

Large-System Estimation Performance in Noisy Compressed Sensing with Random Support - a Bayesian Analysis

Rémy Boyer, Romain Couillet, Bernard-Henry Fleury, Pascal Larzabal

► **To cite this version:**

Rémy Boyer, Romain Couillet, Bernard-Henry Fleury, Pascal Larzabal. Large-System Estimation Performance in Noisy Compressed Sensing with Random Support - a Bayesian Analysis. IEEE Transactions on Signal Processing, Institute of Electrical and Electronics Engineers, 2016, 64 (21), pp.5525-5535. 10.1109/TSP.2016.2591511 . hal-01346039

HAL Id: hal-01346039

<https://hal-centralesupelec.archives-ouvertes.fr/hal-01346039>

Submitted on 18 Jul 2016

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.

Oracle Large-System Estimation Performance in Noisy Compressed Sensing with Random Support - a Bayesian Analysis

Rémy Boyer, Romain Couillet, Bernard-Henry Fleury and Pascal Larzabal

Abstract—Compressed sensing (CS) enables measurement reconstruction by using sampling rates below the Nyquist rate, as long as the amplitude vector of interest is sparse. In this work, we first derive and analyze the Bayesian Cramér-Rao Bound (BCRB) for the amplitude vector when the set of indices (the support) of its non-zero entries is known. We consider the following context: (i) The dictionary is non-stochastic but randomly generated; (ii) the number of measurements and the support cardinality grow to infinity in a controlled manner, *i.e.* the ratio of these quantities converges to a constant; (iii) the support is random; and (iv) the vector of non-zero amplitudes follow a multidimensional generalized normal distribution. Using results from random matrix theory, we obtain closed-form approximations of the BCRB. These approximations can be formulated in a very compact form in low and high SNR regimes. Secondly, we provide a statistical analysis of the variance and the statistical efficiency of the oracle linear mean-square-error (LMMSE) estimator. Finally, we present results from numerical investigations in the context of non-bandlimited finite-rate-of-innovation (FRI) signal sampling. We show that the performance of Bayesian mean square error (BMSE) estimators that are aware of the cardinality of the support, such as OMP and CoSaMP, are in good agreement with the developed lower bounds in the high SNR regime. Conversely, sparse estimators exploiting only the knowledge of the parameter vector and the noise variance in form of a-priori distributions of these parameters, like LASSO and BPDN, are not efficient at high SNR. However, at low SNR their BMSE is lower than that of the former estimators and may be close to the BCRB.

I. INTRODUCTION

Modern data acquisition systems involve massive array technology and a deluge of data provided by large and highly interconnected networks, multi-databases, multidimensional data streaming, etc. Compressive sensing [1]–[3] is a promising solution to tackle the related new challenges, in that it allows for retrieving sparse signals with fewer samples than classical data acquisition theory requires [4]. CS has been successfully exploited in many realistic applications, such as channel estimation [5,6], equalization [7], sampling in magnetic resonance imaging [8], high-resolution radar imaging [9], and array processing [10]. Along with the CS theory,

Rémy Boyer and Pascal Larzabal are with Université Paris-Sud XI in Laboratoire des Signaux et Systèmes (L2S) and SATIE (ENS-Cachan), respectively. Romain Couillet is with SUPELEC in Laboratoire des Signaux et Systèmes (L2S). Bernard H. Fleury is with Aalborg University. Department of Electronic Systems.

This work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMMunications NEWCOM# (Grant agreement no. 318306). Rémy Boyer and Pascal Larzabal are supported by the following projects: ANR MAGELLAN, MI-CNRS DISPLAY and TITAN.

a large collection of sparse recovery estimators has emerged [1,2]. These include matching pursuit (MP) [11], orthogonal matching pursuit (OMP) [12]–[14], and basis pursuit (BP) [15] devised for pilot-assisted channel estimation as an effective alternative to the traditional least squared estimator (LSE). Convex optimization solvers such as SpaRSA [16], SPGL1 [17], YALL1 and GPSR [18] or again [19] are also efficient numerical implementations alternatives to solve the CS problem.

These estimators however have disparate performance. In particular, each one yields a different Bayesian Mean Squared Error (BMSE) as being based on specific model prior assumptions (e.g., knowledge of the number of non-zero amplitudes, of the prior distribution of these amplitudes, of the noise variance). To study the CS estimation problem in its generality, lower bounds on the BMSE provide the minimal estimation performance that a statistically efficient sparse estimator can achieve. The Bayesian lower bounds [20] are particularly well adapted as a benchmark for performance evaluation of Bayesian estimators, such as the Minimum Mean Squared Error (MMSE) and Maximum A-Posteriori probability (MAP) estimators, as well as for global performance evaluation of non-Bayesian estimators, such as the maximum-likelihood (ML) estimator. Bayesian bounds can be partitioned into two categories: the Ziv-Zakai family [21], derived from a binary hypothesis testing problem and the Weiss-Weinstein family [22], derived from the covariance inequality. The Ziv-Zakai class contains the Ziv-Zakai, Bellini-Tartara [23], Chazan-Zakai-Ziv, Weinstein, extended-Ziv-Zakai and Bell bounds [24]. The Weiss-Weinstein class contains the Bayesian Cramér-Rao, Bayesian Bhattacharyya [25], Bobrovsky-Zakai, Reuven-Messer, Weiss-Weinstein, Bayesian Abel bounds and the combined Cramér-Rao/Weiss-Weinstein bounds.

In this contribution we focus our work on the Van Trees' Bayesian Cramér-Rao Bound (BCRB) [20] which we shall show to be well tailored to the CS problem and relatively easy to derive. Alternatively, authors have derived the deterministic Cramér-Rao Bound (CRB) in [26]–[32] adapted to deterministic sparse amplitude vector estimation, the Fisher Information Matrix (FIM) (the matrix inverse of the CRB) in [33] for a source vector parametrized by the variance of each sources. The FIM derived in [33] belongs to the family of the stochastic CRBs [34], whereas the bound considered in the present article is obtained within a Bayesian context. In [35], the authors propose a Bayesian lower bound for sparse *hierarchical* model. The sparsity is controlled by the selection

of the probability distribution of a set of hyper-parameters. In the present work, no assumption on such a hierarchical models is made. Our probabilistic model induces less parameters to be estimated than a hierarchical model [36].

In addition, it is assumed in this work that the parameters of interest are the non-zero amplitudes after the shrinkage operation of a compressible¹ signal. This is not the case in [33,35] where the authors derive a BCRB to estimate more parameters of interest than the amount of measurements by using the Bayesian regularization philosophy. This approach leads to trivial bounds as demonstrated in the sequel. In addition, reference [38] discusses some identifiability issues for sparse signals. In [10], a BCRB is derived for the CS model but in the particular scenario of the off-grid problem.

To simplify the expressions of the BCRB, we shall consider in this article the regime where the number of measurements M and the number K of non-zero amplitudes are both large. More specifically, assuming that the amplitudes belonging to the support set admit a sparse representation into a large $M \times K$ dictionary matrix composed of random and independent entries, we shall rely on the notion of “deterministic equivalents” from random matrix theory [39]–[42] to obtain closed-form approximations of the BCRB. These closed-form expressions have the advantages of providing useful information on the behavior of the CS estimation problem irrespective of the dictionary and to decrease the computational complexity of the proposed analytical bounds, which are in general only mathematically tractable for Gaussian priors [43,44].

The remainder of the article is organized as follows. Section II presents the CS model. Section III introduces the derivation of the asymptotic BCRB, in particular in the context of extreme signal-to-noise ratio and when the amplitudes are independent and identically distributed according to a generalized normal distribution. Section IV presents the same analysis for the oracle-LMMSE estimator and its statistical efficiency is discussed. Finally, in Section V, we apply our results in the context of the FRI signal sampling.

II. THE COMPRESSED SENSING FRAMEWORK

Let \mathbf{y} be the $M \times 1$ noisy measurement vector in a (standard) compressed sensing (CS) model [1,2]:

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

where \mathbf{n} is centered circular white Gaussian noise of unknown variance σ^2 and $\mathbf{\Psi}$ is the $M \times N$ measurement matrix. Let $\mathbf{s} \stackrel{\text{def.}}{=} \mathbf{\Phi}\boldsymbol{\theta}$ where the matrix $\mathbf{\Phi}$ is a $N \times N$ orthonormal basis and $\boldsymbol{\theta}$ is the $N \times 1$ amplitude vector. By defining the $M \times N$ dictionary matrix $\mathbf{H} \stackrel{\text{def.}}{=} \mathbf{\Psi}\mathbf{\Phi}$, model (1) can be recast as

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n}. \quad (2)$$

Define $\rho_{\text{mes.}} \stackrel{\text{def.}}{=} \frac{M}{N}$ which quantifies the dictionary redundancy level. Classical sampling theory says that, to ensure no loss of information, the number of measurements, M , should be at least equal to N , or equivalently $\rho_{\text{mes.}} = 1$. In contrast, in CS theory this goal is reached for $\rho_{\text{mes.}} \ll 1$ as long as

the $N \times 1$ amplitude vector $\boldsymbol{\theta}$ is sparse in a given basis $\mathbf{\Phi}$ (e.g., the canonical basis of \mathbb{R}^N , the Fourier basis, etc.) [45]. This allows one to consider the CS theory to solve the ill-posed problem where the dictionary \mathbf{H} is a redundant matrix. A fundamental question in CS is to determine how large M must be to enable the recovery of $\boldsymbol{\theta}$. This will depend on the structural properties of \mathbf{H} . In particular, if \mathbf{H} satisfies the Restricted Isometry Property (RIP) of order qK [2,46]², i.e., if there exists $\epsilon_{qK} \in (0, 1)$ such that

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

for all $\boldsymbol{\theta} \in \mathcal{W}_K \stackrel{\text{def.}}{=} \{\boldsymbol{\theta} \in \mathbb{R}^N, \|\boldsymbol{\theta}\|_0 \leq K\}$ with $\|\cdot\|_0$ denoting the pseudo-norm l_0 , then practical recovery algorithms exist that can efficiently recover any K -sparse amplitude vector if

$$\rho_{\text{dic.}} = -O(\log \rho_{\text{spar.}}) \quad (4)$$

with $\rho_{\text{spar.}} \stackrel{\text{def.}}{=} \frac{K}{N}$ being the proportion of non-zero entries in $\boldsymbol{\theta}$, $\rho_{\text{dic.}} \stackrel{\text{def.}}{=} \frac{M}{K}$ the dictionary aspect ratio and $O(\cdot)$ denoting the “big- O ” notation. Determining whether a matrix \mathbf{H} satisfies the RIP is combinatorially complex but it was shown that, if $\mathbf{\Phi}$ has independent and identically distributed sub-Gaussian [47] entries of zero mean and variance $1/M$, then \mathbf{H} satisfies the RIP with high probability [48].³

Let \mathcal{S} be the support of $\boldsymbol{\theta}$, i.e. $\mathcal{S} \stackrel{\text{def.}}{=} \{i \in [1 : N], \theta_i \neq 0\}$ and $K \stackrel{\text{def.}}{=} \text{card}(\mathcal{S})$. The $K \times 1$ vector $\boldsymbol{\theta}_{\mathcal{S}}$ contains the K non-zero entries in $\boldsymbol{\theta}$, i.e. $\boldsymbol{\theta}_{\mathcal{S}} \stackrel{\text{def.}}{=} \{\theta_i, i \in \mathcal{S}\}$. With these definitions, the CS model (2) now reduces to

$$\mathbf{y} = \mathbf{H}_{\mathcal{S}}\boldsymbol{\theta}_{\mathcal{S}} + \mathbf{n} \quad \text{with} \quad \mathbf{H}_{\mathcal{S}} \stackrel{\text{def.}}{=} \mathbf{\Psi}\mathbf{\Phi}_{\mathcal{S}}. \quad (5)$$

The $N \times K$ matrix $\mathbf{\Phi}_{\mathcal{S}}$ results by dropping in $\mathbf{\Phi}$ the columns corresponding to the zero entries in $\boldsymbol{\theta}$.

We specify the context of our investigations in the following three assumptions:

- A_1 . The amplitude vector $\boldsymbol{\theta}$ is K -sparse with K *a priori* known. The vector $\boldsymbol{\theta}_{\mathcal{S}}$ is drawn from a distribution with pdf $p(\boldsymbol{\theta}_{\mathcal{S}})$. In Section III.A.2., we assume a generalized normal prior but other priors can be considered with marginal modifications of the presented results.
- A_2 . The support \mathcal{S} of cardinality K is modeled as a random (discrete) random variable specified by its distribution on set Ω which contains all K -sized subset of the indices $\{1, \dots, M\}$ with no repetitions of indices. Notice that since the support is a discrete variable, the regularity conditions necessary to derive the CRB are not fulfilled. The problem of joint support identification and amplitude estimation is traditionally solved as a model detection problem followed by a parameter estimation process. This detection/estimation scheme has been studied in [50]. In addition, notice that [26,28,29] have investigated oracle or genie-aided lower bounds for a deterministic amplitude vector.
- A_3 . To ensure the RIP property with high probability, \mathbf{H} is assumed to be a non-stochastic matrix in the noise domain

²Integer q depends on the considered estimator.

³Specifically, $\Pr\left(\left|\|\mathbf{H}\boldsymbol{\theta}\|^2 - \|\boldsymbol{\theta}\|^2\right| \geq \epsilon_{qK} \|\boldsymbol{\theta}\|^2\right) \leq 2e^{-MR\epsilon_{qK}^2}$ for some $R > 0$ [49].

¹Compressible means that the entries of $\boldsymbol{\theta}$ sorted in decreasing order are upper-bounded by a power law [37].

[28,32] whose entries are drawn independently from a sub-Gaussian distribution with zero mean and variance $1/M$. In the sequel, \mathbf{H} is treated as a known matrix in the context of the large system regime where $M, K \rightarrow \infty$ with $\rho_{\text{mes.}} \rightarrow \bar{\rho}_{\text{mes.}} \in (0, 1)$, $\rho_{\text{spar.}} \rightarrow \bar{\rho}_{\text{spar.}} \in (0, 1)$ and $\rho_{\text{dic.}} \rightarrow \bar{\rho}_{\text{dic.}} = -O(\log \bar{\rho}_{\text{spar.}}) \in (1, \infty)$ using relation (4).

Under these assumptions, we shall derive Bayesian lower bounds on the accuracy of estimating the amplitude vector.

III. ASYMPTOTIC BAYESIAN LOWER BOUNDS

Let $\hat{\boldsymbol{\theta}}_S(\mathbf{y})$ be an oracle-estimator of $\boldsymbol{\theta}_S$ that is derived assuming that the support \mathcal{S} is known (see e.g., [6,28,51,52] and Section IV for examples).

We define the conditional Bayesian covariance matrix and the conditional Bayesian mean-square error (BMSE) of $\hat{\boldsymbol{\theta}}_S(\mathbf{y})$ given \mathcal{S} :

$$\text{Var}(\hat{\boldsymbol{\theta}}_S) \stackrel{\text{def.}}{=} \mathbb{E}_{\mathbf{y}, \boldsymbol{\theta} | \mathbf{H}_S, \mathcal{S}} \left[(\boldsymbol{\theta}_S - \hat{\boldsymbol{\theta}}_S(\mathbf{y})) (\boldsymbol{\theta}_S - \hat{\boldsymbol{\theta}}_S(\mathbf{y}))^T \right] \quad (6)$$

$$\text{BMSE}_S \stackrel{\text{def.}}{=} \text{Tr} \left[\text{Var}(\hat{\boldsymbol{\theta}}_S) \right] = \mathbb{E}_{\mathbf{y}, \boldsymbol{\theta} | \mathbf{H}_S, \mathcal{S}} \left\| \boldsymbol{\theta}_S - \hat{\boldsymbol{\theta}}_S(\mathbf{y}) \right\|^2, \quad (7)$$

where $\mathbb{E}_{z|q}(\cdot)$ is the expectation over the conditional distribution of the random variable z given q and $\text{Tr}[\cdot]$ denoted the matrix trace operator. Averaging over the distribution of \mathcal{S} yields the global covariance matrix and the global BMSE of $\hat{\boldsymbol{\theta}}_S(\mathbf{y})$:

$$\text{Var}(\hat{\boldsymbol{\theta}}_S) \stackrel{\text{def.}}{=} \mathbb{E}_S \left[\text{Var}(\hat{\boldsymbol{\theta}}_S) \right] = \sum_{\mathcal{S} \in \Omega} \text{Pr}(\mathcal{S}) \text{Var}(\hat{\boldsymbol{\theta}}_S) \quad (8)$$

$$\text{BMSE}_\Omega \stackrel{\text{def.}}{=} \mathbb{E}_S \left[\text{BMSE}_S \right] = \sum_{\mathcal{S} \in \Omega} \text{Pr}(\mathcal{S}) \text{BMSE}_S. \quad (9)$$

The above quantities are coined *global* in the sense that they involve taking an expectation over the distribution of the support \mathcal{S} .

A. Global Bayesian lower bounds

1) *Derivation of the van Trees' lower bound:* We first define the $K \times K$ Bayesian information matrix (BIM) \mathbf{B}_S with entries

$$[\mathbf{B}_S]_{ij} = \left[\text{Var} \left(\frac{\partial \log p(\mathbf{y}, \boldsymbol{\theta}_S)}{\partial \theta_S} \right) \right]_{ij} \quad (10)$$

$$\stackrel{\text{def.}}{=} \mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}_S} \left(\frac{\partial \log p(\mathbf{y}, \boldsymbol{\theta}_S)}{\partial \theta_i} \frac{\partial \log p(\mathbf{y}, \boldsymbol{\theta}_S)}{\partial \theta_j} \right) \quad (11)$$

$$- \mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}_S} \left(\frac{\partial \log p(\mathbf{y}, \boldsymbol{\theta}_S)}{\partial \theta_i} \right) \mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}_S} \left(\frac{\partial \log p(\mathbf{y}, \boldsymbol{\theta}_S)}{\partial \theta_j} \right), \quad (12)$$

where $\theta_i, i \in \mathcal{S}$ is the i -th entry of vector $\boldsymbol{\theta}_S$. We have [20]

$$\text{Var}(\hat{\boldsymbol{\theta}}_S) \geq \mathbf{C}_S \stackrel{\text{def.}}{=} \mathbf{B}_S^{-1} \quad (13)$$

$$\text{BMSE}_S \geq \mathcal{C}_S \stackrel{\text{def.}}{=} \text{Tr}[\mathbf{C}_S] \quad (14)$$

$$\text{Var}(\hat{\boldsymbol{\theta}}_S) \geq \mathbb{E}_S[\mathbf{C}_S] \quad (15)$$

$$\text{BMSE}_\Omega \geq \mathcal{C}_\Omega \stackrel{\text{def.}}{=} \mathbb{E}_S[\mathcal{C}_S]. \quad (16)$$

Assume that the joint pdf $p(\mathbf{y}, \boldsymbol{\theta}_S)$ fulfills some mild regularity conditions [53]. Making use of the identity $\log p(\mathbf{y}, \boldsymbol{\theta}_S) = \log p(\mathbf{y} | \boldsymbol{\theta}_S) + \log p(\boldsymbol{\theta}_S)$ we can express the BIM as the sum of two terms:

$$\mathbf{B}_S = \mathbf{J}_S + \mathbf{G}_S, \quad (17)$$

where \mathbf{J}_S is the “data”-part and \mathbf{G}_S is the “prior”-part of the BIM.

Due to the model assumptions, the conditional distribution of the observation \mathbf{y} given $\boldsymbol{\theta}_S$ is Gaussian: $\mathbf{y} | \boldsymbol{\theta}_S \sim \mathcal{N}(\mathbf{H}_S \boldsymbol{\theta}_S, \sigma^2 \mathbf{I})$. So, $\log p(\mathbf{y} | \boldsymbol{\theta}_S)$ follows a log-normal distribution, meaning that its second derivative exists and that $\mathbb{E} \left(\frac{\partial \log p(\mathbf{y} | \boldsymbol{\theta}_S)}{\partial \theta_i} \right) = 0$. Consequently, for $i, j \in \mathcal{S} \times \mathcal{S}$, the “data”-part of the BIM is

$$[\mathbf{J}_S]_{ij} = \mathbb{E}_{\mathbf{y}, \boldsymbol{\theta}_S} \left(- \frac{\partial^2 \log p(\mathbf{y} | \mathbf{H}_S, \boldsymbol{\theta}_S)}{\partial \theta_i \partial \theta_j} \right) = \mathbb{E}_{\boldsymbol{\theta}_S} [\mathbf{F}_S]_{ij} \quad (18)$$

with the Fisher information matrix (FIM)

$$[\mathbf{F}_S]_{ij} = \mathbb{E}_{\mathbf{y} | \boldsymbol{\theta}_S} \left(- \frac{\partial^2 \log p(\mathbf{y} | \boldsymbol{\theta}_S)}{\partial \theta_i \partial \theta_j} \right). \quad (19)$$

For $\mathbf{y} | \boldsymbol{\theta}_S$ following a normal distribution $\mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$ we can invoke the Slepian-Bang formula [54]

$$[\mathbf{F}_S]_{ij} = \left(\frac{\partial \mathbf{m}}{\partial \theta_i} \right)^T \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{m}}{\partial \theta_j} + \frac{1}{2} \text{Tr} \left[\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \theta_j} \boldsymbol{\Sigma}^{-1} \right] \quad (20)$$

$$= \frac{1}{\sigma^2} [\mathbf{H}_S^T \mathbf{H}_S]_{ij}. \quad (21)$$

The last expression is obtained from $\frac{\partial \mathbf{m}}{\partial \theta_S} = [\dots \frac{\partial \mathbf{m}}{\partial \theta_i} \dots] = \mathbf{H}_S$ and $\frac{\partial \boldsymbol{\Sigma}}{\partial \theta_i} = \mathbf{0}$ for $i \in \mathcal{S}$.

The “prior”-part of the BIM has the form

$$[\mathbf{G}_S]_{ij} = \left[\text{Var} \left(\frac{\partial \log p(\boldsymbol{\theta}_S)}{\partial \theta_S} \right) \right]_{ij}, \quad (22)$$

where the prior distribution of $\boldsymbol{\theta}_S$ is left unspecified for the moment, see Section III-A3. Making use of the above results, we obtain:

$$\mathbf{B}_S = \frac{\mathbf{H}_S^T \mathbf{H}_S}{\sigma^2} + \mathbf{G}_S \quad (23)$$

$$\mathcal{C}_S = \text{Tr}(\mathbf{B}_S^{-1}) = \sigma^2 \text{Tr} \left(\left(\mathbf{H}_S^T \mathbf{H}_S + \sigma^2 \mathbf{G}_S \right)^{-1} \right). \quad (24)$$

2) *Maximal support set cardinal and model identifiability:* It is usual (see [35] for instance) to assume $K > M$ in a Bayesian framework, meaning that we wish to estimate more non-zero amplitudes than the number of measurements. In this case, $\mathbf{H}_S^T \mathbf{H}_S$ is rank deficient so that \mathbf{G}_S is selected to have full rank to ensure invertibility of \mathbf{B}_S . This is usually referred to as “Bayesian Regularization”. This strategy however leads to meaningless lower bounds especially in the low noise variance regime. We next address the low noise variance regime. We first show that no estimator with finite BMSE_S exists if $K > M$ and then provide a closed-form expression of the limit of BMSE_S when the noise variance converges towards 0.

a) *Numerical analysis of the ‘‘Bayesian Regularization’’ approach:* Define the condition number of the BIM: $\kappa(\mathbf{B}_S) \stackrel{\text{def.}}{=} \|\mathbf{B}_S^{-1}\|_2 \|\mathbf{B}_S\|_2 = \|\mathbf{C}_S\|_2 \|\mathbf{B}_S\|_2 \in [1, \infty)$ [55]. We have the following lemma.

Lemma 3.1: Assume that the 2-norm of the BIM is finite. Then BMSE_S is upper bounded according to:

$$\text{BMSE}_S \geq \frac{\kappa(\mathbf{B}_S)}{\|\mathbf{B}_S\|_2}. \quad (25)$$

Proof Invoking the covariance inequality property in (14) and $\text{Tr}[\mathbf{C}_S] > \|\mathbf{C}_S\|_2^4$, we obtain $\text{BMSE}_S > \|\mathbf{C}_S\|_2$. The above expression results then from the definition of the condition number of the BIM.

We can, now, apply the above Lemma in the context of the ‘‘Bayesian Regularization’’ approach and we are ready to show the following result.

Proposition 3.2: Assume that $K > M$ and $\|\mathbf{B}_S\|_2$ is finite. In the low noise variance regime, meaning $\sigma^2 \rightarrow 0$, no estimator $\hat{\theta}_S(\mathbf{y})$ with a finite BMSE_S exists.

Proof We can write for the condition number of the BIM

$$\kappa(\mathbf{B}_S) = \kappa\left(\frac{\mathbf{H}_S^T \mathbf{H}_S}{\sigma^2} + \mathbf{G}_S\right) = \kappa(\mathbf{H}_S^T \mathbf{H}_S + \sigma^2 \mathbf{G}_S). \quad (26)$$

Using this expression, we have

$$\kappa(\mathbf{B}_S) \xrightarrow{\sigma^2 \rightarrow 0} \kappa(\mathbf{H}_S^T \mathbf{H}_S) = \infty \quad (27)$$

because matrix $\mathbf{H}_S^T \mathbf{H}_S$ is singular when $K > M$. Now using Lemma 3.1 yields $\text{BMSE}_S \xrightarrow{\sigma^2 \rightarrow 0} \infty$ provided the 2-norm of \mathbf{B}_S is finite.

As we can see, computing the inverse of \mathbf{B}_S is an ill-conditioned problem in the low noise variance regime if $K > M$. In addition, an infinite condition number implies that the class of estimators considered in [35] exhibits infinite BMSE_S . This also means that the ‘‘Bayesian Regularization’’ philosophy yields useless Bayesian lower bounds in realistic operational context.

Deriving lower bounds with singular FIM is an important problem tackled recently in e.g., [56,57]. The singularity of the FIM can be interpreted as a lack of identifiability of the observation model due to parameters ambiguities, e.g. when the model is underdetermined (as just considered), but also due to scaling and permutation [58].

b) *BMSE limit where $K > M$ and independent amplitudes:* Assume $K > M$, thus $\mathbf{H}_S^T \mathbf{H}_S$ is a M -rank matrix and its SVD is $\mathbf{H}_S^T \mathbf{H}_S = \sum_{i=1}^K d_i \mathbf{u}_i \mathbf{u}_i^T$ where \mathbf{u}_i and d_i are the i -th singular vector and value, respectively. Due to the singularity of matrix $\mathbf{H}_S^T \mathbf{H}_S$, we have $d_i = 0$, $M+1 \leq i \leq K$. Now consider a prior that stipulates the entries of the amplitude vector to be independent, so that $\mathbf{G}_S = c \mathbf{I}_K$ for some positive c . Then the BCRB is given by

$$\mathbf{C}_S = \text{Tr} \left(\left(\frac{\mathbf{H}_S^T \mathbf{H}_S}{\sigma^2} + c \mathbf{I}_K \right)^{-1} \right) = \text{Tr}(\mathbf{W}^{-1}), \quad (28)$$

⁴Notice that \mathbf{B}_S is assumed to be non-singular. Thus, $\mathbf{C}_S = \mathbf{B}_S^{-1}$ exists and is full-rank. In this case, the inequality is strict.

where \mathbf{W} is a diagonal matrix with diagonal entries $\mathbf{W}_{ii} = \frac{d_i}{\sigma^2} + c$ for $1 \leq i \leq M$ and $\mathbf{W}_{ii} = c$ for $M+1 \leq i \leq K$. Thus, the conditional BMSE is lower bounded by the ‘‘regularized’’ BCRB according to

$$\text{BMSE}_S \geq \sum_{i=1}^M \frac{\sigma^2}{d_i + \sigma^2 c} + \frac{K-M}{c} \geq \frac{K-M}{c}. \quad (29)$$

According to this result the BMSE_S of any estimator is lower-bounded by $\frac{K-M}{c}$ and therefore does not converge to 0 as the noise variance converges towards 0. In our view, this is a limitation of the ‘‘Bayesian regularization’’ approach. As a conclusion, we claim that regularized lower bounds are irrelevant. In the sequel, it is always assumed that $K < M$.

3) *Amplitude vector prior:* The expressions for \mathbf{C}_S and its low and high noise variance approximations depend on the deterministic matrix \mathbf{G}_S which is determined by the prior on θ_S , i.e. $p(\theta_S)$. In this section, we specialize this prior to be a centered generalized normal distribution and compute explicit expressions for \mathbf{C}_S for it.

Specifically, we assume that θ_S follows a centralized generalized normal distribution [59]–[62], i.e. $\theta_S \sim \mathcal{GN}(\mathbf{0}, \mathbf{D}_S, \beta)$ where β is the real positive shape parameter and \mathbf{D}_S is the dispersion matrix of the distribution. We further assume that the entries of θ_S are independent, i.e. \mathbf{D}_S is diagonal and of the form $\mathbf{D}_S = \text{diag}\{(\alpha_i^2)_{i \in \mathcal{S}}\}$, where $\alpha_i > 0$ is the scale parameter of entry θ_i , $i \in \mathcal{S}$. The pdf of θ_S reads

$$p(\theta_S) \propto e^{-|\theta_S^T \mathbf{D}_S^{-1} \theta_S|^\beta} \quad (30)$$

where \propto means ‘‘proportional to’’. The scale parameters can be written as $\alpha_i^2 = \frac{\sigma_{\theta_i}^2}{\text{SDR}_\beta}$ where $\text{SDR}_\beta = \frac{\Gamma(3/\beta)}{\Gamma(1/\beta)}$ is the Signal to Distortion Ratio (SDR) [63] with $\Gamma(\cdot)$ the Gamma function [64] and $\sigma_{\theta_i}^2$ is the variance of θ_i , $i \in \mathcal{S}$.

The generalized normal distribution encompasses the Laplacian, Gaussian, uniform pdfs and degenerate Dirac distributions as special cases for $\beta = 1$, $\beta = 2$, $\beta \rightarrow \infty$ and $\beta \rightarrow 0$, respectively, see Fig. 1. We denote in the following $\mathbf{R}_{\theta_S} = \text{diag}(\sigma_{\theta_i}^2)_{i \in \mathcal{S}}$. Notice that $\mathbf{R}_{\theta_S} = \mathbf{D}_S$ when $\beta = 2$. We choose this prior due to its universality. Other priors can be adopted, though, with marginal modification of our derivations and results.

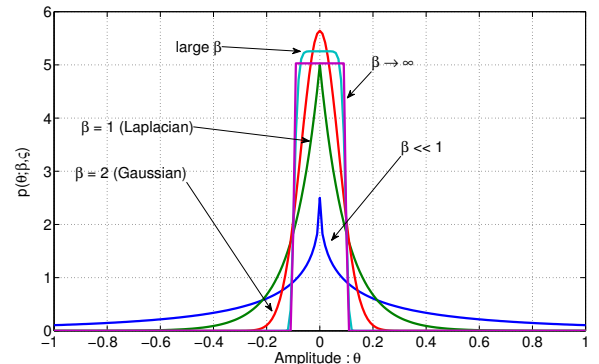


Fig. 1. Graph of the generalized normal pdf for different values of the shape parameter β .

We now provide the explicit expression for \mathbf{G}_S for this setting. Since the entries of $\boldsymbol{\theta}_S$ are independent, we have

$$[\mathbf{G}_S]_{ij} = \left[\text{Var} \left(\frac{\partial \log p(\boldsymbol{\theta}_S)}{\partial \boldsymbol{\theta}_S} \right) \right]_{ij} \quad (31)$$

$$= \begin{cases} \text{Var} \left(\frac{\partial \log p(\theta_i)}{\partial \theta_i} \right) & \text{for } i = j, \\ 0 & \text{otherwise} \end{cases} \quad (32)$$

Moreover from (30), $p(\theta_i) \propto e^{-\frac{|\theta_i|^\beta}{\alpha_i^\beta}}$. The score function of θ_i is thus given by

$$\frac{\partial \log p(\theta_i)}{\partial \theta_i} = -\frac{\partial}{\partial \theta_i} \frac{|\theta_i|^\beta}{\alpha_i^\beta} = -\frac{\beta}{\alpha_i^\beta} \text{sgn}(\theta_i) |\theta_i|^{\beta-1} \quad (33)$$

with $\text{sgn}(\cdot)$ denoting the sign function. It is straightforward to show the first moment of the score function is zero.

Because of the identity $\text{sgn}^2(\theta_i) = 1$ for $\theta_i \neq 0$ we can write

$$\text{Var} \left(\frac{\partial \log p(\theta_i)}{\partial \theta_i} \right) = \mathbb{E} \left(\left(\frac{\partial \log p(\theta_i)}{\partial \theta_i} \right)^2 \right) \quad (34)$$

$$= \frac{\beta^2}{\alpha_i^{2\beta}} \mathbb{E} \left(|\theta_i|^{2(\beta-1)} \right). \quad (35)$$

Assuming that $2(\beta - 1)$ is integer-valued, we have $\mathbb{E} \left(|\theta_i|^{2(\beta-1)} \right) = \alpha_i^{2(\beta-1)} \frac{\Gamma(2-1/\beta)}{\Gamma(1/\beta)}$ [60]. This finally entails

$$\mathbf{G}_S = g_\beta \text{SDR}_\beta \mathbf{R}_{\boldsymbol{\theta}_S}^{-1}, \quad (36)$$

with $g_\beta \stackrel{\text{def.}}{=} \beta^2 \frac{\Gamma(2-1/\beta)}{\Gamma(1/\beta)}$. Making use of the above propositions and definitions we obtain for the conditional BCRB

$$\mathcal{C}_S = \sigma^2 \text{Tr} \left(\left(\mathbf{H}_S^T \mathbf{H}_S + \sigma^2 g_\beta \text{SDR}_\beta \mathbf{R}_{\boldsymbol{\theta}_S}^{-1} \right)^{-1} \right). \quad (37)$$

B. Asymptotic analysis

The above expressions for \mathcal{C}_S still depend on the non-stochastic, known and randomly generated matrix \mathbf{H}_S . In this section, we leverage on results from the theory of random matrices to analyze the (almost sure) convergence of the BCRB in the asymptotic limit when $M, K \rightarrow \infty$ such that $\rho_{\text{dic.}} = M/K \rightarrow \bar{\rho}_{\text{dic.}} \in (1, \infty)$.

1) *Existence of a finite variance estimator in the asymptotic scenario:* As discussed in Section III-A2, a key factor influencing the behavior of the BCRB is the ‘‘numerical stability’’ of the inverse of matrix $\mathbf{H}_S^T \mathbf{H}_S$. More specifically, if matrix $\mathbf{H}_S^T \mathbf{H}_S$ is singular, then no estimator with finite variance [57] exists and the estimation of the entire parameter vector is impossible. To quantify this point, we have to study the condition number of $\mathbf{H}_S^T \mathbf{H}_S$. We can show that

$$\kappa(\mathbf{H}_S^T \mathbf{H}_S) = \frac{\lambda_{\max}(\mathbf{H}_S^T \mathbf{H}_S)}{\lambda_{\min}(\mathbf{H}_S^T \mathbf{H}_S)} \rightarrow \left(\frac{1 + \sqrt{1/\bar{\rho}_{\text{dic.}}}}{1 - \sqrt{1/\bar{\rho}_{\text{dic.}}}} \right)^2 \quad (38)$$

almost surely. Thus, the condition number remains low (‘‘numerical stability’’) if the number of measurements is much larger than the cardinality of the support set. Conversely, as $\bar{\rho}_{\text{dic.}}$ decreased towards 1, the condition number goes to infinity. This means that if the number of measurements

and the cardinality of the support set tend to be close, then the inversion of matrix $\mathbf{H}_S^T \mathbf{H}_S$ is in fact an ill-conditioned problem and no finite-variance estimator exists.

2) *Asymptotic Bayesian lower bound:* First, we shall require the following regularity assumption as the system dimensions grow: as $M \rightarrow \infty$ with $\rho_{\text{dic.}} \rightarrow \bar{\rho}_{\text{dic.}}$,

$$0 < \liminf_K \lambda_1(\mathbf{R}_{\boldsymbol{\theta}_S}) \leq \limsup_K \lambda_K(\mathbf{R}_{\boldsymbol{\theta}_S}) < \infty. \quad (39)$$

Based on [39], we have the following proposition.

Proposition 3.3: Assume that (39) holds. Then, as $M \rightarrow \infty$ with $\rho_{\text{dic.}} \rightarrow \bar{\rho}_{\text{dic.}}$,

$$\frac{1}{K} \mathcal{C}_S - \sigma^2 e_M \rightarrow 0 \quad (40)$$

almost surely, where e_M is the unique positive solution to the equation in e

$$e = \frac{1}{K} \text{Tr} \left(\frac{1}{1 + e \bar{\rho}_{\text{dic.}}^{-1}} \mathbf{I}_K + \sigma^2 g_\beta \text{SDR}_\beta \mathbf{R}_{\boldsymbol{\theta}_S}^{-1} \right)^{-1}. \quad (41)$$

Proposition 3.3 states that, as $M \rightarrow \infty$ with $\rho_{\text{dic.}} \rightarrow \bar{\rho}_{\text{dic.}}$, the random variable $\frac{1}{K} \mathcal{C}_S$ can be asymptotically well approximated by the deterministic, but not necessarily converging, quantity $\sigma^2 e_M$ where e_M is implicit. Observe that we can rewrite (41) as

$$e = \int \left(\frac{1}{1 + e \bar{\rho}_{\text{dic.}}^{-1}} + \sigma^2 g_\beta \text{SDR}_\beta t^{-1} \right)^{-1} \mu_M(dt), \quad (42)$$

where $\mu_M \triangleq \frac{1}{K} \sum_{i=1}^K \delta_{\lambda_i(\mathbf{R}_{\boldsymbol{\theta}_S})}$ is the empirical (normalized) counting measure of the eigenvalues of $\mathbf{R}_{\boldsymbol{\theta}_S}$. Assuming now that $\mu_M \rightarrow \mu$ weakly for some measure μ with support $\text{supp}(\mu) \subset (0, \infty)$ and using the fact that $e_M \leq 1/(\sigma^2 g_\beta \text{SDR}_\beta) \limsup_K \lambda_K(\mathbf{R}_{\boldsymbol{\theta}_S})$, we find that $e_M \rightarrow e^{(\infty)}$, unique positive solution to the equation in e

$$e = \int \left(\frac{1}{1 + e \bar{\rho}_{\text{dic.}}^{-1}} + \sigma^2 g_\beta \text{SDR}_\beta t^{-1} \right)^{-1} \mu(dt). \quad (43)$$

In particular,

$$\frac{1}{K} \mathcal{C}_S \rightarrow \mathcal{C}^{(\infty)} \triangleq \sigma^2 e^{(\infty)} \quad (44)$$

almost surely.

a) *Asymptotic BCRB in extreme noise variance regimes:* Although Proposition 3.3 does not provide any closed-form expression. We study below its approximation in the low noise variance regime.

Proposition 3.4: For $M \rightarrow \infty$ with $\rho_{\text{dic.}} \rightarrow \bar{\rho}_{\text{dic.}}$ and $\sigma \rightarrow 0$ we have

$$\frac{1}{\sigma^2 K} \mathcal{C}_S - \frac{\bar{\rho}_{\text{dic.}}}{\bar{\rho}_{\text{dic.}} - 1} \rightarrow 0 \quad (45)$$

almost surely.

Proof To prove the above proposition recall from [65] that the extension of e_M to the mapping $\mathbb{C} \rightarrow \mathbb{C} \setminus \text{supp}(\mu_M)$, $z \mapsto e_M(z)$, unique solution of (41) with $\sigma^2 = -z$, is the Stieltjes transform of the measure with support $\text{supp}(\mu_M)$. Since $\bar{\rho}_{\text{dic.}} > 1$ and (39) is in place, $\bigcup_{M=1}^\infty \text{supp}(\mu_M) \subset (0, \infty)$, so that $e_M(z)$ is analytic at $z = 0$. In particular, $e_M \rightarrow e_M^\circ$ as $\sigma \rightarrow 0$ with e_M° the unique positive solution

to $e_M^\circ = 1 + e_M^\circ \bar{\rho}_{\text{dic.}}^{-1}$, which here is explicitly given by $e_M^\circ = \frac{\bar{\rho}_{\text{dic.}}}{\bar{\rho}_{\text{dic.}} - 1}$. Now, from [40], $\liminf_n \lambda_{\min}(\mathbf{H}_S^T \mathbf{H}_S) > 0$ almost surely, so that $\sigma^{-2} \frac{1}{K} \mathcal{C}_S$ is uniformly (across M) continuous at $\sigma = 0$.

The above proposition means that the BCRB in the large dimensional regime is a function of the noise variance and of $\bar{\rho}_{\text{dic.}}$ only. That is, the actual support set \mathcal{S} only intervenes through its cardinality K . An alternative expression involves the ratio between the sparsity ratio of the amplitude vector over the dictionary redundancy level.

Remark 3.5: For small noise variance, the asymptotic BCRB can be closely approximated by

$$\mathcal{C}^{(\infty)} \approx \sigma^2 \frac{\rho_{\text{dic.}}}{\rho_{\text{dic.}} - 1} = \frac{\sigma^2}{1 - \frac{\rho_{\text{spar.}}}{\rho_{\text{mes.}}}}. \quad (46)$$

Remark 3.6: In the high noise variance regime, irrespective of random matrix considerations, it is immediate that

$$\frac{1}{K} \mathcal{C}_S - g_\beta \text{SDR}_\beta \frac{1}{K} \text{Tr}(\mathbf{R}_{\theta_S}) \rightarrow 0 \quad (47)$$

as $\sigma \rightarrow \infty$.

The above remark is evident since the measurements carry no information when the noise variance is large, only the prior on the parameters is relevant in that regime. In the next section, we derive the BCRB for an arbitrary noise variance when the prior is a centralized generalized normal distribution with identical dispersion.

b) Generalized normal prior with identical dispersion:

When $\mathbf{R}_{\theta_S} = \sigma^2 \text{SNR} \mathbf{I}_K$ (identical dispersion regime) with

$$\text{SNR} \stackrel{\text{def.}}{=} \frac{\sigma_\theta^2}{\sigma^2} = \frac{\Gamma(3/\beta) \alpha^2}{\Gamma(1/\beta) \sigma^2} \quad (48)$$

Proposition 3.3 particularizes as follows:

Proposition 3.7: Assume a generalized normal prior with identical dispersion for the amplitudes. Then, as $M \rightarrow \infty$ with $\rho_{\text{dic.}} \rightarrow \bar{\rho}_{\text{dic.}}$

$$\frac{1}{K} \mathcal{C}_S \rightarrow \mathcal{C}^{(\infty)} \stackrel{\text{def.}}{=} r \left(\frac{\text{SNR}}{g_\beta \text{SDR}_\beta} \right) \quad (49)$$

almost surely, where

$$r(x) = \frac{\mathcal{U}(x)}{2} \left(\sqrt{1 + \frac{4\mathcal{V}(x)}{\mathcal{U}(x)^2}} - 1 \right) \quad (50)$$

with $\mathcal{U}(x) = \sigma^2 (\bar{\rho}_{\text{dic.}} + (\bar{\rho}_{\text{dic.}} - 1)x)$ and $\mathcal{V}(x) = \sigma^4 \bar{\rho}_{\text{dic.}} x$.

Proof Based on Proposition 3.3, solving (41) for e leads to the resolution of a quadratic polynomial in e defined by

$$P(e; x) = e^2 + \mathcal{U}(x) e - \mathcal{V}(x), \quad (51)$$

where $x = \frac{\text{SNR}}{g_\beta \text{SDR}_\beta}$. Choosing the positive root leads to the above proposition.

We now derive a closed-form expression of the BCRB for the important case where the amplitude vector is highly sparse.

Proposition 3.8: Assume a generalized normal prior with identical dispersion for the amplitudes. If the amplitude vector is highly sparse, i.e., $\bar{\rho}_{\text{dic.}} \gg 1$, we have

$$\mathcal{C}^{(\infty)} = \frac{\sigma_\theta^2}{\text{SNR} + g_\beta \text{SDR}_\beta} + O(\bar{\rho}_{\text{dic.}}^{-2}). \quad (52)$$

Proof The first-order Taylor approximation of $r(x)$ in (50) for large $\bar{\rho}_{\text{dic.}}$ reads $r(x) = \sigma^2 \frac{x}{1+x} + O(\bar{\rho}_{\text{dic.}}^{-2})$. Substituting $x = \frac{\text{SNR}}{g_\beta \text{SDR}_\beta}$ yields $r(\frac{\text{SNR}}{g_\beta \text{SDR}_\beta}) = \frac{\sigma_\theta^2}{\text{SNR} + g_\beta \text{SDR}_\beta} + O(\bar{\rho}_{\text{dic.}}^{-2})$.

c) Asymptotic BCRB: In the large system limit the BCRB defined in (16) takes a very simple expression.

Proposition 3.9: As $M \rightarrow \infty$ with $\rho_{\text{dic.}} \rightarrow \bar{\rho}_{\text{dic.}}$ and $\sigma \rightarrow 0$

$$\frac{1}{K} \mathcal{C}_\Omega - \mathcal{C}^{(\infty)} \rightarrow 0. \quad (53)$$

Proof As noted in Proposition 3.4, the asymptotic BCRB, $\mathcal{C}^{(\infty)}$, derived for a given realization of the support \mathcal{S} is only a function of the ratio $\bar{\rho}_{\text{dic.}}$. Thus,

$$\mathcal{C}_\Omega = \sum_{\mathcal{S} \in \Omega} \Pr(\mathcal{S}|L) \mathcal{C}_S \rightarrow \mathcal{C}^{(\infty)} \lim_{M, K \rightarrow \infty} \underbrace{\sum_{\mathcal{S} \in \Omega} \Pr(\mathcal{S}|L)}_1 = \mathcal{C}^{(\infty)}. \quad (54)$$

The above proposition is important from a computational point of view since the brute force computation of \mathcal{C}_Ω , which involves a costly numerical matrix inversion, becomes rapidly intractable as M and K grow large. Moreover, (53) is valid for any prior distribution of the random support.

IV. STATISTICAL ANALYSIS OF THE ORACLE LMMSE ESTIMATOR

Oracle estimators can be viewed as a gold standard against which practical sub-optimal approaches are compared [51,66]. This class of estimators is also called genie-aided estimators since they assume perfect knowledge of the support set. This assumption seems severe at first glance but in practice, it can be checked by means of numerical simulations that if the noise variance is sufficiently small, the assumption of perfect estimation of the support set is realistic.

The optimal estimator for the Bayesian linear model is the Minimum Mean Squared Error (MMSE) estimator [44]. Unfortunately, the variance of the MMSE estimator admits no closed-form expression unless the amplitude vector follows a Gaussian prior. For other priors, as for instance the Laplacian distribution, the analytic derivation of the variance of the MMSE estimator is intractable. Thus, we focus our analysis on the oracle linear MMSE (LMMSE) estimator which we denote in the sequel as $\hat{\theta}_S(\mathbf{y})$.

The Bayesian Gauss-Markov Theorem [44] provides conditional variance of the oracle LMMSE estimator:

$$\text{Var}(\hat{\theta}_S) = \text{Tr} \left(\left(\frac{1}{\sigma^2} \mathbf{H}_S^T \mathbf{H}_S + \mathbf{R}_{\theta_S}^{-1} \right)^{-1} \right). \quad (55)$$

When the amplitude prior is the generalized normal prior with identical dispersion

$$\text{Var}(\hat{\theta}_S) = \sigma_\theta^2 \text{Tr} \left((\text{SNR} \mathbf{H}_S^T \mathbf{H}_S + \mathbf{I}_K)^{-1} \right). \quad (56)$$

A. Asymptotic variance

Notice that expression (56) is formally similar to that of the BCRB in (37). Consequently, by using the same method as applied to prove Proposition 3.7 we obtain the following proposition.

Proposition 4.1: For $M \rightarrow \infty$ with $\rho_{\text{dic.}} \rightarrow \bar{\rho}_{\text{dic.}}$.

$$\frac{1}{K} \text{Var}(\hat{\theta}_S) \rightarrow \text{Var}^{(\infty)}(\hat{\theta}) \stackrel{\text{def.}}{=} r(\text{SNR}) \quad (57)$$

almost surely where $r(x)$ is defined in (50).

Remark 4.2: The variance of the oracle LMMSE estimator is given by

$$\text{Var}_{\Omega}(\hat{\theta}) = \sum_{S \in \Omega} \Pr(S) \text{Var}(\hat{\theta}_S). \quad (58)$$

In the large dimensional limit the variance of the oracle LMMSE estimator has the same invariance property towards the support as given in Proposition 3.9 for the asymptotic BCRB. Thus, in the sequel we only study $\text{Var}^{(\infty)}(\hat{\theta})$.

In the following proposition, we derive a closed-form expression for the variance of the oracle-LMMSE for the important case where the amplitude vector is highly sparse.

Proposition 4.3: Provided the amplitude vector is highly sparse, i.e., $\bar{\rho}_{\text{dic.}} \gg 1$, we have

$$\text{Var}^{(\infty)}(\hat{\theta}) = \frac{\sigma_{\theta}^2}{\text{SNR} + 1} + O(\bar{\rho}_{\text{dic.}}^2). \quad (59)$$

Proof The proof is straightforwardly derived from the proof of Proposition 3.8 for $x = \text{SNR}$.

B. Asymptotic statistical efficiency of the oracle-LMMSE estimator

Using the derived closed-form expressions, it is now easy to compare the asymptotic variance of the oracle-LMMSE estimator with the Bayesian lower bound. With the definition

$$C_{\beta} = \frac{\mathcal{C}^{(\infty)}}{\text{Var}^{(\infty)}(\hat{\theta})} = \frac{r\left(\frac{\text{SNR}}{g_{\beta} \text{SDR}_{\beta}}\right)}{r(\text{SNR})}. \quad (60)$$

we obtain the next proposition.

Proposition 4.4:

\mathcal{P}_1 . For a sufficiently large SNR ratio C_{β} is given by

$$C_{\beta} \begin{cases} = 1, & \text{for } \beta = 2, \\ < 1, & \text{for } \beta \neq 2. \end{cases} \quad (61)$$

\mathcal{P}_2 . At small SNR ratio C_{β} is given by

$$C_{\beta} = \frac{1}{g_{\beta} \text{SDR}_{\beta}} + O(\text{SNR}^2). \quad (62)$$

Proof \mathcal{P}_1 . First, notice that if the prior is Gaussian, $g_2 \text{SDR}_2 = 1$. Thus $\mathcal{C}^{(\infty)} = \text{Var}^{(\infty)}(\hat{\theta})$.

Second, it is straightforward to see that $r(x)$ is an increasing function and since $\frac{\text{SNR}}{g_{\beta} \text{SDR}_{\beta}} < \text{SNR}$ for all $\beta \neq 2$ we conclude that $C_{\beta} < 1$ and therefore that $\mathcal{C}^{(\infty)} < \text{Var}^{(\infty)}(\hat{\theta})$.

\mathcal{P}_2 . Assuming that x is small, we have

$$f(x) = \frac{4\bar{\rho}_{\text{dic.}}x}{(\bar{\rho}_{\text{dic.}} + (\bar{\rho}_{\text{dic.}} - 1)x)^2} = \frac{4}{\bar{\rho}_{\text{dic.}}}x + O(x^2) \quad (63)$$

$$g(x) = \sqrt{1 + f(x)} - 1 = \frac{1}{2}f(x) + O(f(x)^2). \quad (64)$$

Using these approximations and (50), we obtain an approximated root for small x according to

$$r(x) = \sigma^2 x + O(x^2). \quad (65)$$

Thus, at small SNR we have

$$\mathcal{C}^{(\infty)} = \frac{\sigma_{\theta}^2}{g_{\beta} \text{SDR}_{\beta}} + O(\text{SNR}^2) \quad (66)$$

$$\text{Var}^{(\infty)}(\hat{\theta}) = \sigma_{\theta}^2 + O(\text{SNR}^2). \quad (67)$$

Now, inserting these approximations in (60) yields (62).

From Proposition 4.4, we draw the following conclusions:

- At small SNR the asymptotic variance of the oracle LMMSE estimator is proportional to $\mathcal{C}^{(\infty)}$. The proportionality coefficient C_{β} expressed in dB reads

$$C_{\beta} [\text{dB}] = -10 \log_{10}(g_{\beta}) - 10 \log_{10}(\text{SDR}_{\beta}). \quad (68)$$

It is tabulated in Table I for selected values of β .

An important result in the context of the CS is that the oracle LMMSE estimator for the Laplacian prior ($\beta = 1$) is not statistically efficient at small SNR and the proportionality coefficient is about -3 dB.

- At high SNR the variance of the oracle LMMSE estimator, $\text{Var}^{(\infty)}(\hat{\theta})$, and bound $\mathcal{C}^{(\infty)}$ have the same asymptotic behavior wrt. the SNR. Specifically, these quantities are almost identical for $\beta \neq 2$ and equal if $\beta = 2$. When the prior is Gaussian the oracle LMMSE estimator, which coincides with the oracle MMSE estimator, is statistically efficient. However, for any other prior the oracle LMMSE estimator is never statistically efficient. This is also true for the Laplacian prior.

V. APPLICATION TO CS OF FINITE-RATE-OF-INNOVATION (FRI) SIGNALS

A. The FRI model

Consider a non-bandlimited continuous-time signal with a finite number of weighted Dirac impulses:

$$x(t) = \sum_{\ell \in S} \theta_{\ell} \delta(t - \tau_{\ell}), \quad (69)$$

where τ_{ℓ} and a_{ℓ} are respectively the time-delay and the amplitude of the ℓ -th Dirac. Signals of this form are sparse in time and encompass a wide range of realistic signals. Notice that signal $x(t)$ is non-bandlimited and thus cannot be sampled in the Shannon framework without error. However, a major theory has been developed in [67,68] which allows to overcome the Shannon theory [4]. So, we consider the following estimation problem described in Fig. 2. Consider a normalized sinc sampling kernel defined by $g(t) = \frac{1}{T_S} \text{sinc}\left(\frac{t}{T_S}\right)$ where $1/T_S$ is the sampling rate. Then, uniform sampling at rate $1/T_S$ of signal $x(t)$ yields the samples

$$s_k = \int_{-\infty}^{\infty} g(t - kT_S)x(t)dt \quad (70)$$

$$= \frac{1}{T_S} \sum_{\ell \in S} \theta_{\ell} \text{sinc}\left(\frac{\tau_{\ell}}{T_S} - k\right), \quad k \in [1 : N]. \quad (71)$$

TABLE I
RATIO C_β IN DB

β	1 (Lap.)	2 (Gauss.)	3	4	5	6	7	8	9	10
C_β [dB]	-3	0	-0.5	-1.4	-2.2	-2.9	-3.5	-4	-4.6	-5

Define the orthogonal basis matrix (see Fig. 2-(a)) as

$$[\Phi]_{kk'} = \frac{1}{T_S} \text{sinc} \left(\frac{\tau_{k'}}{T_S} - k \right), \quad k, k' \in [1 : N]. \quad (72)$$

and assume that vector θ is K -sparse. We then have

$$\mathbf{s} = \Phi \theta = \Phi_S \theta_S, \quad (73)$$

where \mathcal{S} is the support set. Vector \mathbf{s} is a stream of filtered Dirac impulses. It is well-known that this class of signals has a finite rate of innovation where the rate of innovation is defined in function of the number of degrees of freedom ($2K$) per unit of time (N). In our scenario, this rate is given by $\rho = 2\bar{\rho}_{\text{spar}}$. This means that in the FRI framework, the number of pulse, K , grows at the same rate as the window length N or equivalently as M . In addition, in the sequel, we consider sampling rates given by $1/T_S \rightarrow 2\bar{\rho}_{\text{spar}}$.

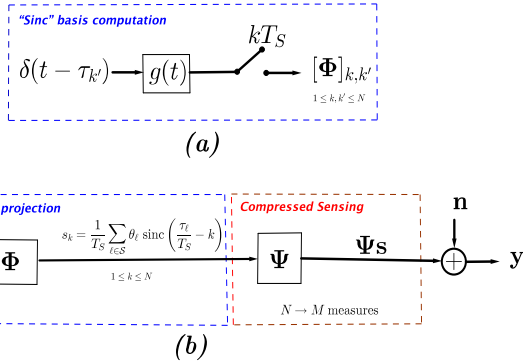


Fig. 2. (a) Derivation of the basis matrix based on the “sinc” sampling kernel $g(t)$, (b) CS for FRI signals.

Define a $M \times N$ non-stochastic but randomly generated measurement matrix Ψ with $M < N$, see also Fig. 2-(b). The under-sampled observation vector with ratio ρ_{mes} is given by

$$\mathbf{y} = \Psi \mathbf{s} + \mathbf{n} = \Psi \Phi_S \theta_S + \mathbf{n}, \quad (74)$$

where \mathbf{s} is the uncompressed measurement vector of size N . First notice that due to the orthonormality of the “sinc” basis, matrix $\Psi \Phi$ satisfies the RIP conditions with high probability. Due to identification constraint, we impose that $K < M$. As indicated before K grows as the same rate as N , which implies that M grows at the same rate as K .

B. Numerical investigations

We assume a Laplacian prior θ_S with location vector $\mathbf{0}$ and scale matrix $\frac{1}{2} \mathbf{R}_{\theta_S}$, i.e., $\theta_S \sim \mathcal{GN}(\mathbf{0}, \frac{1}{2} \mathbf{R}_{\theta_S}, 1)$.

In Fig. 3, we compare the asymptotic BCRB and the variance of the LMMSE estimator with the corresponding matrix-based expressions given by (37) and (56) normalized

by $K = 5$. We can see that that in the asymptotic regime is already reached for this small value of K . This means that RMT provides accurate results already for practical scenarios. We also notice that for the Laplacian prior the oracle-LMMSE is suboptimal in the low SNR regime.

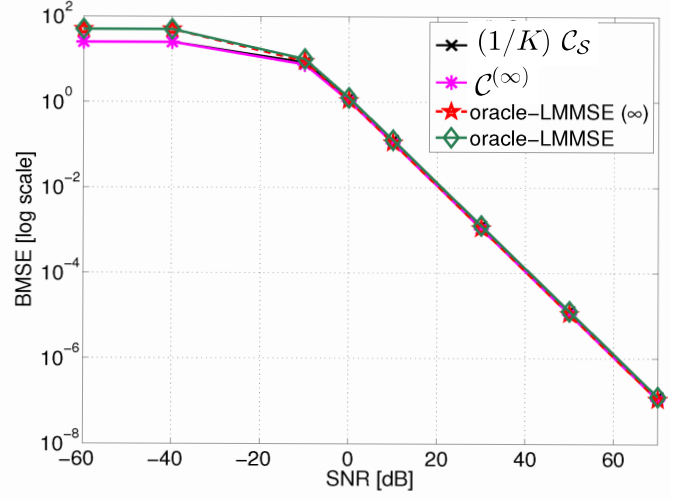


Fig. 3. Two BCRBs and variance of the oracle LMMSE estimator vs. SNR: $K = 5$, $M = 50$, $N = 100$.

We now compare the asymptotic BCRB and the variance of the oracle LMMSE estimator to those of three popular sparse estimators, namely basis pursuit denoising (BPDN), orthogonal matching pursuit (OMP), and compressive sampling matching pursuit (CoSaMP):

- BPDN solves the following optimization problem:

$$\hat{\theta}_S = \min_{\theta} \|\theta\|_1^2 \text{ subject to } \|\mathbf{y} - \mathbf{H}\theta\|^2 \leq \sigma^2. \quad (75)$$

The algorithm exploits the following knowledge: (i) the amplitude prior follows a Laplacian prior and (ii) the noise variance is known. The cardinality of the support has to be estimated. The minimization problem (75) is solved by using the standard SPGL1 MatLab toolbox [17] with a maximal number of iterations of 1000.

- OMP and the CoSaMP [69,70] belong to the family of greedy algorithms. In contrast to BPDN, they are faster and generic in the sense that the amplitude prior does not need to be specified. In addition, they do not need to know the noise variance but assume the knowledge of the cardinality of the support.

In Fig. 4, we compare the estimation accuracy of the above sparse estimators with the asymptotic BCRB and the variance of the oracle-LMMSE estimator in the context of CS of FRI signals. We perform our analysis for the high and low SNR regimes.

- *High SNR regime:* We can see that OMP and CoSaMP have a BMSE very close to the BCRB in the high SNR regime. This property can be explained as follows. At high SNR the support set is estimated with high accuracy, provided its cardinality is known; thus OMP inherits the optimality of the least square estimator in case of Gaussian noise. CoSaMP is a modified OMP, so it has globally the same behavior as the latter. In addition, since the specific shape of the amplitude prior has a marginal impact on the BCRB at sufficiently high SNR and the noise variance is decoupled from the other parameter estimates, it is natural to conclude that the proposed lower bound can well predict the performance of OMP and CoSaMP in this regime. Conversely, the poor accuracy of BPDN in the high SNR regime is due to the error in the estimation of the cardinality of the support.
- *Low SNR regime:* In this regime, the BCRB is solely governed by the *a priori* distribution of the amplitudes since the available measurements are heavily corrupted by noise. Thus, in this regime BPDN reaches the BCRB derived for a Laplacian prior. OMP and CoSaMP do not exploit the knowledge of the prior. This explains why their MSE is very far from the BCRB and goes to infinity as $\sigma^2 \rightarrow \infty$. So, in the low SNR regime the BCRB provides a good prediction of the performance of algorithms based on criterion (75).

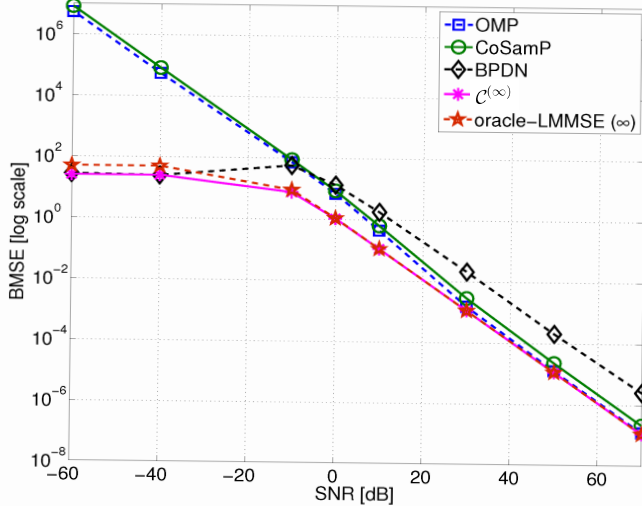


Fig. 4. Asymptotic BCRB, variance of the oracle LMMSE estimator, and BMSE of three benchmark sparse estimators vs. SNR: $K = 5$, $M = 50$, $N = 100$.

VI. CONCLUSION

In the context of the CS problem, we derive and study the Bayesian performance estimation for K -sparse generalized normal amplitudes belonging to a random support of known cardinality K for large non-stochastic but randomly generated dictionaries. By “large”, we assume in this work that the dimensions of the dictionary grows at the same rate. This context is well adapted to exploit results from Random Matrix

Theory. Compact closed-form expressions of the BCRB are derived in (i) extreme SNR regimes, (ii) for highly-sparse amplitude vector and (iii) for generalized normal amplitudes with identical dispersion matrix. In particular, we show that in the asymptotic context our Bayesian lower bound is valid for any support priors. This result is important from a computational point of view. The second part of this contribution presents a statistical efficiency analysis of the oracle LMMSE estimator. We show that the oracle-LMMSE for any priors (excepted the Gaussian one) is never efficient in particular at low SNR. Finally, we apply our results in the context of FRI signal sampling and the derived bounds are compared to the performance of the most popular sparse estimators such as OMP, CoSaMP and the BPDN.

REFERENCES

- [1] E. J. Candes and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [2] R. Baraniuk, “Compressive sensing [lecture notes],” *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.
- [3] M. Elad and I. Yavneh, “A plurality of sparse representations is better than the sparsest one alone,” *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4701–4714, 2009.
- [4] M. Unser, “Sampling-50 years after Shannon,” *Proceedings of the IEEE*, vol. 88, no. 4, pp. 569–587, 2000.
- [5] N. L. Pedersen, C. N. Manchón, D. Shutin, and B. H. Fleury, “Application of bayesian hierarchical prior modeling to sparse channel estimation,” in *IEEE International Conference on Communications (ICC)*. IEEE, 2012, pp. 3487–3492.
- [6] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, “Compressed channel sensing: A new approach to estimating sparse multipath channels,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.
- [7] S. F. Cotter and B. D. Rao, “Sparse channel estimation via matching pursuit with application to equalization,” *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 374–377, 2002.
- [8] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing MRI,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, 2008.
- [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] 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.
- [11] S. Mallat, *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [12] Y. C. Pati, R. Rezaifar, and P. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition,” in *Conference Record of The Twenty-Seventh Asilomar Conference*. IEEE, 1993, pp. 40–44.
- [13] P. Schniter, L. C. Potter, and J. Ziniel, “Fast Bayesian matching pursuit,” in *IEEE Information Theory and Applications Workshop*, 2008, pp. 326–333.
- [14] J. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [15] 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.
- [16] S. J. Wright, R. D. Nowak, and M. A. Figueiredo, “Sparse reconstruction by separable approximation,” *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, 2009.
- [17] 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>.
- [18] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, “Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.
- [19] M. Zibulevsky and M. Elad, “L1-L2 optimization in signal and image processing,” *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 76–88, 2010.

- [20] H. L. Van Trees and K. L. Bell, "Bayesian bounds for parameter estimation and nonlinear filtering/tracking," *AMC*, vol. 10, p. 12, 2007.
- [21] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Transactions on Information Theory*, vol. 15, no. 3, pp. 386–391, May 1969.
- [22] A. J. Weiss and E. Weinstein, "A lower bound on the mean-square error in random parameter estimation (corresp.)," *IEEE Transactions on Information Theory*, vol. 31, no. 5, pp. 680–682, 1985.
- [23] E. Weinstein, "Relations between Belini-Tartara, Chazan-Zakai-Ziv, and Wax-Ziv lower bounds," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 342–343, Mar 1988.
- [24] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, "Extended Ziv-Zakai lower bound for vector parameter estimation," *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 624–637, 1997.
- [25] D. Fraser and I. Guttman, "Bhattacharyya bounds without regularity assumptions," *The Annals of Mathematical Statistics*, pp. 629–632, 1952.
- [26] 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.
- [27] P. Pakrooh, L. L. Scharf, A. Pezeshki, and Y. Chi, "Analysis of fisher information and the Cramér-Rao bound for nonlinear parameter estimation after compressed sensing," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 6630–6634.
- [28] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramér-Rao bound in noisy compressed sensing," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 518–526, 2012.
- [29] 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.
- [30] R. Boyer, B. Babadi, N. Kalouptsidis, and V. Tarokh, "Errata to " Asymptotic Achievability of the Cramér-Rao Bound for Noisy Compressive Sampling "," Apr. 2016, working paper or preprint. [Online]. Available: <https://hal-supelec.archives-ouvertes.fr/hal-01299004>
- [31] A. Florescu, E. Chouzenoux, J.-C. Pesquet, and S. Ciochina, "Cramér-Rao bound for a sparse complex model," in *International Conference on Communications (COMM)*. IEEE, 2014, pp. 1–4.
- [32] M. Shaghghi, S. Vorobyov *et al.*, "Cramér-Rao bound for sparse signals fitting the low-rank model with small number of parameters," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1497–1501, 2015.
- [33] K. Qiu and A. Dogandzic, "Variance-component based sparse signal reconstruction and model selection," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 2935–2952, 2010.
- [34] P. Stoica, G. Larsson, and A. B. Gershman, "The stochastic CRB for array processing: A textbook derivation," *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 148–150, 2001.
- [35] R. Prasad and C. R. Murthy, "Cramér-Rao-type bounds for sparse Bayesian learning," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 622–632, 2013.
- [36] M. N. E. Korso, R. Boyer, P. Larzabal, and B. H. Fleury, "Estimation performance for the Bayesian Hierarchical linear model," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 488–492, April 2016.
- [37] V. Cevher, "Learning with compressible priors," in *Advances in Neural Information Processing Systems*, 2009, pp. 261–269.
- [38] P. Pal and P. Vaidyanathan, "Parameter identifiability in sparse bayesian learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 1851–1855.
- [39] J. W. Silverstein and Z. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate analysis*, vol. 54, no. 2, pp. 175–192, 1995.
- [40] Z. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices," *Annals of probability*, pp. 316–345, 1998.
- [41] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*, 1st ed. New York, NY, USA: Cambridge University Press, 2011.
- [42] W. Hachem, P. Loubaton, X. Mestre, J. Najim, and P. Vallet, "Large information plus noise random matrix models and consistent subspace estimation in large sensor networks," *Random Matrices: Theory and Applications*, vol. 1, no. 02, 2012.
- [43] J. S. Turek, I. Yavneh, and M. Elad, "On MMSE and MAP denoising under sparse representation modeling over a unitary dictionary," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3526–3535, 2011.
- [44] S. M. Kay, "Fundamentals of statistical signal processing: estimation theory," *PTR Prentice-Hall, Englewood Cliffs, NJ*, 1993.
- [45] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [46] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [47] V. Buldygin and Y. Kozachenko, *Metric characterization of random variables and random processes*. American Mathematical Soc., 2000, vol. 188.
- [48] 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.
- [49] 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.
- [50] E. Chaumette, P. Larzabal, and P. Forster, "On the influence of a detection step on lower bounds for deterministic parameter estimation," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4080–4090, 2005.
- [51] Z. Ben-Haim and Y. C. Eldar, "Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1032–1047, 2011.
- [52] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "Compressed sensing block map-lms adaptive filter for sparse channel estimation and a bayesian Cramér-Rao bound," in *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on*. IEEE, 2009, pp. 1–6.
- [53] E. L. Lehmann and G. Casella, *Theory of point estimation*. Springer Science & Business Media, 1998, vol. 31.
- [54] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson/Prentice Hall Upper Saddle River, NJ, 2005.
- [55] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012, vol. 3.
- [56] P. Stoica and B. C. Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 177–179, 1998.
- [57] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 87–90, 2001.
- [58] B. M. Sadler, R. J. Kozick, and T. Moore, "Bounds on bearing and symbol estimation with side information," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 822–834, 2001.
- [59] S. Nadarajah, "A generalized normal distribution," *Journal of Applied Statistics*, vol. 32, no. 7, pp. 685–694, 2005.
- [60] S. Yu, A. Zhang, and H. Li, "A review of estimating the shape parameter of generalized Gaussian distribution," *Journal of Computational Information Systems*, vol. 8, no. 21, pp. 9055–9064, 2012.
- [61] M. Novey, T. Adali, and A. Roy, "A complex Generalized Gaussian distribution - Characterization, generation, and estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1427–1433, 2010.
- [62] T. Zhang, A. Wiesel, and M. S. Greco, "Multivariate Generalized Gaussian distribution: Convexity and graphical models," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 4141–4148, 2013.
- [63] R. Gribonval, V. Cevher, and M. E. Davies, "Compressible distributions for high-dimensional statistics," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5016–5034, 2012.
- [64] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Corporation, 1964, no. 55.
- [65] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in correlated mimo broadcast channels under limited feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.
- [66] Z. Ben-Haim, Y. C. Eldar, and M. Elad, "Coherence-based performance guarantees for estimating a sparse vector under random noise," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5030–5043, 2010.
- [67] 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.
- [68] J. A. Urugiien, Y. C. Eldar, P. Dragoty *et al.*, "Sampling at the rate of innovation: Theory and applications," *Compressed Sensing: Theory and Applications*, p. 148, 2012.
- [69] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 310–316, 2010.

- [70] 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.