ON SATURATION OF THE CRAMÉR RAO BOUND FOR SPARSE BAYESIAN LEARNING

Ali Koochakzadeh and Piya Pal

Dept. of Electrical and Computer Engineering,
University of California, San Diego, USA
E-mail: alik@eng.ucsd.edu, pipal@eng.ucsd.edu

ABSTRACT

This paper analyzes the Cramér-Rao Bound associated with the estimation of certain sparse hyper-parameters in the Sparse Bayesian Learning (SBL) framework, that crucially control the sparsity of the desired signal. The CRB is shown to exhibit saturation with respect to the number of measurements, i.e., it can be lower bounded by a non-negative quantity that does not go to zero even when the number of measurements tends to infinity. Moreover, the CRB corresponding to the nonzero and zero elements of the sparse hyper-parameter can exhibit different behaviors. While the CRB for the non-zero elements always *saturate* regardless of the type of dictionary, saturation of the CRB for zero elements provably happens when the dictionary has normalized columns. For an unnormalized dictionary, singular values of certain sub-dictionaries determine if saturation can happen, prompting future research into this interesting phenomenon. [1]

Index Terms— Sparse Bayesian Learning, Cramér-Rao Bounds, Compressed Sensing, hyper-parameter estimation, mean squared error.

1. INTRODUCTION

Sparse Bayesian Learning (SBL) [1, 2, 3, 4] constitutes an important family of Bayesian algorithms where the goal is to estimate a sparse signal $\mathbf{x} \in \mathbb{F}^N$, from compressed measurement $\mathbf{y} \in \mathbb{F}^M$ acquired as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} \tag{1}$$

Here $\mathbf{A} \in \mathbb{F}^{M\times N}$ ($M < N$) denotes an underdetermined dictionary and $\mathbf{w} \in \mathbb{F}^{M\times 1}$ denotes the additive noise. Throughout this paper, $\mathbb{F}$ can be either the set of real ($\mathbb{R}$) or complex ($\mathbb{C}$) numbers. Unlike traditional Compressed Sensing algorithms [5, 6] that only exploit the sparsity of $\mathbf{x}$ to solve the ill-posed problem (1), SBL algorithms impose a suitable prior distribution on $\mathbf{x}$ (that also models its sparsity) and computes the corresponding posterior estimate. Alongside recovering $\mathbf{x}$, SBL algorithms also allow estimation of certain hyper-parameters characterizing the prior distribution of $\mathbf{x}$ that crucially control its sparsity as well as correlation structure [7].

The authors in [8] investigated fundamental performance limits of the SBL framework by deriving appropriate Cramér Rao Bounds (CRB) on the mean-squared error (MSE) of SBL estimators for $\mathbf{x}$ and associated hyper-parameters. However, the analytical behavior of these bounds as a function of number of measurements M, have not been investigated so far. Of particular interest would be to understand if increasing the number of measurements M enables us to estimate the hyper-parameters with proportionately decreasing MSE that converge to 0? As we will show in this paper, the answer is a surprising "no", implying that even when M goes to infinity, the CRB does not decrease below a certain positive quantity, and it essentially *saturates*. Hence, even with infinite measurements, no *unbiased* estimator exists that can *exactly* recover the hyper-parameter. We mathematically characterize this *saturation* behavior of the CRB corresponding to both zero and nonzero elements of the sparse hyper-parameter.

Related Work. Cramér-Rao Bounds for estimating sparse signals in presence of noise have been derived in [9, 10]. However, these results do not consider a stochastic model (or prior) for $\mathbf{x}$ and hence cannot be applied for analyzing SBL. In [8], for the first time, CRB expressions for the SBL framework were derived, assuming different statistical models. In [11], the authors proved that the sparse hyper-parameter can be identifiable even when the number of non-zero elements of $\mathbf{x}$ exceed M, and derived corresponding CRB expressions. This paper conducts further analysis of the CRB for sparse hyper-parameters, and mathematically justifies its saturation behavior (which manifests differently for the non-zero and zero elements).

Notations. Throughout this paper, $(.)^T, (.)^H, (.)^*$ represent matrix transpose, Hermitian, and complex conjugate operators, respectively. Furthermore, $\circ, \odot, \otimes$ denote Hadamard, Khatri-Rao and Kronecker products, respectively.

2. STATISTICAL MODEL FOR SBL

Assume that the signal $\mathbf{x}$ is a random vector distributed as $\mathbf{x} \sim \mathbb{F}\mathcal{N}(\mathbf{0}, \mathbf{P})$ where $\mathbf{P} = \mathrm{diag}(p_1, p_2, \cdots, p_N)$, and we denote $\mathbf{p} = [p_1, \cdots, p_N]^T$ as the vector of hyper-parameters, representing the power of the elements of $\mathbf{x}$. The vector $\mathbf{p}$ is assumed to be sparse where $\mathcal{S}$ denotes the support of $\mathbf{p}$, with $|\mathcal{S}| = K$, i.e., $\mathbf{p}$ contains only K nonzero numbers. Furthermore, let $\mathbf{w}$ represent white Gaussian noise with distribution $\mathbb{F}\mathcal{N}(\mathbf{0}, \sigma_w^2\mathbf{I})$, which is uncorrelated with $\mathbf{x}$. We henceforth assume that the noise variance σ_w^2 is also known. Under these assumptions, one can write the probability density function (pdf) of $\mathbf{y}$ as

$$p_{\mathbf{Y};\mathbf{p}}(\mathbf{y};\mathbf{P}) = [(2\pi)^M(\det \mathbf{R})]^{-\eta_{\mathbb{F}}} e^{-\eta_{\mathbb{F}}\mathbf{y}^H\mathbf{R}^{-1}\mathbf{y}} \tag{2}$$

[1]This work was supported in parts by the NSF CAREER award (ECCS 1553954), and the University of California, San Diego.

where $\mathbf{R}$ is the covariance matrix of the random variable $\mathbf{y}$, given by

$$\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma_w^2\mathbf{I},$$

and $\eta_{\mathbb{F}} = 1$ for $\mathbb{F} = \mathbb{C}$, and $\eta_{\mathbb{F}} = \frac{1}{2}$ for $\mathbb{F} = \mathbb{R}$.

In SBL, both the signal $\mathbf{x}$, and the hyper-parameter $\mathbf{p}$ can be recovered by respectively solving the so-called Type I, and Type II estimation problems. In this paper, we are primarily interested in estimating the underlying hyper-parameter $\mathbf{p}$ that characterize the signal distribution (as well as its sparsity).

2.1. Review of Cramér Rao Bounds For Hyperparameter Estimation

It is well known that the Cramér Rao Bound (CRB) serves as a fundamental lower bound on the Mean Squared Error (MSE) of *any* unbiased estimator for a (deterministic) parameter. In [8], various CRB expressions (such as Hybrid, Bayesian and Marginalized CRBs) are derived under different statistical assumptions and models. Since our goal in this paper is to analyze the CRB for the hyperparameter $\mathbf{p}$, we consider the marginalized CRB (MCRB) for $\mathbf{p}$ derived in [8]. The authors in [8] also show that among all CRB expressions, the MCRB provides the tightest lower bound.

The MCRB for $\mathbf{p}$ can be derived using the marginalized distribution of $\mathbf{y}$ given by (2) and assuming that the noise power σ_w^2 is known. In this case, $\mathbf{p}$ is the only unknown parameter characterizing $p_{\mathbf{Y};\mathbf{P}}(\mathbf{y};\mathbf{P})$, and the corresponding Fisher Information Matrix (FIM) $\mathbf{J}$ can be shown to be of the form [11]

$$\mathbf{J} = \eta_{\mathbb{F}}\mathbf{A}_{\text{ca}}^H\mathbf{W}\mathbf{A}_{\text{ca}} \tag{3}$$

where $\mathbf{A}_{\text{ca}} = \mathbf{A}^* \odot \mathbf{A}$, $\mathbf{W} = \mathbf{R}^{-T} \otimes \mathbf{R}^{-1}$. The matrix $\mathbf{A}_{\text{ca}}$ denotes the Khatri-Rao product (or column-wise Kronecker product) of the dictionary $\mathbf{A}$ and crucially controls important properties of $\mathbf{J}$. By considering the rank of $\mathbf{A}_{\text{ca}}$, the authors in [11] have been able to provide the following necessary and sufficient condition under which the MCRB for $\mathbf{p}$ exists:

Theorem 1. *[11] The FIM* $\mathbf{J}$ *given in* (3) *is non-singular if and only if* $N = \text{rank}(\mathbf{A}_{\text{ca}})$.

Hence, as long as $N = \text{rank}(\mathbf{A}_{\text{ca}})$ (which can imply $N = O(M^2)$ for certain dictionaries), the CRB exists and can be used to lower bound the MSE of any unbiased estimate of $\mathbf{p}$.

3. SATURATION OF THE MCRB

For many overdetermined estimation problems ($N \le M$), the CRB typically converges to 0 asymptotically as the number of measurements $M \to \infty$, implying that the parameter can be estimated with zero MSE (as $M \to \infty$) using appropriate estimators (such as Maximum Likelihood Estimator). However, we will now show that the MCRB for SBL (that typically involves a compressive measurement model with $N > M$) can saturate at a value strictly bounded away from zero, even when $M \to \infty$. This behavior implies that it is not possible to find an unbiased estimator that can recover $\mathbf{p}$ with zero MSE as the number of measurements grows infinitely large. In this regard, we will show that the *non-zero and zero* elements of $\mathbf{p}$ exhibit different saturation behavior as follows:

(i) CRB of Non-Zero Elements: For all values of N and M and all choices of the dictionary $\mathbf{A}$, the CRB corresponding to the nonzero elements of $\mathbf{p}$ always exhibit a saturation effect, meaning that we can find a lowerbound for the CRB (in terms of M) that tends to a strictly positive limit as $M \to \infty$.

(ii) CRB of Zero Elements: The CRB corresponding to the zero elements of $\mathbf{p}$ can be lower bounded by a *non zero quantity* (even when $M \to \infty$) as long as the columns of the dictionary $\mathbf{A}$ are normalized. If the columns of $\mathbf{A}$ are not normalized, saturation of the CRB is shown to be determined by the singular values of certain submatrices of $\mathbf{A}$.

3.1. Saturation of the CRB Corresponding to Nonzero Elements

Let $\mathbf{C} = \mathbf{J}^{-1}$, where the ith diagonal element of $\mathbf{C}$ provides a lower bound on the MSE of any unbiased estimator for $[\mathbf{p}]_i$, i.e., given any unbiased estimate $\hat{\mathbf{p}}(\mathbf{y})$ (which is a function of only the measurement $\mathbf{y}$) of $\mathbf{p}$, we have

$$\mathbb{E}_{\mathbf{y}}|[\mathbf{p}]_i - [\hat{\mathbf{p}}(\mathbf{y})]_i|^2 \ge [\mathbf{C}]_{ii}$$

The following theorem shows that if $i \in \mathcal{S}$ (i.e., $[\mathbf{p}]_i > 0$), then $[\mathbf{C}]_{ii}$ is strictly bounded away from zero.

Theorem 2. *Consider the model (1), where the measurement* $\mathbf{y}$ *is distributed according to (2). If* $N = \text{rank}(\mathbf{A}_{\text{ca}})$*, the CRB corresponding to the unknown parameter* $\mathbf{p}$ *satisfies*

$$[\mathbf{C}]_{ii} \ge \eta_{\mathbb{F}}^{-1} p_i^2, \quad i \in \mathcal{S}$$

Proof. Since $N = \text{rank}(\mathbf{A}_{\text{ca}})$, $\mathbf{J}$ is invertible and the CRB exists. Following the analysis in [12] and (3), it can be shown that ith diagonal element of the $\mathbf{C}$ can be written as

$$[\mathbf{C}]_{ii}^{-1} = \eta_{\mathbb{F}}\|\Pi^{\perp}_{\mathbf{W}^{1/2}\mathbf{A}_{\text{ca}}^{(-i)}}\mathbf{W}^{1/2}\mathbf{a}_{\text{ca}}^{(i)}\|^2$$

where $\mathbf{a}_{\text{ca}}^{(i)} = \mathbf{a}_i^* \otimes \mathbf{a}_i$,

$$\mathbf{A}_{\text{ca}}^{(-i)} = [\mathbf{a}_{\text{ca}}^{(1)}, \mathbf{a}_{\text{ca}}^{(2)}, \cdots, \mathbf{a}_{\text{ca}}^{(i-1)}, \mathbf{a}_{\text{ca}}^{(i+1)}, \cdots, \mathbf{a}_{\text{ca}}^{(N)}].$$

In other words, $\mathbf{A}_{\text{ca}}^{(-i)}$ contains a total of $N-1$ columns that excludes the column $\mathbf{a}_{\text{ca}}^{(i)}$. Furthermore, given any matrix $\mathbf{B}$ with full column rank, $\Pi_{\mathbf{B}}^{\perp} = \mathbf{I} - \mathbf{B}(\mathbf{B}^H\mathbf{B})^{-1}\mathbf{B}^H$ denotes the projection onto the orthogonal complement of range space of $\mathbf{B}$. Therefore, we can write

$$\begin{aligned}\eta_{\mathbb{F}}^{-1}[\mathbf{C}]_{ii}^{-1} &= (\mathbf{a}_{\text{ca}}^{(i)})^H\Big(\mathbf{W} - \\ &\quad \mathbf{W}\mathbf{A}_{\text{ca}}^{(-i)}((\mathbf{A}_{\text{ca}}^{(-i)})^H\mathbf{W}\mathbf{A}_{\text{ca}}^{(-i)})^{-1}(\mathbf{A}_{\text{ca}}^{(-i)})^H\mathbf{W}\Big)\mathbf{a}_{\text{ca}}^{(i)} \\ &\le (\mathbf{a}_{\text{ca}}^{(i)})^H\mathbf{W}\mathbf{a}_{\text{ca}}^{(i)} = |\mathbf{a}_i^H\mathbf{R}^{-1}\mathbf{a}_i|^2. \end{aligned} \tag{4}$$

where the last equality can be verified using algebraic properties of the Kronecker product. Since $i \in \mathcal{S}$, we can decompose $\mathbf{R}$ as

$$\mathbf{R} = \tilde{\mathbf{A}}_i\tilde{\mathbf{P}}_i\tilde{\mathbf{A}}_i^H + p_i\mathbf{a}_i\mathbf{a}_i^H + \sigma_w^2\mathbf{I}$$

where $\tilde{\mathbf{A}}_i$ is comprised of columns of $\mathbf{A}$ indexed by $\mathcal{S} \setminus i$, and $\tilde{\mathbf{P}}_i$ is a diagonal matrix composed of the corresponding elements of $\mathbf{p}$. Let us denote $\mathbf{R}_i := p_i \mathbf{a}_i \mathbf{a}_i^H + \sigma_w^2 \mathbf{I}$. Using Woodbury's matrix identity [13], we have

$$\mathbf{R}^{-1} = \mathbf{R}_i^{-1} - \mathbf{R}_i^{-1} \tilde{\mathbf{A}}_i (\tilde{\mathbf{P}}_i^{-1} + \tilde{\mathbf{A}}_i^H \mathbf{R}_i^{-1} \tilde{\mathbf{A}}_i)^{-1} \tilde{\mathbf{A}}_i^H \mathbf{R}_i^{-1}. \tag{5}$$

We can further use the Sherman-Morison [14] formula to get

$$\mathbf{R}_i{}^{-1} = \sigma_w^{-2} \mathbf{I} - \frac{\sigma_w^{-4} p_i \mathbf{a}_i \mathbf{a}_i^H}{1 + \sigma_w^{-2} \|\mathbf{a}_i\|^2 p_i}$$

Therefore,

$$\mathbf{a}_i^H \mathbf{R}_i{}^{-1} \mathbf{a}_i = \frac{\|\mathbf{a}_i\|^2}{\sigma_w^2} - \frac{\sigma_w^{-4} \|\mathbf{a}_i\|^4 p_i}{1 + \sigma_w^{-2} \|\mathbf{a}_i\|^2 p_i} = \frac{\sigma_w^{-2} \|\mathbf{a}_i\|^2}{1 + \sigma_w^{-2} \|\mathbf{a}_i\|^2 p_i}$$

Using (4), (5), and the fact that

$$\mathbf{a}_i^H \mathbf{R}_i^{-1} \tilde{\mathbf{A}}_i (\tilde{\mathbf{P}}_i^{-1} + \tilde{\mathbf{A}}_i^H \mathbf{R}_i^{-1} \tilde{\mathbf{A}}_i)^{-1} \tilde{\mathbf{A}}_i^H \mathbf{R}_i^{-1} \mathbf{a}_i \geq 0,$$

we conclude that

$$\begin{aligned} [\mathbf{C}]_{ii} &\geq \frac{\eta_{\mathbb{F}}^{-1}}{|\mathbf{a}_i^H \mathbf{R}^{-1} \mathbf{a}_i|^2} \\ &\geq \frac{\eta_{\mathbb{F}}^{-1}}{|\mathbf{a}_i^H \mathbf{R}_i^{-1} \mathbf{a}_i|^2} = \eta_{\mathbb{F}}^{-1} \left(\frac{\sigma_w^2 + \|\mathbf{a}_i\|^2 p_i}{\|\mathbf{a}_i\|^2} \right)^2 \quad (6) \\ &\geq \eta_{\mathbb{F}}^{-1} p_i^2 \quad (B_{nz}) \end{aligned}$$

where the label B_{nz} stands for the bound for nonzero entries. The bounds for zero entries (B_{z1}, B_{z2}) will be studied later. □

Remark 1. *The theorem indicates that for all admissible values of M, N and sparsity K, the CRB corresponding to the non zero elements of $\mathbf{p}$ is strictly greater than 0, regardless of the structure of the dictionary. This happens in both overdetermined ($N \leq M$) and underdetermined settings ($N > M$), implying that the non zero elements of $\mathbf{p}$ cannot be estimated with zero MSE even when $M \to \infty$.*

Remark 2. *For the special case when $\mathbf{A}^H \mathbf{A} = M\mathbf{I}$ (which holds only if $N \leq M$), the authors in [8] show that the inequality in (6) holds with equality. Our result generalizes this observation for any dictionary $\mathbf{A}$ and for all values of M and N.*

3.2. Lower Bounds on the CRB Corresponding to Zero Elements

We will now show that for $i \notin \mathcal{S}$ (i.e. $[\mathbf{p}]_i = 0$), saturation of the CRB may or may not happen, depending on the structure of the dictionary $\mathbf{A}$ and the normalization of its columns.

3.2.1. Saturation Effect for Normalized Dictionaries

Let $\mathbf{A}$ be a dictionary with normalized columns such that

$$\|\mathbf{a}_i\|_2^2 = c, \quad 1 \leq i \leq N$$

where c is a constant that does not depend on M or N. In this case, the CRB corresponding to the zero elements of $\mathbf{p}$ will saturate, as given by the following theorem:

Theorem 3. *Consider the model (1), where the measurement $\mathbf{y}$ is distributed according to (2), and the columns of $\mathbf{A}$ are normalized such that $\|\mathbf{a}_i\|_2^2 = c, 1 \leq i \leq N$ where c is a universal constant that does not depend on M or N. If $N = rank(\mathbf{A}_{\text{ca}})$, the CRB corresponding to the unknown parameter $\mathbf{p}$ satisfies*

$$[\mathbf{C}]_{ii} \geq \frac{\sigma_w^4 \eta_{\mathbb{F}}^{-1}}{c^2} \quad i \notin \mathcal{S}$$

Proof. Similar to the proof of Theorem 2, we use Woodbury's matrix identity on $\mathbf{R}^{-1}$, but in a different form. In particular, we can write

$$\mathbf{R}^{-1} = \sigma_w^{-2} (\mathbf{I} - \tilde{\mathbf{A}} (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \sigma_w^2 \tilde{\mathbf{P}}^{-1})^{-1} \tilde{\mathbf{A}}^H)$$

where $\tilde{\mathbf{A}}$ is the matrix comprised of columns of $\mathbf{A}$ indexed by $\mathcal{S}$, and $\tilde{\mathbf{P}}$ is a diagonal matrix containing only the non zero elements of $\mathbf{p}$. Since $i \notin \mathcal{S}$, we have

$$\begin{aligned} \mathbf{a}_i^H \mathbf{R}_i{}^{-1} \mathbf{a}_i &= \sigma_w^{-2} \mathbf{a}_i^H (\mathbf{I} - \tilde{\mathbf{A}} (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \sigma_w^2 \mathbf{P}^{-1})^{-1} \tilde{\mathbf{A}}^H) \mathbf{a}_i \\ &\leq \sigma_w^{-2} \|\mathbf{a}_i\|^2 \end{aligned}$$

which follows from the fact that

$$\mathbf{a}_i^H \tilde{\mathbf{A}} (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}} + \sigma_w^2 \tilde{\mathbf{P}}^{-1})^{-1} \tilde{\mathbf{A}}^H \mathbf{a}_i \geq 0.$$

Using (4), for $i \notin \mathcal{S}$, we can always write

$$[\mathbf{C}]_{ii} \geq \frac{\sigma_w^4 \eta_{\mathbb{F}}^{-1}}{\|\mathbf{a}_i\|^4} = \frac{\sigma_w^4 \eta_{\mathbb{F}}^{-1}}{c^2} \tag{B_{z1}}$$

□

Notice that the first inequality in (B_{z1}) provides a valid lower bound for any dictionary $\mathbf{A}$ (regardless of normalization of columns). However, for unnormalized dictionaries, if $\|\mathbf{a}_i\|^2$ grows monotonically with M, the lower bound $\frac{\sigma_w^4 \eta_{\mathbb{F}}^{-1}}{\|\mathbf{a}_i\|^4}$ in (B_{z1}) converges to a trivial value of 0 (as $M \to \infty$) which does not shed any light into the asymptotic behavior of the CRB.

3.2.2. Lower Bound for Unnormalized Dictionaries, and $K \geq M$

To better understand the behavior of CRB for dictionaries with unnormalized columns, we consider a special case when $K \geq M$ and the non-zero hyper-parameters are all equal to p, i.e, $[\mathbf{p}]_i = p$ for $i \in \mathcal{S}$, and $[\mathbf{p}]_i = 0$ for $i \notin \mathcal{S}$. We further assume that $\tilde{\mathbf{A}}$ has full row rank M (which is possible since $K \geq M$). Consider the singular value decomposition of $\tilde{\mathbf{A}} \tilde{\mathbf{A}}^H$ as $\tilde{\mathbf{A}} \tilde{\mathbf{A}}^H = \mathbf{U} \mathbf{\Sigma} \mathbf{U}^H$ where $\mathbf{\Sigma} = \text{diag}(\sigma_1, \cdots, \sigma_M)$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_M > 0$. Thus, we can write

$$\mathbf{R} = p \tilde{\mathbf{A}} \tilde{\mathbf{A}}^H + \sigma_w^2 \mathbf{I} = \mathbf{U} [p \mathbf{\Sigma} + \sigma_w^2 \mathbf{I}] \mathbf{U}^H$$

Therefore, we have

$$\mathbf{R}^{-1} = \mathbf{U} \mathbf{\Gamma} \mathbf{U}^H$$

where $\mathbf{\Gamma} = \text{diag}(\frac{1}{p\sigma_1+\sigma_w^2}, \frac{1}{p\sigma_2+\sigma_w^2}, \cdots, \frac{1}{p\sigma_M+\sigma_w^2})$. Hence,

$$\mathbf{a}_i^H \mathbf{R}^{-1}\mathbf{a}_i \le \sigma_{\max}(\mathbf{R}^{-1})\|\mathbf{a}_i\|^2 = \frac{\|\mathbf{a}_i\|^2}{p\sigma_M + \sigma_w^2}$$

with $\sigma_{\max}(.)$ denoting the maximum singular value of a matrix. Using (4), we get

$$[\mathbf{C}]_{ii} \ge \eta_{\mathbb{F}}^{-1} \frac{(p\sigma_{\min}^2(\tilde{\mathbf{A}}) + \sigma_w^2)^2}{\|\mathbf{a}_i\|^4} \qquad (B_{z2})$$

where $\sigma_M = \sigma_{\min}^2(\tilde{\mathbf{A}})$, and $\sigma_{\min}(.)$ indicates the smallest nonzero singular value of a matrix.

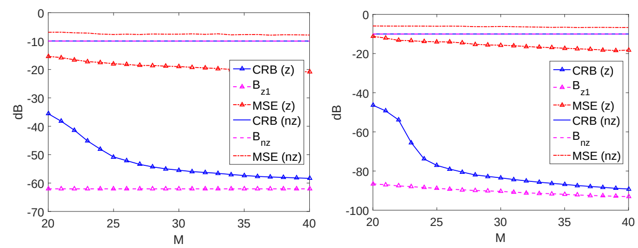
Proposition 1. *If* $\frac{\sigma_{\min}(\tilde{\mathbf{A}})}{\|\mathbf{a}_i\|} = \mathcal{O}(1)$ *(i.e. does not scale with* M *or* N*),* $[\mathbf{C}]_{ii}$ *in* (B_{z2}) *for* $i \notin \mathcal{S}$*, will be bounded below by a positive quantity as* $M \to \infty$*.*

Therefore, when the columns $\mathbf{A}$ are not normalized, saturation may or may not happen. This depends on the asymptotic behavior of the smallest singular value of $\tilde{\mathbf{A}}$ with respect to its column norm, as we increase the sizes M, N and K.

4. SIMULATIONS

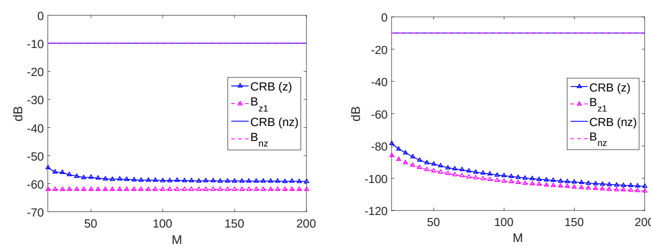
We conduct numerical experiments to examine the behavior of the CRB for zero and nonzero elements of $\mathbf{p}$, as we increase the size of dictionary $\mathbf{A}$. We generate a matrix $\mathbf{A_0}$ with i.i.d. standard normal entries, and let $\mathbf{A}$ be a submatrix of $\mathbf{A_0}$ by choosing the first M (resp. N) rows (resp. columns) of $\mathbf{A_0}$. In each simulation, we generate $\mathbf{p}$ such that the support corresponding to a smaller sparsity level K is a subset of the support corresponding to the larger value of K. We consider $L = 20$ i. i. d. realizations of the vector $\mathbf{x}$ with the same support. This essentially scales the CRB values by a factor of $\frac{1}{L}$ and does not affect our analysis, yet it can slightly improve the performance of our estimator (discussed later), whose error is compared with CRB. The noise variance is assumed to be $\sigma_w = 0.05$, and all the nonzero values of $\mathbf{p}$ are equal to 1.

We consider three different experimental settings, and for each case, we consider both normalized and unnormalized $\mathbf{A}$. In "Setting 1", we fix $N = 100$, $K = 10$ and increase M. In "Setting 2", we also let K and N grow as we increase M, such that $K = \lfloor \frac{M}{4} \rfloor$, $N = 4M$. "Setting 3" differs from "Setting 2" only in the fact that K can be larger than M. In particular, we let $K = 2M$, $N = 4M$. In all cases, we let the starting M to be $M = 20$, to ensure nonsingularity of the FIM. The experimental results for each scenario are plotted in Figure 2, where we compare the CRB with the lower-bounds established in this paper. In Fig. 1 (a,b) we also show the Mean Square Error (MSE) of the Maximum Likelihood (ML) estimator, and compare it with the CRB corresponding to Setting 1. The MSE of the ML estimator is computed by averaging over 2000 Monte-Carlo simulations for each M. We observe that the saturation effect always happens for both zero and non-zero elements when the dictionary is normalized, thereby validating our theoretical claims. When $\mathbf{A}$ is not normalized, the CRB corresponding to the zero elements seem to decrease monotonically for Settings 1 and 2 (where we have $K < M$). In future, we will explore the behavior of the CRB of zero elements in greater detail.

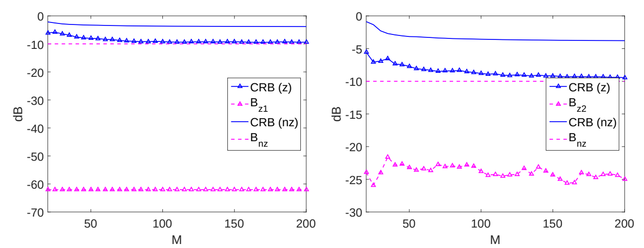


(a) Setting 1, Normalized (b) Setting 1, Unnormalized

Fig. 1. Comparisons between the CRB, the lower bounds established in this paper (indicated by their corresponding labels), and the MSE of ML algorithm. The label "(z)" indicates zero elements and "(nz)" represents nonzero elements.



(c) Setting 2, Normalized (d) Setting 2, Unnormalized



(e) Setting 3, Normalized (f) Setting 3, Unnormalized

Fig. 2. Comparison between the CRB, and the lower bounds, established in this paper. The labels in this Figure are the same as those in Fig. 1.

5. CONCLUSION

We considered the Marginalized Cramér Rao bound associated with hyper-parameter estimation in Sparse Bayesian Learning. We showed that the CRB corresponding to the nonzero elements is always bounded below by a positive quantity which does not go to zero as we increase the number measurements, thereby exhibiting saturation. However, for the zero elements, saturation of the CRB may or may not happen, depending on the column norm as well as the algebraic structure of the dictionary. We will further investigate this phenomena in future by deriving suitable upper bounds for the CRB corresponding to zero elements.

6. REFERENCES

[1] Michael E Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of machine learning research*, vol. 1, no. Jun, pp. 211–244, 2001.

[2] Michael E Tipping, Anita C Faul, et al., "Fast marginal likelihood maximisation for sparse bayesian models.," in *AISTATS*, 2003.

[3] David P Wipf and Bhaskar D Rao, "Sparse bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2153–2164, 2004.

[4] David P Wipf and Bhaskar D Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, 2007.

[5] Emmanuel J Candès et al., "Compressive sampling," in *Proceedings of the international congress of mathematicians*. Madrid, Spain, 2006, vol. 3, pp. 1433–1452.

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

[7] Zhilin Zhang and Bhaskar D Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 912–926, 2011.

[8] Ranjitha Prasad and Chandra R Murthy, "Cramér-rao-type bounds for sparse bayesian learning," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 622–632, 2013.

[9] Behtash Babadi, Nicholas Kalouptsidis, and Vahid 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.

[10] Zvika Ben-Haim and Yonina C Eldar, "The cramér-rao bound for estimating a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3384–3389, 2010.

[11] Piya Pal and P.P. Vaidyanathan, "Parameter identifiability in sparse bayesian learning," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 1851–1855.

[12] Louis L Scharf and LT McWhorter, "Geometry of the cramér–rao bound," *Signal Processing*, vol. 31, no. 3, pp. 301–311, 1993.

[13] Max A Woodbury, "Inverting modified matrices," *Memorandum report*, vol. 42, pp. 106, 1950.

[14] Jack Sherman and Winifred J Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *The Annals of Mathematical Statistics*, vol. 21, no. 1, pp. 124–127, 1950.