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 follows 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.

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

$$y = \Psi s + n,$$

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

$$y = H \theta + n.$$  

Define $\rho_{\text{mes}} \defeq \frac{M}{N}$ 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

1Compressible means that the entries of $\theta$ 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, $H$ is treated as a known matrix in the context of the large system regime where $M, K \to \infty$ with $\rho_{\text{mes.}} \to \rho_{\text{mes.}} \in (0,1)$, $\rho_{\text{par.}} \to \rho_{\text{par.}} \in (0,1)$ and $\rho_{\text{dic.}} \to \rho_{\text{dic.}} = -O(\log \rho_{\text{par.}}) \in (1,\infty)$ 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 $\tilde{\theta}_S(y)$ be an oracle-estimator of $\theta_S$ that is derived assuming that the support $S$ 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 $\tilde{\theta}_S(y)$ given $S$:

$$\text{Var} \left( \tilde{\theta}_S \right) \overset{\text{def}}{=} \mathbb{E}_{\theta \mid H_S,S} \left[ (\theta_S - \tilde{\theta}_S(y))(\theta_S - \tilde{\theta}_S(y))^T \right]$$

BMSE$S \overset{\text{def}}{=} \text{Tr} \left[ \text{Var} \left( \tilde{\theta}_S \right) \right] = \mathbb{E}_{\theta \mid H_S,S} \left\| \theta_S - \tilde{\theta}_S(y) \right\|^2,$$

where $\mathbb{E}_{z \mid q(\cdot)}$ 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 $S$ yields the global covariance matrix and the global BMSE of $\tilde{\theta}_S(y)$:

$$\text{Var} \left( \tilde{\theta}_S \right) \overset{\text{def}}{=} \mathbb{E}_S \left[ \text{Var} \left( \tilde{\theta}_S \right) \right] = \sum_{S \in \Omega} \text{Pr}(S) \text{Var} \left( \tilde{\theta}_S \right)$$

BMSE$_\Omega \overset{\text{def}}{=} \mathbb{E}_S [\text{BMSE}_S] = \sum_{S \in \Omega} \text{Pr}(S) \text{BMSE}_S.$

The above quantities are coined global in the sense that they involve taking an expectation over the distribution of the support $S$.

A. Global Bayesian lower bounds

1) Derivation of the van Trees’ lower bound: We first define the $K \times K$ Bayesian information matrix (BIM) $B_S$ with entries

$$[B_S]_{ij} \overset{\text{def}}{=} \text{Var} \left( \frac{\partial \log p(y, \theta_S)}{\partial \theta_i} \right)_{ij}$$

$$\overset{\text{def}}{=} \mathbb{E}_{y \theta_S} \left( \frac{\partial \log p(y, \theta_S)}{\partial \theta_i} \right) \frac{\partial \log p(y, \theta_S)}{\partial \theta_j}$$

$$- \mathbb{E}_{y \theta_S} \left( \frac{\partial \log p(y, \theta_S)}{\partial \theta_i} \right) \mathbb{E}_{y \theta_S} \left( \frac{\partial \log p(y, \theta_S)}{\partial \theta_j} \right),$$

where $\theta_i$, $i \in S$ is the $i$-th entry of vector $\theta_S$. We have [20]

$$\text{Var} \left( \tilde{\theta}_S \right) \geq C_S \overset{\text{def}}{=} B_S^{-1}$$

BMSE$S \geq C_S \overset{\text{def}}{=} \text{Tr} [ C_S ]$

$$\text{Var} \left( \tilde{\theta}_S \right) \geq \mathbb{E}_S [ C_S ]$$

BMSE$_\Omega \geq C_\Omega \overset{\text{def}}{=} \mathbb{E}_S [ C_S ]$.

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

$$B_S = J_S + G_S,$$

where $J_S$ is the “data”-part and $G_S$ is the “prior”-part of the BIM.

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

$$[J_S]_{ij} = \mathbb{E}_{y \theta_S} \left( -\frac{\partial^2 \log p(y \mid H_S \theta_S)}{\partial \theta_i \partial \theta_j} \right) = \mathbb{E}_{\theta_S} [ F_S ]_{ij}$$

with the Fisher information matrix (FIM)

$$[F_S]_{ij} = \mathbb{E}_{y \theta_S} \left( -\frac{\partial^2 \log p(y \mid \theta_S)}{\partial \theta_i \partial \theta_j} \right).$$

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

$$[F_S]_{ij} = \left( \frac{\partial m}{\partial \theta_i} \right)^T \Sigma^{-1} \frac{\partial m}{\partial \theta_j} + \frac{1}{2} \text{Tr} \left[ \frac{\partial \Sigma}{\partial \theta_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \theta_j} \Sigma^{-1} \right]$$

$$= \frac{1}{\sigma^2} [H_S^T H_S]_{ij}.$$

The last expression is obtained from $\frac{\partial m}{\partial \theta_S} = [\ldots \frac{\partial m}{\partial \theta_1} \ldots] = H_S$ and $\frac{\partial \Sigma}{\partial \theta_S} = 0$ for $i \in S$.

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

$$[G_S]_{ij} = \left[ \text{Var} \left( \frac{\partial \log p(\theta_S)}{\partial \theta_S} \right) \right]_{ij},$$

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

$$B_S = \frac{H_S^T H_S}{\sigma^2} + G_S$$

$$C_S = \text{Tr}(B_S^{-1}) = \sigma^2 \text{Tr} \left( \left( H_S^T H_S + \sigma^2 G_S \right)^{-1} \right).$$

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, $H_S^T H_S$ is rank deficient so that $G_S$ is selected to have full rank to ensure invertibility of $B_S$. 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:
\[ \kappa(B_S) = \frac{\|B_S^{-1}\|_2}{\|B_S\|_2} \]
In the low noise variance regime, meaning the following result.

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:
\[ \text{BMSE}_S \geq \kappa(B_S) \|B_S\|_2 \]  
(25)

**Proof** Invoking the covariance inequality property in (14) and \( \text{Tr}(C_S) > \|C_S\|_F^4 \), we obtain \( \text{BMSE}_S > \|C_S\|_2 \). 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 \( \|B_S\|_2 \) is finite. In the low noise variance regime, meaning \( \sigma^2 \to 0 \), no estimator \( \hat{\theta}_S(y) \) with a finite BMSE_S exists.

**Proof** We can write for the condition number of the BIM
\[ \kappa(B_S) = \kappa \left( \frac{H_S^T H_S}{\sigma^2} + G_S \right) = \kappa \left( H_S^T H_S + \sigma^2 G_S \right). \]  
(26)
Using this expression, we have
\[ \kappa(B_S) \xrightarrow{\sigma^2 \to 0} \kappa(H_S^T H_S) = \infty \]  
(27)
because matrix \( H_S^T H_S \) is singular when \( K > M \). Now using Lemma 3.1 yields \( \text{BMSE}_S \xrightarrow{\sigma^2 \to 0} \infty \) provided the 2-norm of \( B_S \) is finite.

As we can see, computing the inverse of \( B_S \) 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 \( K > M \), then \( H_S^T H_S \) is a \( M \)-rank matrix and its SVD is \( H_S^T H_S = \sum_{i=1}^{K} d_i u_i u_i^T \) where \( u_i \) and \( d_i \) are the \( i \)-th singular vector and value, respectively. Due to the singularity of matrix \( H_S^T H_S \), 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 \( G_S = cI_K \) for some positive \( c \). Then the BCRB is given by
\[ C_S = \text{Tr} \left( \frac{H_S^T H_S}{\sigma^2} + cI_K \right)^{-1} = \text{Tr} \left( W^{-1} \right), \]  
(28)
\(^{\text{Notice that } B_S \text{ is assumed to be non-singular. Thus, } C_S = B_S^{-1} \text{ exists and is full-rank. In this case, the inequality is strict.}}\)

where \( W \) is a diagonal matrix with diagonal entries \( W_{ii} = \frac{\sigma^2}{\sigma^2} + c \) for \( 1 \leq i \leq M \) and \( W_{ii} = c \) for \( M+1 \leq i \leq K \). Thus, the conditional BMSE is lower bounded by the “regularized” BCRB according to
\[ \text{BMSE}_S \geq \sum_{i=1}^{M} \frac{\sigma^2}{d_i} + \frac{K - M}{c} = \frac{K - M}{c}. \]  
(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 \( C_S \) and its low and high noise variance approximations depend on the deterministic matrix \( G_S \) which is determined by the prior on \( \theta_S, \) i.e. \( p(\theta_S) \). In this section, we specialize this prior to be a centered generalized normal distribution and compute explicit expressions for \( C_S \) for it.

Specifically, we assume that \( \theta_S \) follows a centralized generalized normal distribution [59]–[62], i.e. \( \theta_S \sim \mathcal{G}N(0, D_S, \beta) \) where \( \beta \) is the real positive shape parameter and \( D_S \) is the dispersion matrix of the distribution. We further assume that the entries of \( \theta_S \) are independent, i.e. \( D_S \) is diagonal and of the form \( D_S = \text{diag}\{(\alpha_i^2)_{i \in S}\} \), where \( \alpha_i > 0 \) is the scale parameter of entry \( \theta_i, \: i \in S \). The pdf of \( \theta_S \) reads
\[ p(\theta_S) \propto e^{-\frac{\beta}{2} D_S^{-1} \theta_S^T \theta_S} \]  
(30)
where \( \propto \) means “proportional to”. The scale parameters can be written as \( \alpha_i^2 = \frac{\sigma^2_i}{\text{SDR}_i} \) where \( \text{SDR}_\beta = \frac{\Gamma(\beta)}{\Gamma(1/\beta)} \) is the Signal to Distortion Ratio (SDR) [63] with \( \Gamma(\cdot) \) the Gamma function [64] and \( \sigma^2_i \) is the variance of \( \theta_i, \: i \in S \).

The generalized normal distribution encompasses the Laplacian, Gaussian, uniform pdfs and degenerate Dirac distributions as special cases for \( \beta = 1, \: \: \beta = 2, \: \: \beta \to \infty \) and \( \beta \to 0 \), respectively, see Fig. 1. We denote in the following \( R_{\theta_S} = \text{diag}\{\sigma^2_i\}_{i \in S} \). Notice that \( R_{\theta_S} = D_S \) 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 \( \beta \).](image-url)
We now provide the explicit expression for $G_S$ for this setting. Since the entries of $\theta_S$ are independent, we have

$$[G_S]_{ij} = \var\left(\frac{\partial \log p(\theta_S)}{\partial \theta_S}\right)_{ij}$$

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

(31)

(32)

Moreover from (30), $p(\theta_i) \propto e^{-|\theta_i|^2}$. The score function of $\theta_i$ is thus given by

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

(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

$$\var\left(\frac{\partial \log p(\theta_i)}{\partial \theta_i}\right) = \mathbb{E}\left(\left(\frac{\partial \log p(\theta_i)}{\partial \theta_i}\right)^2\right)$$

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

(34)

(35)

Assuming that $2(\beta - 1)$ is integer-valued, we have

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

(60)

This finally entails

$$G_S = g_\beta \text{SDR}_\beta R^{-1}_{\theta_S},$$

(36)

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

$$C_S = \sigma^2 \text{Tr}\left((H_S^T H_S + \sigma^2 g_\beta \text{SDR}_\beta R^{-1}_{\theta_S})^{-1}\right).$$

(37)

B. Asymptotic analysis

The above expressions for $C_S$ still depend on the non-stochastic, known and randomly generated matrix $H_S$. 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 $M, K \to \infty$ such that $\rho_{\text{dic.}} = M/K \to \rho_{\text{dic.}} \in (0, 1)$.

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 $H_S^2 H_S$. More specifically, if matrix $H_S^2 H_S$ 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 $H_S^2 H_S$. We can show that

$$\kappa(H_S^2 H_S) = \frac{\lambda_{\max}(H_S^2 H_S)}{\lambda_{\min}(H_S^2 H_S)} \to \left(1 + \frac{1}{\sqrt{1/\rho_{\text{dic.}}} - 1}\right)^2$$

(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 $\rho_{\text{dic.}}$ 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 $H_S^2 H_S$ 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 $M \to \infty$ with $\rho_{\text{dic.}} \to \rho_{\text{dic.}}$,

$$0 < \lim \inf \lambda_1(\mathbf{R}_\theta) \leq \lim \sup \lambda_K(\mathbf{R}_\theta) < \infty.$$  

(39)

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

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

$$\frac{1}{K} C_S - \sigma^2 e_M \to 0$$

(40)

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

$$e = \frac{1}{K} \text{Tr}\left(\frac{1}{1 + e\rho_{\text{dic.}}} \mathbf{I}_K + \sigma^2 g_\beta \text{SDR}_\beta R^{-1}_{\theta_S})\right).$$

(41)

Proposition 3.3 states that, as $M \to \infty$ with $\rho_{\text{dic.}} \to \rho_{\text{dic.}}$, the random variable $\frac{1}{K} C_S$ 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

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

(42)

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

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

(43)

In particular,

$$\frac{1}{K} C_S \to C(\infty) \triangleq \sigma^2 e(\infty)$$

(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 $M \to \infty$ with $\rho_{\text{dic.}} \to \rho_{\text{dic.}}$, and $\sigma \to 0$ we have

$$\frac{1}{\sigma^2 K} C_S \to \frac{\rho_{\text{dic.}}}{\rho_{\text{dic.}} - 1} \to 0$$

(45)

almost surely.

Proof To prove the above proposition recall from [65] that the extension of $e_M$ to the mapping $C \to \mathbb{C} \setminus \text{supp}(\mu_M)$, $z \mapsto e_M(z)$, unique solution of (41) with $\sigma^2 = -z$, is the Stieltjes transform of the measure with support $\text{supp}(\mu_M)$. Since $\rho_{\text{dic.}} > 1$ 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, $e_M \to e_M^*$ as $\sigma \to 0$ with $e_M^*$ the unique positive solution
to $e^*_M = 1 + e^*_M \hat{\rho}_{\text{dic}}^{-1}$, which here is explicitly given by $e^*_M = \frac{\hat{\rho}_{\text{dic}}}{\rho_{\text{dic}} - \sigma^2}$. Now, from [40], $\lim_{\sigma \to 0} \lambda_{\min}(H_S^T H_S) > 0$ almost surely, so that $\sigma^{-2} \frac{1}{K} C_S$ 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 $\rho_{\text{dic}}$ only. That is, the actual support set $S$ 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

$$C(\infty) \approx \sigma^2 \frac{\rho_{\text{dic}}}{\rho_{\text{dic}} - 1} = \frac{\sigma^2}{1 - \frac{\rho_{\text{max}}}{\rho_{\text{min}}}}.$$

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

$$\frac{1}{K} C_S - g_\beta \text{SDR}_\beta \frac{1}{K} \text{Tr}(R_{\theta_S}) \to 0$$

as $\sigma \to \infty$.

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 $R_{\theta_S} = \sigma^2 \text{SNR}_K$ (identical dispersion regime) with

$$\text{SNR} \overset{\text{def}}{=} \frac{\sigma^2}{\sigma^2} = \frac{\Gamma(3/\beta)}{\Gamma(1/\beta)} \alpha^2 \beta^2$$

**Proposition 3.7** particularizes as follows:

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

$$\frac{1}{K} C_S \to C(\infty) \overset{\text{def}}{=} \frac{\text{SNR}}{g_\beta \text{SDR}_\beta}$$

almost surely, where

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

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

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

$$P(e; x) = e^2 + U(x) e - V(x),$$

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., $\hat{\rho}_{\text{dic}} \gg 1$, we have

$$C(\infty) = \frac{\sigma^2}{\text{SNR} + g_\beta \text{SDR}_\beta} + O(\hat{\rho}_{\text{dic}}^2).$$

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

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

**Proposition 3.9:** As $M \to \infty$ with $\rho_{\text{dic}} \to \hat{\rho}_{\text{dic}}$ and $\sigma \to 0$

$$\frac{1}{K} C_\Omega - C(\infty) \to 0.$$

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

$$C_\Omega = \sum_{S \in \Omega} \text{Pr}(S|L) C_S \to C(\infty) \lim_{M,K \to \infty} \sum_{S \in \Omega} \text{Pr}(S|L) = C(\infty).$$

The above proposition is important from a computational point of view since the brute force computation of $C_\Omega$, 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 $\hat{\theta}_S(y)$.

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

$$\text{Var}\left(\hat{\theta}_S\right) = \text{Tr}\left(\frac{1}{\sigma^2} H_S^T H_S + R_{\theta_S}^{-1}\right)^{-1}.$$  \hspace{1cm} (55)

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

$$\text{Var}\left(\hat{\theta}_S\right) = \sigma^2 \text{Tr}\left(\frac{1}{\sigma^2} H_S^T H_S + I_K\right)^{-1}.$$  \hspace{1cm} (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 $M \to \infty$ with $\rho_{\text{dic.}} \to \rho_{\text{dic.}}$,
\[ \frac{1}{K} \text{Var} \left( \hat{\theta}_S \right) \to \text{Var}^{(\infty)} \left( \hat{\theta} \right) \overset{\text{def.}}{=} r(\text{SNR}) \] almost surely where $r(x)$ is defined in (50).

Remark 4.2: The variance of the oracle LMMSE estimator is given by
\[ \text{Var}_{\Omega}(\hat{\theta}) = \sum_{S \in \Omega} \text{Pr}(S) \text{Var} \left( \hat{\theta}_S \right). \] (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 $\text{Var}^{(\infty)} \left( \hat{\theta} \right)$.

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., $\rho_{\text{dic.}} \gg 1$, we have
\[ \text{Var}^{(\infty)} \left( \hat{\theta} \right) = \frac{\sigma_0^2}{\text{SNR} + 1} + O(\rho_{\text{dic.}}^2). \] (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
\[ C_\beta = \frac{\text{C}^{(\infty)}}{\text{Var}^{(\infty)} \left( \hat{\theta} \right)} = \frac{r \left( \frac{\text{SNR}}{g_2 \text{SDR}_\beta} \right)}{r(\text{SNR})}. \] (60)
we obtain the next proposition.

Proposition 4.4:
\[ \mathcal{P}_1. \text{ For a sufficiently large SNR ratio } C_\beta \text{ is given by } \]
\[ C_\beta = \begin{cases} 1, & \text{for } \beta = 2, \\ < 1, & \text{for } \beta \neq 2. \end{cases} \] (61)

\[ \mathcal{P}_2. \text{ At small SNR ratio } C_\beta \text{ is given by } C_\beta = \frac{1}{g_2 \text{SDR}_\beta} + O(\text{SNR}^2). \] (62)

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

Second, it is straightforward to see that $r(x)$ is an increasing function and since $\frac{\text{SNR}}{g_2 \text{SDR}_\beta} < \text{SNR}$ for all $\beta \neq 2$ we conclude that $C_\beta < 1$ and therefore that $C^{(\infty)} < \text{Var}^{(\infty)} \left( \hat{\theta} \right)$.

$\mathcal{P}_2$. Assuming that $x$ is small, we have
\[ f(x) = \frac{4\rho_{\text{dic.}} x}{(\rho_{\text{dic.}} + (\rho_{\text{dic.}} - 1)x)^2} = \frac{4}{\rho_{\text{dic.}}^2} x + O(x^2) \] (63)
\[ g(x) = \sqrt{1 + f(x)} - 1 = \frac{1}{2} f(x) + O(f(x)^2). \] (64)

Using these approximations and (50), we obtain an approximated root for small $x$ according to
\[ r(x) = x^2 + O(x^2). \] (65)

Thus, at small SNR we have
\[ C^{(\infty)} = \frac{\sigma_0^2}{g_2 \text{SDR}_\beta} + O(\text{SNR}^2). \] (66)
\[ \text{Var}^{(\infty)} \left( \hat{\theta} \right) = \sigma_0^2 + O(\text{SNR}^2). \] (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 $C^{(\infty)}$. The proportionality coefficient $C_\beta$ expressed in dB reads
\[ C_\beta [\text{dB}] = -10 \log_{10}(g_2) - 10 \log_{10}(\text{SDR}_\beta). \] (68)

It is tabulated in Table I for selected values of $\beta$. 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, $\text{Var}^{(\infty)} \left( \hat{\theta} \right)$, and bound $C^{(\infty)}$ have the same asymptotic behavior wrt. the SNR. Specifically, these quantities are almost identical for $\beta = 2$ 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:
\[ x(t) = \sum_{\ell \in S} \theta_\ell \delta(t - \tau_\ell), \] (69)
where $\tau_\ell$ and $\theta_\ell$ are respectively the time-delay and the amplitude of the $\ell$-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 $g(t) = \frac{1}{T_S} \text{sinc} \left( \frac{k}{T_S} \right)$ where $1/T_S$ is the sampling rate. Then, uniform sampling at rate $1/T_S$ of signal $x(t)$ yields the samples
\[ s_k = \int_{-\infty}^{\infty} g(t - kT_S) x(t) dt \] (70)
\[ = \frac{1}{T_S} \sum_{\ell \in S} \theta_\ell \text{sinc} \left( \frac{\tau_\ell - k}{T_S} \right), \quad k \in [1 : N]. \] (71)
Define the orthogonal basis matrix (see Fig. 2-(a)) as
\[ [\Phi]_{kk'} = \frac{1}{T_S} \text{sinc} \left( \frac{\tau_{kk'}}{T_S} - 1 \right), \quad k, k' \in [1 : N]. \] (72)
and assume that vector \( \theta \) is \( K \)-sparse. We then have
\[ s = \Phi \theta = \Phi_S \theta_S, \] (73)
where \( S \) is the support set. Vector \( s \) 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 \( 1/T_S \to 2 \bar{\rho}_{\text{spar}} \).

Define a \( M \times N \) non-stochastic but randomly generated measurement matrix \( \Psi \) with \( M < N \), see also Fig. 2-(b). The under-sampled observation vector with ratio \( \rho_{\text{mes}} \) is given by
\[ y = \Psi s + n = \Psi \Phi S \theta_S + n, \] (74)
where \( s \) 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 \( \theta_S \) with location vector \( 0 \) and scale matrix \( \frac{1}{2} \mathbf{R}_{\theta_S} \), i.e., \( \theta_S \sim \mathcal{G} \mathcal{N} \left( 0, \frac{1}{2} \mathbf{R}_{\theta_S} \right) \).

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.

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>1 (Lap.)</th>
<th>2 (Gauss.)</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_B ) [dB]</td>
<td>-3</td>
<td>0</td>
<td>-0.5</td>
<td>-1.4</td>
<td>-2.2</td>
<td>-2.9</td>
<td>-3.5</td>
<td>-4</td>
<td>-4.6</td>
<td>-5</td>
</tr>
</tbody>
</table>

Table I: Ratio \( C_B \) in dB

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:
  \[ \hat{\theta}_S = \min_{\theta} ||\theta||^2_1 \text{ subject to } ||y - \mathbf{H}\theta||^2 \leq \sigma^2. \] (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 $\sigma^2 \to \infty$. So, in the low SNR regime the BCRB provides a good prediction of the performance of algorithms based on criterion (75).

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 (except 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


