta}_S + \mathbf{n} \right) \end{aligned}$$

and the estimation error is

$$\hat{\boldsymbol{\theta}}_S - \boldsymbol{\theta}_S \approx \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S} \left( \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S)^2 \boldsymbol{\theta}_S + \mathbf{n} \right).$$

The BMSE depending on the corrupted dictionary matrix is given by

$$\text{BMSE}_{\text{BiCE}} \approx (m_4^o \sigma_\theta^2 + \sigma^2) \text{Tr} \left[ \left( \tilde{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \tilde{\mathbf{H}}_S \right)^{-1} \right]. \quad (45)$$

However, this expression is not a closed-form expression since the (45) depends on the random corrupted matrices  $\tilde{\mathbf{H}}_S$  and  $\dot{\mathbf{H}}_S$ . Therefore, we aim to approximate this expression to find a closed-form expression depending only on the uncorrupted dictionary matrices  $\mathbf{H}_S$  and  $\dot{\mathbf{H}}_S$ . To do this, we exploit the first-order Taylor approximation of the orthogonal projector [54] given by  $\mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \approx \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \boldsymbol{\Lambda}$  where

$$\boldsymbol{\Lambda} = \dot{\mathbf{H}}_S^{\dagger T} \text{diag}(\mathbf{e}_S) \ddot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp + \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S) \dot{\mathbf{H}}_S^\dagger.$$

Based on the above approximation and following the methodology introduced in [43], we can derive after some calculus the following first-order Taylor approximations:

$$\begin{aligned} \tilde{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \tilde{\mathbf{H}}_S &= \left( \mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S) \right)^T \cdot \\ \left( \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \dot{\mathbf{H}}_S^{\dagger T} \text{diag}(\mathbf{e}_S) \ddot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S) \dot{\mathbf{H}}_S^\dagger \right) \\ \cdot \left( \mathbf{H}_S + \dot{\mathbf{H}}_S \text{diag}(\mathbf{e}_S) \right) &\approx \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \mathbf{H}_S \end{aligned}$$

and

$$\begin{aligned} \tilde{\mathbf{H}}_S^\dagger \mathbf{E}_{\tilde{\mathbf{H}}_S \dot{\mathbf{H}}_S} &\approx \left( \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \tilde{\mathbf{H}}_S^T \mathbf{P}_{\tilde{\mathbf{H}}_S}^\perp \\ &\approx \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \dot{\mathbf{H}}_S} + \text{diag}(\mathbf{e}_S) \left( \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \\ &\quad \times \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^\dagger \right). \end{aligned}$$

Based on the above expressions, we get that the approximation of the estimation error vector:

$$\begin{aligned} \hat{\boldsymbol{\theta}}_S - \boldsymbol{\theta}_S &\approx \left[ \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \dot{\mathbf{H}}_S} + \text{diag}(\mathbf{e}_S) \left( \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \right. \\ &\quad \left. \times \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^\dagger \right) \right] \mathbf{n} \\ &\quad + \text{diag}(\mathbf{e}_S)^2 \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \dot{\mathbf{H}}_S} \ddot{\mathbf{H}}_S \boldsymbol{\theta}_S. \end{aligned}$$

Using that the noise is centered,  $\mathbb{E}(\text{diag}(\mathbf{e}_S)^2) = \sigma_e^2 \mathbf{I}$ , and tacking the expectation conditionally to  $\boldsymbol{\theta}_S$  of the estimation error vector, the bias vector is given by (30).

For the purpose of brevity and as the derivations are tedious but straightforward, only a coarse sketch of the derivation of the BMSE is provided here (see the derivations in (46)). By discarding the terms in  $O(\sigma_e^2 \sigma^2)$  in expression (46) gives, after calculus, expression (31).

### B. Proof of Lemma 4.3

Let  $\hat{\boldsymbol{\theta}}_S \left( \dot{\mathbf{U}}_S^T \mathbf{y} \right)$  be a Bayesian estimator based on the projected observation  $\dot{\mathbf{U}}_S^T \mathbf{y}$ . Note that  $\dot{\mathbf{U}}_S^T \mathbf{y} | \boldsymbol{\theta}_S, \mathbf{H}_S \sim \mathcal{N} \left( \dot{\mathbf{U}}_S^T \mathbf{H}_S \boldsymbol{\theta}_S, \sigma^2 \mathbf{I} \right)$ . The corresponding BMSE verifies the following inequality:

$$\mathbb{E} \left( \left\| \boldsymbol{\theta}_S - \hat{\boldsymbol{\theta}}_S \left( \dot{\mathbf{U}}_S^T \mathbf{y} \right) \right\|^2 \right) \geq \mathcal{C}_{S | \dot{\mathbf{U}}_S^T \mathbf{y}}$$

where  $\mathcal{C}_{S | \dot{\mathbf{U}}_S^T \mathbf{y}}$  stands for the biased Interfering-ECRB (I-ECRB)<sup>2</sup> which is defined by

$$\begin{aligned} \mathcal{C}_{S | \dot{\mathbf{U}}_S^T \mathbf{y}} &= \text{Tr} \left[ \left( \mathbf{I} + \frac{\partial \mathbf{b}(\boldsymbol{\theta}_S)}{\partial \boldsymbol{\theta}_S} \right) \mathbf{F}(\boldsymbol{\theta}_S | \dot{\mathbf{U}}_S^T \mathbf{y})^{-1} \right. \\ &\quad \left. \times \left( \mathbf{I} + \frac{\partial \mathbf{b}(\boldsymbol{\theta}_S)}{\partial \boldsymbol{\theta}_S} \right)^T \right] + \mathbb{E}_{\boldsymbol{\theta}_S} \left( \|\mathbf{b}(\boldsymbol{\theta}_S)\|^2 \right) \end{aligned}$$

where we have remarking that there is no need to the expectation over the amplitude vector since  $\mathbf{F}(\boldsymbol{\theta}_S | \dot{\mathbf{U}}_S^T \mathbf{y})^{-1} = \sigma^2 \left( \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1}$  and  $\frac{\partial \mathbf{b}(\boldsymbol{\theta}_S)}{\partial \boldsymbol{\theta}_S} \approx \sigma_e^2 \mathbf{H}_S^\dagger \mathbf{E}_{\mathbf{H}_S \dot{\mathbf{H}}_S} \ddot{\mathbf{H}}_S$  are not functions of the amplitude vector. The expected norm of the bias in the above biased I-ECRB is given by

$$\mathbb{E}_{\boldsymbol{\theta}_S} \left( \|\mathbf{b}(\boldsymbol{\theta}_S)\|^2 \right) \approx \sigma_e^4 \sigma_\theta^2 \bar{F}_S.$$

Discarding the terms in  $O(\sigma_e^2 \sigma^2)$  or in  $O(\sigma_e^4 \sigma^2)$  with respect to the term in  $O(\sigma^2)$  for sufficiently small  $\sigma_e^2 \sigma^2$ , we obtain after calculus,

$$\begin{aligned} \text{Tr} \left[ \left( \mathbf{I} + \frac{\partial \mathbf{b}(\boldsymbol{\theta}_S)}{\partial \boldsymbol{\theta}_S} \right) \mathbf{F}(\boldsymbol{\theta}_S | \dot{\mathbf{U}}_S^T \mathbf{y})^{-1} \right. \\ \left. \times \left( \mathbf{I} + \frac{\partial \mathbf{b}(\boldsymbol{\theta}_S)}{\partial \boldsymbol{\theta}_S} \right)^T \right] \approx \sigma^2 F_S \end{aligned}$$

where  $F_S$  has been given in (32). Consequently, the biased I-ECRB takes expression (36).

<sup>2</sup>The biased I-ECRB is the expected bound over vector  $\boldsymbol{\theta}_S$  of the biased I-CRB given in [43] and defined with respect to the projected measurement vector  $\dot{\mathbf{U}}_S^T \mathbf{y}$ .

$$\begin{aligned}
\text{BMSE}_{\text{BiCE}} &\approx \sigma^2 \text{Tr} \left[ \left( \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \dot{\mathbf{H}}_S} \right)^2 \right] + \sigma_e^2 \sigma^2 \text{Tr} \left[ \left( \left( \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^\dagger \right) \right)^2 \right] \\
&\quad + m_4^e \sigma_\theta^2 \text{Tr} \left[ \left( \mathbf{H}^\dagger \mathbf{E}_{\mathbf{H}_S \ddot{\mathbf{H}}_S} \right)^2 \right] \\
&\approx \sigma^2 \text{Tr} \left[ \left( \left( \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \right)^2 \right] + \sigma_e^2 \sigma^2 \text{Tr} \left[ \left( \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-2} \left( \dot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \ddot{\mathbf{H}}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp - \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \dot{\mathbf{H}}_S^\dagger \right)^2 \right] \\
&\quad + m_4^e \sigma_\theta^2 \text{Tr} \left[ \left( \left( \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \mathbf{H}_S \right)^{-1} \mathbf{H}_S^T \mathbf{P}_{\dot{\mathbf{H}}_S}^\perp \ddot{\mathbf{H}}_S \right)^2 \right] \tag{46}
\end{aligned}$$

## REFERENCES

- [1] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] R. Baraniuk, "Compressive sensing," *IEEE signal processing magazine*, vol. 24, no. 4, 2007.
- [3] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [4] N. Wagner, Y. C. Eldar, and Z. Friedman, "Compressed beamforming in ultrasound imaging," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4643–4657, 2012.
- [5] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [6] J. L. Paredes, G. R. Arce, and Z. Wang, "Ultra-wideband compressed sensing: channel estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 383–395, 2007.
- [7] Y. Wang, G. Leus, and A. Pandharipande, "Direction estimation using compressive sampling array processing," in *IEEE Workshop on Statistical Signal Processing (SSP'09)*. IEEE, 2009, pp. 626–629.
- [8] D. Malioutov, M. Cetin, and A. 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.
- [9] M. Herman and T. Strohmer, "High-resolution radar via compressed sensing," *IEEE Transactions on Signal Processing*, vol. 57, no. 6, pp. 2275–2284, June 2009.
- [10] M. Unser, "Sampling-50 years after Shannon," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.
- [11] Y. C. Pati, R. Rezaifar, Y. C. Pati, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. of the 27th Annual Asilomar Conference on Signals, Systems, and Computers*, 1993, pp. 40–44.
- [12] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [13] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [14] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [15] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [16] M. Rosenbaum and A. B. Tsybakov, "Sparse recovery under matrix uncertainty," *The Annals of Statistics*, vol. 38, no. 5, pp. 2620–2651, 2010.
- [17] D. Chae, P. Sadeghi, and R. Kennedy, "Effects of basis-mismatch in compressive sampling of continuous sinusoidal signals," in *2nd International Conference on Future Computer and Communication (ICFCC)*, vol. 2, 2010, pp. V2–739–V2–743.
- [18] M. Herman and T. Strohmer, "General Deviants: An analysis of perturbations in compressed sensing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 342–349, 2010.
- [19] S. Bernhardt, R. Boyer, B. Zhang, S. Marcos, and P. Larzabal, "Performance analysis for sparse based biased estimator: Application to line spectra analysis," in *IEEE Sensor Array and Multichannel Signal Processing (SAM)*. invited paper, 2014, pp. 365–368.
- [20] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to Basis Mismatch in Compressed Sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.
- [21] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1127–1141, 1999.
- [22] F. A. Petitcolas, R. Anderson, and M. Kuhn, "Information hiding-a survey," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1062–1078, Jul 1999.
- [23] K. Hayashi, M. Nagahara, and T. Tanaka, "A user's guide to compressed sensing for communications systems," *IEICE transactions on communications*, vol. 96, no. 3, pp. 685–712, 2013.
- [24] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proceedings of the 5th international conference on Information processing in sensor networks*. ACM, 2006, pp. 134–142.
- [25] C. Delpha, S. Hijazi, and R. Boyer, "A compressive sensing based quantized watermarking scheme with statistical transparency constraint," in *Digital-Forensics and Watermarking*, ser. Lecture Notes in Computer Science, Y. Q. Shi, H.-J. Kim, and F. Prez-Gonzlez, Eds. Springer Berlin Heidelberg, 2014, pp. 409–422.
- [26] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2002–2016, 2011.
- [27] Z. Yang, C. Zhang, and L. Xie, "Robustly stable signal recovery in compressed sensing with structured matrix perturbation," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4658–4671, 2012.
- [28] Z. Tan, P. Yang, and A. Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4997–5008, Oct 2014.
- [29] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse bayesian inference," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 38–43, 2013.
- [30] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramer-Rao bound in noisy compressed sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 518–526, 2012.
- [31] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér-Rao bound for noisy compressive sampling," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1233–1236, 2009.
- [32] Z. Ben-Haim and Y. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3384–3389, June 2010.
- [33] R. Jagannath and K. Hari, "Block sparse estimator for grid matching in single snapshot DoA estimation," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1038–1041, 2013.
- [34] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [35] M. A. Davenport, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "A simple proof that random matrices are democratic," *arXiv preprint arXiv:0911.0736*, 2009.
- [36] S. Bernhardt, R. Boyer, S. Marcos, and P. Larzabal, "Compressed Sensing with uncertainty - The Bayesian estimation perspective," in *IEEE International Workshop on Computational Advances (CAMSA'15)*, 2015.
- [37] H. L. Van Trees and K. L. Bell, "Bayesian bounds for parameter estimation and nonlinear filtering/tracking," *AMC*, vol. 10, p. 12, 2007.
- [38] E. L. Lehmann and G. Casella, *Theory of point estimation*. Springer Science & Business Media, 1998, vol. 31.

- [39] A. Wiesel, Y. Eldar, and A. Yeredor, "Linear regression with gaussian model uncertainty: Algorithms and bounds," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2194–2205, June 2008.
- [40] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge University Press, 2011.
- [41] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004.
- [42] R. T. Behrens and L. L. Scharf, "Signal processing applications of oblique projection operators," *IEEE Transactions on Signal Processing*, vol. 42, no. 6, pp. 1413–1424, 1994.
- [43] R. Boyer, "Oblique projection for source estimation in a competitive environment: algorithm and statistical analysis," *Signal Processing*, vol. 89, no. 12, pp. 2547–2554, 2009.
- [44] S. Nadarajah, "A generalized normal distribution," *Journal of Applied Statistics*, vol. 32, no. 7, pp. 685–694, 2005.
- [45] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6202–6216, 2012.
- [46] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.
- [47] R. Tur, Y. C. Eldar, and Z. Friedman, "Innovation rate sampling of pulse streams with application to ultrasound imaging," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1827–1842, 2011.
- [48] J. A. Uriguen, T. Blu, and P. L. Dragotti, "FRI sampling with arbitrary kernels," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5310–5323, 2013.
- [49] S. Bernhardt, R. Boyer, S. Marcos, Y. C. Eldar, and P. Larzabal, "Cramér-Rao bound for finite streams of pulses," in *European Signal Processing Conference (EUSIPCO'14)*, 2014, pp. 984–988.
- [50] S. Bernhardt, R. Boyer, S. Marcos, Y. Eldar, and P. Larzabal, "Sampling FRI signals with the SoS kernel: Bounds and optimal kernel," in *European Signal Processing Conference (EUSIPCO'15)*, 2015.
- [51] P. L. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang–Fix," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1741–1757, 2007.
- [52] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [53] E. van den Berg and M. P. Friedlander, "SPGL1: A solver for large-scale sparse reconstruction," June 2007, <http://www.cs.ubc.ca/labs/scl/spgl1>.
- [54] H. Krim, P. Forster, and J. G. Proakis, "Operator approach to performance analysis of root-MUSIC and root-min-norm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, pp. 1687–1696, 1992.