Cramér-Rao-Type Bounds for Sparse Bayesian Learning

Ranjitha Prasad and Chandra R. Murthy, *Senior Member, IEEE*

***Abstract*—In this paper, we derive Hybrid, Bayesian and Marginalized Cramér-Rao lower bounds (HCRB, BCRB and MCRB) for the single and multiple measurement vector Sparse Bayesian Learning (SBL) problem of estimating compressible vectors and their prior distribution parameters. We assume the unknown vector to be drawn from a compressible Student-t prior distribution. We derive CRBs that encompass the deterministic or random nature of the unknown parameters of the prior distribution and the regression noise variance. We extend the MCRB to the case where the compressible vector is distributed according to a general compressible prior distribution, of which the generalized Pareto distribution is a special case. We use the derived bounds to uncover the relationship between the compressibility and Mean Square Error (MSE) in the estimates. Further, we illustrate the tightness and utility of the bounds through simulations, by comparing them with the MSE performance of two popular SBL-based estimators. We find that the MCRB is generally the tightest among the bounds derived and that the MSE performance of the Expectation-Maximization (EM) algorithm coincides with the MCRB for the compressible vector. We also illustrate the dependence of the MSE performance of SBL based estimators on the compressibility of the vector for several values of the number of observations and at different signal powers.**

***Index Terms*—Cramér-Rao lower bounds, expectation maximization, mean square error, sparse Bayesian learning.**

I. Introduction

RECENT results in the theory of compressed sensing have generated immense interest in sparse vector estimation problems, resulting in a multitude of successful practical signal recovery algorithms. In several applications, such as the processing of natural images, audio, and speech, signals are not exactly sparse, but *compressible*, i.e., the magnitudes of the sorted coefficients of the vector follow a power law decay [1]. In [2] and [3], the authors show that random vectors drawn from a special class of probability distribution functions (pdf) known as *compressible priors* result in compressible vectors. Assuming that the vector to be estimated (henceforth referred to as the unknown vector) has a compressible prior distribution enables one to formulate the compressible vector recovery problem in the Bayesian framework, thus allowing the use of Sparse Bayesian Learning (SBL) techniques [4]. In his seminal work, Tipping proposed an SBL algorithm for estimating the unknown vector, based on the Expectation Maximization (EM) and McKay updates [4]. Since these update rules are known to be slow, fast update techniques are proposed in [5]. A duality based algorithm for solving the SBL cost function is proposed in [6], and $\ell_1 - \ell_2$ based reweighting schemes are explored in [7]. Such algorithms have been successfully employed for image/visual tracking [8], neuro-imaging [9], [10], beamforming [11], and joint channel estimation and data detection for OFDM systems [12].

Many of the aforementioned papers study the complexity, convergence and support recovery properties of SBL based estimators (e.g., [5], [6]). In [3], the general conditions required for the so-called instance optimality of such estimators are derived. However, it is not known whether these recovery algorithms are optimal in terms of the Mean Square Error (MSE) in the estimate or by how much their performance can be improved. In the context of estimating *sparse* signals, Cramér-Rao lower bounds on the MSE performance are derived in [13]–[15]. However, to the best of our knowledge, none of the existing works provide a lower bound on the MSE performance of *compressible* vector estimation. Such bounds are necessary, as they provide absolute yardsticks for comparative analysis of estimators, and may also be used as a criterion for minimization of MSE in certain problems [16]. In this paper, we close this gap in theory by providing Cramér-Rao type lower bounds on the MSE performance of estimators in the SBL framework.

As our starting point, we consider a linear Single Measurement Vector (SMV) SBL model given by

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}, \tag{1}$$

where the observations $\mathbf{y} \in \mathbb{R}^N$ and the measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{N\times L}$ are known, and $\mathbf{x} \in \mathbb{R}^L$ is the unknown sparse/compressible vector to be estimated [17]. Each component of the additive noise $\mathbf{n} \in \mathbb{R}^N$ is white Gaussian, distributed as $\mathcal{N}(0, \sigma^2)$, where the variance σ^2 may be known or unknown. The SMV-SBL model in (1) can be generalized to a linear Multiple Measurement Vector (MMV) SBL model given by

$$\mathbf{T} = \mathbf{\Phi}\mathbf{W} + \mathbf{V}. \tag{2}$$

Here, $\mathbf{T} \in \mathbb{R}^{N\times M}$ represents the M observation vectors, the columns of $\mathbf{W} \in \mathbb{R}^{L\times M}$ are the M sparse/compressible vectors with a common underlying distribution, and each column of $\mathbf{V} \in \mathbb{R}^{N\times M}$ is modeled similar to $\mathbf{n}$ in (1)[18].

In typical compressible vector estimation problems, $\mathbf{\Phi}$ is underdetermined $(N < L)$, rendering the problem ill-posed. Bayesian techniques circumvent this problem by using a prior distribution on the compressible vector as a regularization, and computing the corresponding posterior estimate. To incorporate a compressible prior in (1) and (2), SBL uses a two-stage

Manuscript received January 31, 2012; revised August 10, 2012; accepted October 08, 2012. Date of publication October 23, 2012; date of current version January 14, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jerome Idier.

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore 560012, India (e-mail: ranjitha.p@ece.iisc.ernet.in; cmurthy@ece.iisc.ernet.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2012.2226165

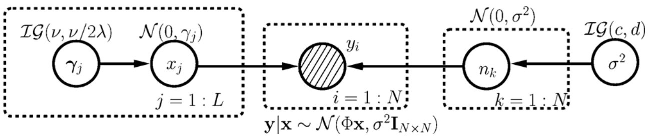

Fig. 1. Graphical model for SBL: Two stage hierarchical model with the compressible vector taking a conditional Gaussian distribution and the hyperparameters taking an Inverse Gamma distribution. The noise is modeled as white Gaussian distributed, with the noise variance modeled as deterministic/random and known or unknown.

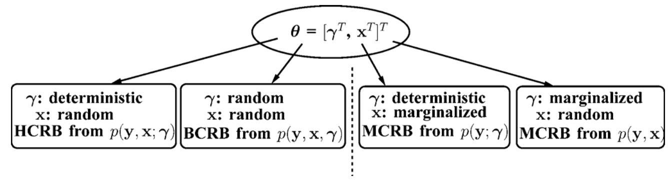

Fig. 2. Summary of the lower bounds derived in this work when noise variance is assumed to be known.

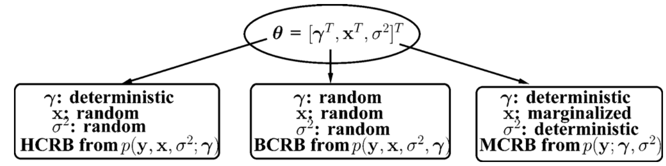

Fig. 3. Different modeling assumptions and the corresponding bounds derived in this work when noise variance is assumed to be unknown.

hierarchical model on the unknown vector, as shown in Fig. 1. Here, $\mathbf{x} \sim \mathcal{N}(0, \boldsymbol{\Upsilon})$, where the diagonal matrix $\boldsymbol{\Upsilon}$ contains the *hyperparameters* $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_L]^T$ as its diagonal elements. Further, an Inverse Gamma (IG) *hyperprior* is assumed for $\boldsymbol{\gamma}$ itself, because it leads to a Student-t prior on the vector $\mathbf{x}$, which is known to be compressible [4].[1] In scenarios where the noise variance is unknown and random, an IG prior is used for the distribution of the noise variance as well. For the system model in (2), every compressible vector $\mathbf{w}_i \sim \mathcal{N}(0, \boldsymbol{\Upsilon})$, i.e., the M compressible vectors are governed by a common $\boldsymbol{\Upsilon}$.

It is well known that the Cramér-Rao Lower Bound (CRLB) provides a fundamental limit on the MSE performance of unbiased estimators [19] for deterministic parameter estimation. For the estimation problem in SBL, an analogous bound known as the Bayesian Cramér-Rao Bound (BCRB) is used to obtain lower bounds [20], by incorporating the prior distribution on the unknown vector. If the unknown vector consists of both deterministic and random components, Hybrid Cramér-Rao Bounds (HCRB) are derived [21].

In SBL, the unknown vector estimation problem can also be viewed as a problem involving nuisance parameters. Since the assumed hyperpriors are conjugate to the Gaussian likelihood, the marginalized distributions have a closed form and the Marginalized Cramér-Rao Bounds (MCRB) [22] can be derived. For example, in the SBL hyperparameter estimation problem, $\mathbf{x}$ itself can be considered a nuisance variable and marginalized from the joint distribution, $p_{\mathbf{Y},\mathbf{X}|\boldsymbol{\gamma}}(\mathbf{y}, \mathbf{x}|\boldsymbol{\gamma})$, to obtain the log likelihood as

$$\log \int_{\mathbf{x}} p_{\mathbf{Y},\mathbf{X}|\Gamma}(\mathbf{y}, \mathbf{x}|\boldsymbol{\gamma}) d\mathbf{x} = \frac{-(\log|\boldsymbol{\Sigma}_y| + \mathbf{y}^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y})}{2}, \quad (3)$$

where $\boldsymbol{\Sigma}_y = \sigma^2 \mathbf{I}_{N\times N} + \boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^T$ [23].

The goal of this paper is to derive Cramér-Rao type lower bounds on the MSE performance of estimators based on the SBL framework. Our contributions are as follows:

- Under the assumption of known noise variance, we derive the HCRB and the BCRB for the unknown vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, as indicated in the left half of Fig. 2.
- When the noise variance is known, we marginalize nuisance variables ($\boldsymbol{\gamma}$ or $\mathbf{x}$) and derive the corresponding MCRB, as indicated in the right half of Fig. 2. Since the MCRB is a function of the parameters of the hyperprior (and hence is an offline bound), it yields insights into the relationship between the MSE performance of the estimators and the compressibility of $\mathbf{x}$.
- In the unknown noise variance case, we derive the BCRB, HCRB and MCRB for the unknown vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T, \sigma^2]^T$, as indicated in Fig. 3.
- We derive the MCRB for a general parametric form of the compressible prior [3] and deduce lower bounds for two of the well-known compressible priors, namely, the Student-t and generalized double Pareto distributions.
- Similar to the SMV-SBL case, we derive the BCRB, HCRB and MCRB for the MMV-SBL model in (2).

Through numerical simulations, we show that the MCRB on the compressible vector $\mathbf{x}$ is the tightest lower bound, and that the MSE performance of the EM algorithm achieves this bound at high SNR and as $N \rightarrow L$. The techniques used to derive the bounds can be extended to handle different compressible prior pdfs used in literature [2]. These results provide a convenient and easy-to-compute benchmark for comparing the performance of the existing estimators, and in some cases, for establishing their optimality in terms of the MSE performance.

The rest of this paper is organized as follows. In Section II, we provide the basic definitions and describe the problem set up. In Sections III and IV, we derive the lower bounds for the cases shown in Figs. 2 and 3, respectively. The bounds are extended to the MMV-SBL signal model in Section V. The efficacy of the lower bounds is graphically illustrated through simulation results in Section VI. We provide some concluding remarks in Section VII. In the Appendix, we provide proofs for the Propositions and Theorems stated in the paper.

Notation: In the sequel, boldface small letters denote vectors and boldface capital letters denote matrices. The symbols $(\cdot)^T$ and $|\cdot|$ denote the transpose and determinant of a matrix, respectively. The empty set is represented by $\emptyset$, and $\Gamma(\cdot)$ denotes the Gamma function. The function $p_X(x)$ represents the pdf of the random variable X evaluated at its realization x. Also, $\text{diag}(\mathbf{a})$ stands for a diagonal matrix with entries on the diagonal given by the vector $\mathbf{a}$. The symbol $\nabla_{\boldsymbol{\theta}}$ is the gradient with respect to (w.r.t.) the vector $\boldsymbol{\theta}$. The expectation w.r.t. a random variable X is denoted as $\mathbb{E}_X(\cdot)$. Also, $\mathbf{A} \succeq \mathbf{B}$ denotes that $\mathbf{A} - \mathbf{B}$ is positive semidefinite, and $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of the two matrices $\mathbf{A}$ and $\mathbf{B}$.

[1]The IG hyperprior is conjugate to the Gaussian pdf [4].

II. Preliminaries

As a precursor to the sections that follow, we define the MSE matrix and the Fisher Information Matrix (FIM) [19], and state the assumptions under which we derive the lower bounds in this paper. Consider a general estimation problem where the unknown vector $\boldsymbol{\theta} \in \mathbb{R}^n$ can be split into sub-vectors $\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T,\ \boldsymbol{\theta}_d^T]^T$, where $\boldsymbol{\theta}_r \in \mathbb{R}^m$ consists of *random* parameters distributed according to a known pdf, and $\boldsymbol{\theta}_d \in \mathbb{R}^{n-m}$ consists of *deterministic* parameters. Let $\hat{\boldsymbol{\theta}}(\mathbf{y})$ denote the estimator of $\boldsymbol{\theta}$ as a function of the observations $\mathbf{y}$. The MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ is defined as

$$\mathbf{E}^{\boldsymbol{\theta}} \triangleq \mathbb{E}_{\mathbf{Y},\Theta_r}\left[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{y}))(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{y}))^T\right], \tag{4}$$

where Θ_r denotes the random parameters to be estimated, whose realization is given by $\boldsymbol{\theta}_r$. The first step in obtaining Cramér-Rao type lower bounds is to derive the FIM $\mathbf{I}^{\boldsymbol{\theta}}$ [19]. Typically, $\mathbf{I}^{\boldsymbol{\theta}}$ is expressed in terms of the individual blocks of submatrices, where the $(ij)^{\text{th}}$ block is given by

$$\mathbf{I}^{\boldsymbol{\theta}}_{ij} \triangleq -\mathbb{E}_{\mathbf{Y},\Theta_r}[\nabla_{\boldsymbol{\theta}_i}\nabla^T_{\boldsymbol{\theta}_j}\log p_{\mathbf{Y},\Theta_r;\Theta_d}(\mathbf{y},\boldsymbol{\theta}_r;\boldsymbol{\theta}_d)]. \tag{5}$$

In this paper, we use the notation $\mathbf{I}^{\boldsymbol{\theta}}$ to represent the FIM under the different modeling assumptions. For example, when $\boldsymbol{\theta}_r \neq \emptyset$ and $\boldsymbol{\theta}_d \neq \emptyset$, $\mathbf{I}^{\boldsymbol{\theta}}$ represents a Hybrid Information Matrix (HIM). When $\boldsymbol{\theta}_r \neq \emptyset$ and $\boldsymbol{\theta}_d = \emptyset$, $\mathbf{I}^{\boldsymbol{\theta}}$ represents a Bayesian Information matrix (BIM). Assuming that the MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ exists and the FIM is non-singular, a lower bound on the MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ is given by the inverse of the FIM:

$$\mathbf{E}^{\boldsymbol{\theta}} \succeq \left(\mathbf{I}^{\boldsymbol{\theta}}\right)^{-1}. \tag{6}$$

It is easy to verify that the underlying pdfs considered in the SBL model satisfy the regularity conditions required for computing the FIM (see Sec. 5.2.3 in [22]).

We conclude this section by making one useful observation about the FIM in the SBL problem. An assumption in the SMV-SBL framework is that $\mathbf{x}$ and $\mathbf{n}$ are independent of each other (for the MMV-SBL model, $\mathbf{T}$ and $\mathbf{W}$ are independent). This assumption is reflected in the graphical model in Fig. 1, where the compressible vector $\mathbf{x}$ (and its attribute $\boldsymbol{\gamma}$) and the noise component $\mathbf{n}$ (and its attribute σ^2) are on unconnected branches. Due to this, a submatrix of the FIM is of the form

$$\begin{aligned}\mathbf{I}^{\boldsymbol{\theta}}_{\gamma\xi} = -\mathbb{E}_{\mathbf{X},\mathbf{Y},\Gamma,\Xi}\big[\nabla_{\boldsymbol{\gamma}}\nabla_{\xi}\{&\log p_{\mathbf{Y}|\mathbf{X},\Xi}(\mathbf{y}|\mathbf{x},\xi)\\ &+\log p_{\mathbf{X},\Gamma}(\mathbf{x},\boldsymbol{\gamma})+\log p_{\Xi}(\xi)\}\big],\end{aligned} \tag{7}$$

where there are no terms in which both $\boldsymbol{\gamma}$ and $\xi = \sigma^2$ are jointly present. Hence, the corresponding terms in the above mentioned submatrix are always zero. This is formally stated in the following Lemma.

Lemma 1: When $\boldsymbol{\theta}_i = \boldsymbol{\gamma}$ and $\boldsymbol{\theta}_j = \sigma^2$, the $(ij)^{\text{th}}$ block matrix of the FIM $\mathbf{I}^{\boldsymbol{\theta}}$ given by (5) simplifies to $\mathbf{I}^{\boldsymbol{\theta}}_{ij} = \mathbf{0}_{L\times 1}$, i.e., to an all zero vector.

III. SMV-SBL: Lower Bounds When σ^2 is Known

In this section, we derive lower bounds for the system model in (1) for the scenarios in Fig. 2, where the unknown vector is $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$. We examine different modeling assumptions on $\boldsymbol{\gamma}$ and derive the corresponding lower bounds.

A. Bounds From the Joint Pdf

1) HCRB for $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$: In this subsection, we consider the unknown variables as a hybrid of a deterministic vector $\boldsymbol{\gamma}$ and a random vector $\mathbf{x}$ distributed according to a Gaussian distribution parameterized by $\boldsymbol{\gamma}$. Using the assumptions and notation in the previous section, we obtain the following proposition.

Proposition 1: For the signal model in (1), the HCRB on the MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ of the unknown vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$ with the parameterized distribution of the compressible signal $\mathbf{x}$ given by $\mathcal{N}(0, \boldsymbol{\Upsilon})$, and with $\boldsymbol{\gamma}$ modeled as unknown and deterministic, is given by $\mathbf{E}^{\boldsymbol{\theta}} \succeq (\mathbf{H}^{\boldsymbol{\theta}})^{-1}$, where

$$\begin{aligned}\mathbf{H}^{\boldsymbol{\theta}} &\triangleq \begin{bmatrix}\mathbf{H}^{\boldsymbol{\theta}}(\mathbf{x}) & \mathbf{H}^{\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\gamma})\\ (\mathbf{H}^{\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\gamma}))^T & \mathbf{H}^{\boldsymbol{\theta}}(\boldsymbol{\gamma})\end{bmatrix}\\ &= \begin{bmatrix}\left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2}+\boldsymbol{\Upsilon}^{-1}\right) & \mathbf{0}_{L\times L}\\ \mathbf{0}_{L\times L} & \operatorname{diag}(2\gamma_1^2,\ 2\gamma_2^2,\ \ldots,\ 2\gamma_L^2)^{-1}\end{bmatrix}.\end{aligned} \tag{8}$$

Proof: See Appendix A.

Note that the lower bound on the estimate of $\mathbf{x}$ depends on the prior information through the diagonal matrix $\boldsymbol{\Upsilon}$. In the SBL problem, the realization of the random parameter $\boldsymbol{\gamma}$ has to be used to compute the bound above, and hence, it is referred to as an online bound. Also, the lower bound on the MSE matrix of $\mathbf{x}$ is $\mathbf{E}^{\boldsymbol{\theta}} \succeq \left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2}+\boldsymbol{\Upsilon}^{-1}\right)^{-1}$, which is the same as the lower bound on the error covariance of the Baye's vector estimator for a linear model (see Theorems 10.2 and 10.3 in [19]), and is achievable by the MMSE estimator when $\boldsymbol{\Upsilon} = \operatorname{diag}(\gamma_1,\ldots,\gamma_L)$ is known.

2) BCRB for $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$: For deriving the BCRB, a hyperprior distribution is considered on $\boldsymbol{\gamma}$, and the resulting $\mathbf{x}$ is viewed as being drawn from a compressible prior distribution. The most commonly used hyperprior distribution in the literature is the IG distribution [4], where $\gamma_i,\ i = 1, 2, \ldots, L$ are distributed as $\mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right)$, given by

$$p_{\Gamma}(\gamma_i) \triangleq \left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1}\left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}}\gamma_i^{\left(-\frac{\nu}{2}-1\right)}\exp\left\{-\frac{\nu}{2\lambda\gamma_i}\right\}, \tag{9}$$

where $\gamma_i \in (0,\infty)$, $\nu,\ \lambda > 0$. Using the definitions and notation in the previous section, we state the following proposition.

Proposition 2: For the signal model in (1), the BCRB on the MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}$ of the unknown random vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, where the conditional distribution of the compressible signal $\mathbf{x}|\boldsymbol{\gamma}$ is $\mathcal{N}(0, \boldsymbol{\Upsilon})$, and the hyperprior distribution on $\boldsymbol{\gamma}$ is $\prod_{i=1}^{L}\mathcal{IG}\left(\frac{\nu}{2}, \frac{\nu}{2\lambda}\right)$, is given by $\mathbf{E}^{\boldsymbol{\theta}} \succeq (\mathbf{B}^{\boldsymbol{\theta}})^{-1}$, where

$$\begin{aligned}\mathbf{B}^{\boldsymbol{\theta}} &\triangleq \begin{bmatrix}\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}) & \mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\gamma})\\ (\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\gamma}))^T & \mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma})\end{bmatrix}\\ &= \begin{bmatrix}\left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2}+\lambda\mathbf{I}_{L\times L}\right) & \mathbf{0}_{L\times L}\\ \mathbf{0}_{L\times L} & \frac{\lambda^2(\nu+2)(\nu+7)}{2\nu}\mathbf{I}_{L\times L}\end{bmatrix}.\end{aligned} \tag{10}$$

Proof: See Appendix B.

It can be seen from $\mathbf{B}^{\boldsymbol{\theta}}$ that the lower bound on the MSE of $\hat{\boldsymbol{\gamma}}(\mathbf{y})$ is a function of the parameters of the IG prior on $\boldsymbol{\gamma}$, i.e., a function of ν and λ, and it can be computed without the knowledge of realization of $\boldsymbol{\gamma}$. Thus, it is an offline bound.

B. Bounds From Marginalized Distributions

1) MCRB for $\boldsymbol{\theta} = [\boldsymbol{\gamma}]$: Here, we derive the MCRB for $\boldsymbol{\theta} = [\boldsymbol{\gamma}]$, where $\boldsymbol{\gamma}$ is an unknown deterministic parameter. This requires the marginalized distribution $p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})$, which is obtained by considering $\mathbf{x}$ as a nuisance variable and marginalizing it out of the joint distribution $p_{\mathbf{X},\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{x},\mathbf{y};\boldsymbol{\gamma})$, to obtain (3). Since $\boldsymbol{\gamma}$ is a deterministic parameter, the pdf $p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})$ must satisfy the regularity condition in [19]. We have the following theorem.

Theorem 1: For the signal model in (1), the log likelihood function $\log p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})$ satisfies the regularity conditions in [19]. Further, the MCRB on the MSE matrix $\mathbf{E}^{\gamma}$ of the unknown deterministic vector $\boldsymbol{\theta} = [\boldsymbol{\gamma}]$ is given by $\mathbf{E}^{\gamma} \succeq (\mathbf{M}^{\gamma})^{-1}$, where the $(ij)^{\text{th}}$ element of $\mathbf{M}^{\gamma}$ is given by

$$\mathbf{M}^{\gamma}_{ij} = \frac{1}{2}(\Phi_j^T \boldsymbol{\Sigma}_y^{-1} \Phi_i)^2, \qquad (11)$$

for $1 \le i,\ j \le L$, where Φ_i is the i^{th} column of $\boldsymbol{\Phi}$, and $\boldsymbol{\Sigma}_y = \sigma^2 \mathbf{I}_{N\times N} + \boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^T$, as defined earlier.

Proof: See Appendix C.

To intuitively understand (11), we consider a special case of $\boldsymbol{\Phi}^T\boldsymbol{\Phi} = N\mathbf{I}_{N\times N}$, and use the Woodbury formula to simplify $\boldsymbol{\Sigma}_y^{-1}$, to obtain the $(ii)^{\text{th}}$ entry of the matrix $\mathbf{M}^{\gamma}$ as

$$\mathbf{M}^{\gamma}_{ii} = 2\left(\frac{\sigma^2}{N} + \gamma_i\right)^{-2}. \qquad (12)$$

Hence, the error in γ_i is bounded as $\mathbf{E}^{\gamma}_{ii} \ge 2\left(\frac{\sigma^2}{N} + \gamma_i\right)^2$. As $N \to \infty$, the bound reduces to $2\gamma_i^2$, which is the same as the lower bound on the estimate of $\boldsymbol{\gamma}$ obtained as the lower-right submatrix in (8). For finite N, the MCRB is tighter than the HCRB.

2) MCRB for $\boldsymbol{\theta} = [\mathbf{x}]$: In this subsection, we assume a hyperprior on $\boldsymbol{\gamma}$, which leads to a joint distribution of $\mathbf{x}$ and $\boldsymbol{\gamma}$, from which $\boldsymbol{\gamma}$ can be marginalized. Further, assuming specific forms for the hyperprior distribution can lead to a compressible prior on $\mathbf{x}$. For example, assuming an IG hyperprior on $\boldsymbol{\gamma}$ leads to an $\mathbf{x}$ with a Student-t distribution. Sampling from a Student-t distribution with parameters ν and λ results in a ν-compressible $\mathbf{x}$ [2]. The Student-t prior is given by

$$p_{\mathbf{X}}(\mathbf{x}) \triangleq \left(\frac{\Gamma\left(\frac{(\nu+1)}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\right)^L \left(\frac{\lambda}{\pi\nu}\right)^{\frac{L}{2}} \prod_{i=1}^{L}\left(1 + \frac{\lambda x_i^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \qquad (13)$$

where $x_i \in (-\infty,\infty),\ \nu,\ \lambda > 0$, ν represents the number of degrees of freedom and λ represents the inverse variance of the distribution. Using the notation developed so far, we state the following theorem.

Theorem 2: For the signal model in (1), the MCRB on the MSE matrix $\mathbf{E}^{\mathbf{x}}$ of the unknown compressible random vector $\boldsymbol{\theta} = [\mathbf{x}]$ distributed as (13), is given by $\mathbf{E}^{\mathbf{x}} \succeq (\mathbf{M}^{\mathbf{x}})^{-1}$, where

$$\mathbf{M}^{\mathbf{x}} = \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \frac{\lambda(\nu+1)}{(\nu+3)}\mathbf{I}_{L\times L}. \qquad (14)$$

Proof: See Appendix D.

We see that the bound derived depends on the parameters of the Student-t pdf. From [3], the prior is "*somewhat*" compressible for $2 < \nu < 4$, and (14) is nonnegative and bounded for $2 < \nu < 4$, i.e., the bound is meaningful in the range of ν used in practice. Note that, by choosing λ to be large (or the variance of $\mathbf{x}$ to be small), the bound is dominated by the prior information, rather than the information from the observations, as expected in Bayesian bounds [19].

It is conjectured in [22] that, in general, the MCRB is tighter than the BCRB. Analytically comparing the MCRB (14) with the BCRB (8), we see that for the SBL problem of estimating a compressible vector, the MCRB is indeed tighter than the BCRB, since

$$\left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \frac{\lambda(\nu+1)}{(\nu+3)}\mathbf{I}_{L\times L}\right)^{-1} \succeq \left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \lambda\mathbf{I}_{L\times L}\right)^{-1}.$$

The techniques used to derive the bounds in this subsection can be applied to any family of compressible distributions. In [3], the authors propose a parametric form of the Generalized Compressible Prior (GCP) and prove that such a prior is compressible for certain values of ν. In the following subsection, we derive the MCRB for the GCP.

C. General Marginalized Bounds

In this subsection, we derive MCRBs for the parametric form of the GCP. The GCP encompasses the double Pareto shrinkage type prior [24] and the Student-t prior (13) as its special cases. We consider the GCP on $\mathbf{x}$ as follows

$$p_{\mathbf{X}}(\mathbf{x}) \triangleq K^L \prod_{i=1}^{L}\left(1 + \frac{\lambda|x_i|^\tau}{\nu}\right)^{\frac{-(\nu+1)}{\tau}}, \qquad (15)$$

where $x_i \in (-\infty,\infty),\ \tau,\ \nu,\ \lambda > 0$, and the normalizing constant $K \triangleq \frac{\tau}{2}\left(\frac{\lambda}{\nu}\right)^{\frac{1}{\tau}} \frac{\Gamma\left(\frac{(\nu+1)}{\tau}\right)}{\Gamma\left(\frac{1}{\tau}\right)\Gamma\left(\frac{\nu}{\tau}\right)}$. When $\tau = 2$, (15) reduces to the Student-t prior in (13), and when $\tau = 1$, it reduces to a generalized double Pareto shrinkage prior [24], [25]. Also, the expression for the GCP in [3] can be obtained from (15) by setting $\lambda = 1$, and defining $\nu \triangleq s - 1$. The following theorem provides the MCRB for the GCP.

Theorem 3: For the signal model in (1), the MCRB on the MSE matrix $\mathbf{E}^{\boldsymbol{\theta}}_{\tau}$ of the unknown random vector $\boldsymbol{\theta} = [\mathbf{x}]$, where $\mathbf{x}$ is distributed as the GCP in (15), is given by $\mathbf{E}^{\boldsymbol{\theta}}_{\tau} \succeq (\mathbf{M}^{\boldsymbol{\theta}}_{\tau})^{-1}$, where

$$\mathbf{M}^{\boldsymbol{\theta}}_{\tau} = \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + T_\tau, \qquad (16)$$

where $T_\tau = \frac{\tau^2(\nu+1)}{(\nu+\tau+1)}\left(\frac{\lambda}{\nu}\right)^{\frac{2}{\tau}} \frac{\Gamma\left(\frac{\nu+2}{\tau}\right)\Gamma\left(2-\frac{1}{\tau}\right)}{\Gamma\left(\frac{1}{\tau}\right)\Gamma\left(\frac{\nu}{\tau}\right)}\mathbf{I}_{L\times L}$.

Proof: See Appendix E.

It is straightforward to verify that for $\tau = 2$, (16) reduces to the MCRB derived in (14) for the Student-t distribution. For $\tau = 1$, the inverse of the MCRB can be reduced to

$$\mathbf{M}^{\boldsymbol{\theta}}_{\tau} = \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \frac{\lambda^2(\nu+1)^2}{\nu(\nu+2)}\mathbf{I}_{L\times L}. \qquad (17)$$

In Fig. 4, we plot the expression in (16). We observe that, in general, the bounds predict an increase in MSE for higher values of τ. Also, for given value of N, the lower bounds at different signal to noise ratios (SNRs) converge as the value

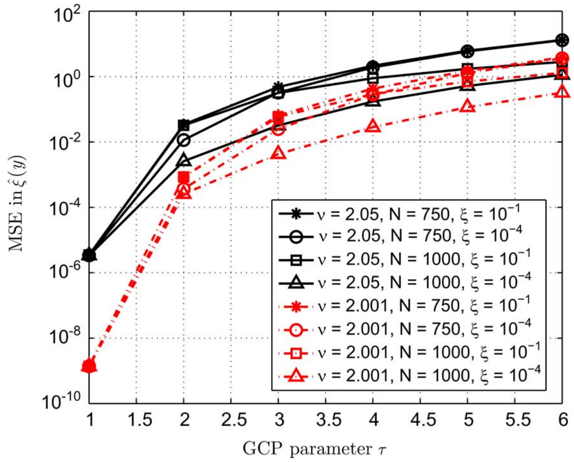

Fig. 4. Behavior of the MCRB (16) for the parametric form of the GCP, as a function of τ, ν, N and noise variance ξ.

of τ increases, indicating that increasing τ renders the bound insensitive to the SNR. The lower bounds also predict a smaller value of MSE for a lower value of ν.

Thus far, we have presented the lower bounds on the MSE in estimating the unknown parameters of the SBL problem when the noise variance is known. In the next section, we extend the results to the case of unknown noise variance.

IV. SMV-SBL: Lower Bounds When σ^2 is Unknown

Let us denote the unknown noise variance as $\xi = \sigma^2$. In the Bayesian formulation, the noise variance is associated with a prior, and since the IG prior is conjugate to the Gaussian likelihood $p_{\mathbf{Y}|\mathbf{X},\Xi}(\mathbf{y}|\mathbf{x},\xi)$, it is assumed that $\sigma^2 \sim \mathcal{IG}(c,d)$[4], i.e., $\xi = \sigma^2$ is distributed as

$$p_{\Xi}(\xi) \triangleq \frac{d^c}{\Gamma(c)}\xi^{(-c-1)}\exp\left\{-\frac{d}{\xi}\right\}; \quad \xi \in (0,\infty),\ c,\ d > 0. \tag{18}$$

Under this assumption, one can marginalize the unknown noise variance and obtain the likelihood $p(\mathbf{y}|\mathbf{x})$ as

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &\triangleq \int_{\xi=0}^{\infty} p(\mathbf{y},\xi|\mathbf{x})\mathrm{d}\xi \\ &= \frac{(2d)^c\Gamma\left(\frac{N}{2}+c\right)}{\Gamma(c)(\pi)^{\frac{N}{2}}} \\ &\quad \times \left((\mathbf{y}-\mathbf{\Phi x})^T(\mathbf{y}-\mathbf{\Phi x})+2d\right)^{-\left(\frac{N}{2}+c\right)}, \end{aligned} \tag{19}$$

which is a multivariate Student-t distribution. It turns out that the straightforward approach of using the above multivariate likelihood to directly compute lower bounds for the various cases given in the previous section is analytically intractable, and that the lower bounds cannot be computed in closed form. Hence, we compute lower bounds from the *joint* pdf, i.e., we derive the HCRB and BCRBs for the unknown vector $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T, \xi]^T$ with the MSE matrix $\mathbf{E}_{\xi}^{\boldsymbol{\theta}}$ defined by (4).[2] Using the assumptions and notation from the previous sections, we obtain the following proposition.

[2]We use the subscript ξ to indicate that the error matrices and bounds are obtained for the case of unknown noise variance.

Proposition 3: For the signal model in (1), the HCRB on the MSE matrix $\mathbf{E}_{\xi}^{\boldsymbol{\theta}}$ of the unknown vector $\boldsymbol{\theta} = [\boldsymbol{\theta}'^T, \xi]^T$, where $\boldsymbol{\theta}' = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, with the distribution of the compressible vector $\mathbf{x}$ given by $\mathcal{N}(0,\mathbf{\Upsilon})$, where $\boldsymbol{\gamma}$ is modeled as a deterministic or as a random parameter distributed as $\prod_{i=1}^{L}\mathcal{IG}\left(\frac{\nu}{2},\frac{\nu}{2\lambda}\right)$, and ξ is modeled as a deterministic parameter, is given by $(\mathbf{H}_{\xi}^{\boldsymbol{\theta}})^{-1}$, where

$$\mathbf{H}_{\xi}^{\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{H}^{\boldsymbol{\theta}'} & \mathbf{0}_{L\times 1} \\ \mathbf{0}_{1\times L} & \frac{N}{2\xi^2} \end{bmatrix}. \tag{20}$$

In the above expression, with a slight abuse of notation, $\mathbf{H}^{\boldsymbol{\theta}'}$ is the FIM given by (8) when $\boldsymbol{\gamma}$ is unknown deterministic and by (10) when $\boldsymbol{\gamma}$ is random.

Proof: See Appendix F.

The lower bound on the estimation of ξ matches with known lower bounds on noise variance estimation (see Sec. 3.5 in [19]). One disadvantage of such a bound on $\hat{\xi}(\mathbf{y})$ is that the knowledge of the noise variance is essential to compute the bound, and hence, it cannot be computed offline. Instead, assigning a hyperprior to ξ would result in a lower bound that only depends on the parameters of the hyperprior, which are assumed to be known, allowing the bound to be computed offline. We state the following proposition in this context.

Proposition 4: For the signal model in (1), the HCRB on the MSE matrix $\mathbf{E}_{\xi}^{\boldsymbol{\theta}}$ of the unknown vector $\boldsymbol{\theta} = [\boldsymbol{\theta}'^T, \xi]^T$, where $\boldsymbol{\theta}' = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, with the distribution of the vector $\mathbf{x}$ given by $\mathcal{N}(0,\mathbf{\Upsilon})$, where $\boldsymbol{\gamma}$ is modeled as a deterministic parameter or as a random parameter distributed as $\prod_{i=1}^{L}\mathcal{IG}\left(\frac{\nu}{2},\frac{\nu}{2\lambda}\right)$, and with the random parameter ξ distributed as $\mathcal{IG}(c,d)$, is given by $(\mathbf{H}_{\xi}^{\boldsymbol{\theta}})^{-1}$, where

$$\mathbf{H}_{\xi}^{\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{H}^{\boldsymbol{\theta}'} & \mathbf{0}_{L\times 1} \\ \mathbf{0}_{1\times L} & \frac{c(c+1)(\frac{N}{2}+c+3)}{d^2} \end{bmatrix}. \tag{21}$$

In (21), $\mathbf{H}^{\boldsymbol{\theta}'}$ is the FIM given in (8) when $\boldsymbol{\gamma}$ is unknown deterministic and by (10) when $\boldsymbol{\gamma}$ is random.

Proof: See Appendix G.

In SBL problems, a non-informative prior on ξ is typically preferred, i.e., the distribution of the noise variance is modeled to be as flat as possible. In [4], it was observed that a non-informative prior is obtained when $c,\ d \rightarrow 0$. However, as $c,\ d \rightarrow 0$, the bound in (21) is indeterminate. In Section VI, we illustrate the performance of the lower bound in (21) for practical values of c and d.

A. Marginalized Bounds

In this subsection, we obtain lower bounds on the MSE of the estimator $\hat{\xi}(\mathbf{y})$, in the presence of nuisance variables in the joint distribution. To start with, we consider the marginalized distributions of $\boldsymbol{\gamma}$ and ξ, i.e., $p_{\mathbf{Y};\boldsymbol{\gamma},\xi}(\mathbf{y};\boldsymbol{\gamma},\xi)$ where both, $\boldsymbol{\gamma}$ and ξ are deterministic variables. Since the unknowns are deterministic, the regularity condition has to be satisfied for $\boldsymbol{\theta} = [\boldsymbol{\gamma}^T, \xi]^T$. We state the following theorem.

Theorem 4: For the signal model in (1), the log likelihood function $\log p_{\mathbf{Y};\boldsymbol{\gamma},\xi}(\mathbf{y};\boldsymbol{\gamma},\xi)$ satisfies the regularity condition [19]. Further, the MCRB on the MSE matrix $\mathbf{E}_{\xi}^{\boldsymbol{\theta}}$ of the unknown

TABLE I
CRAMÉR-RAO TYPE BOUNDS FOR THE MMV-SBL CASE

Bound Derived	Expression
HCRB on $\hat{\boldsymbol{\gamma}}(\mathbf{y})$	$\mathbf{H}_M^{\boldsymbol{\theta}} = \text{diag}\left(\frac{M}{2\gamma_i^2}\right), i = 1, 2 \ldots, L$
BCRB on $\hat{\boldsymbol{\gamma}}(\mathbf{y})$	$\mathbf{B}_M^{\boldsymbol{\theta}} = \frac{\lambda^2(\nu+2)(M+\nu+6)}{2\nu}\mathbf{I}_{L\times L}$
MCRB on $\hat{\boldsymbol{\gamma}}(\mathbf{y})$	$\mathbf{M}_M^{\boldsymbol{\theta}} = [\mathbf{M}_{ij}^{\boldsymbol{\theta}}]$, where $\mathbf{M}_{ij}^{\boldsymbol{\theta}} = \frac{M}{2}(\Phi_j^T \boldsymbol{\Sigma}_y^{-1}\Phi_i)^2$
HCRB on $\hat{\mathbf{w}}(\mathbf{y})$	$\mathbf{H}_M^{\boldsymbol{\theta}} = \left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \boldsymbol{\Upsilon}^{-1}\right) \otimes \mathbf{I}_{M\times M}$
BCRB on $\hat{\mathbf{w}}(\mathbf{y})$	$\mathbf{B}_M^{\boldsymbol{\theta}} = \left(\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \lambda\mathbf{I}_{L\times L}\right) \otimes \mathbf{I}_{M\times M}$
HCRB on $\hat{\xi}(\mathbf{y})$	$\mathbf{H}_{M,\xi}^{\boldsymbol{\theta}} = \left(\frac{MN}{2\xi^2}\right)$
BCRB on $\hat{\xi}(\mathbf{y})$	$\mathbf{B}_{M,\xi}^{\boldsymbol{\theta}} = \frac{c(\frac{MN}{2}+c+3)(c+1)}{d^2}$
MCRB on $[\hat{\boldsymbol{\gamma}}(\mathbf{y})^T, \hat{\xi}(\mathbf{y})]^T$	$\mathbf{M}_{M,\xi}^{\boldsymbol{\theta}} = M \times \mathbf{M}_\xi^{\boldsymbol{\theta}}$

deterministic vector $\boldsymbol{\theta} = [\boldsymbol{\gamma}^T, \xi]^T$ is given by $\mathbf{E}_\xi^{\boldsymbol{\theta}} \succeq (\mathbf{M}_\xi^{\boldsymbol{\theta}})^{-1}$, where

$$\mathbf{M}_\xi^{\boldsymbol{\theta}} \triangleq \begin{bmatrix} \mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) & \mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma}, \xi) \\ \mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi, \boldsymbol{\gamma}) & \mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi) \end{bmatrix}, \tag{22}$$

where the $(ij)^{\text{th}}$ entry of the matrix $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ is given by $(\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma}))_{ij} = \frac{1}{2}\left\{(\Phi_j^T \boldsymbol{\Sigma}_y^{-1}\Phi_i)^2\right\}$, and $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi) = \frac{1}{2}\text{Tr}(\boldsymbol{\Sigma}_y^{-2})$. Further, $(\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma}, \xi))_i = (\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi, \boldsymbol{\gamma}))_i = \frac{\Phi_i^T \boldsymbol{\Sigma}_y^{-2}\Phi_i}{2}$, $i, j = 1, 2, \ldots, L$.

Proof: See Appendix H.

Remark: From the graphical model in Fig. 1, it can be seen that the branches consisting of γ_i and ξ are independent conditioned on $\mathbf{x}$. However, when $\mathbf{x}$ is marginalized, the nodes ξ and γ_i are connected, and hence, Lemma 1 is no longer valid. Due to this, the lower bound on $\boldsymbol{\gamma}$ depends on ξ and vice versa, i.e., $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ and $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi)$ depend on both ξ and $\boldsymbol{\Upsilon} = \text{diag}(\boldsymbol{\gamma})$ through $\boldsymbol{\Sigma}_y = \xi\mathbf{I}_{N\times N} + \boldsymbol{\Phi}\boldsymbol{\Upsilon}\boldsymbol{\Phi}^T$.

Thus far, we have presented several bounds for the MSE performance of the estimators $\hat{\mathbf{x}}(\mathbf{y})$, $\hat{\boldsymbol{\gamma}}(\mathbf{y})$ and $\hat{\xi}(\mathbf{y})$ in the SMV-SBL framework. In the next section, we derive Cramér-Rao type lower bounds for the MMV-SBL signal model.

V. LOWER BOUNDS FOR THE MMV-SBL

In this section, we provide Cramér-Rao type lower bounds for the estimation of unknown parameters in the MMV-SBL model given in (2). We consider the estimation of the compressible vector $\mathbf{w}$ from the vector of observations $\mathbf{t}$, which contain the stacked columns of $\mathbf{W}$ and $\mathbf{T}$, respectively. In the MMV-SBL model, each column of $\mathbf{W}$ is distributed as $\mathbf{w}_i \sim \mathcal{N}(0, \boldsymbol{\Upsilon})$, for $i = 1, \ldots M$, and the likelihood is given by $\prod_{i=1}^M p_{\mathbf{T}|\mathbf{W}_i,\Xi}(\mathbf{t}_i|\mathbf{w}_i, \xi)$, where $p_{\mathbf{T}|\mathbf{W}_i\Xi}(\mathbf{t}_i|\mathbf{w}_i, \xi) = \mathcal{N}(\boldsymbol{\Phi}\mathbf{w}_i, \xi)$ and $\xi = \sigma^2$. The modeling assumptions on $\boldsymbol{\gamma}$ and ξ are the same as in the SMV-SBL case, given by (9) and (18), respectively [18].

Using the notation developed in Section II, we derive the bounds for the MMV SBL case similar to the SMV-SBL cases considered in Sections III and IV. Since the derivation of these bounds follow along the same lines as in the previous sections, we simply state results in Table I.

We see that the lower bounds on $\hat{\boldsymbol{\gamma}}(\mathbf{y})$ and $\hat{\xi}(\mathbf{y})$ are reduced by a factor of M compared to the SMV case, which is intuitively satisfying. It turns out that it is not possible to obtain the MCRB on $\mathbf{w}$ in the MMV-SBL setting, since closed form expressions for the FIM are not available.

In the next section, we consider two popular algorithms for SBL and graphically illustrate the utility of the lower bounds.

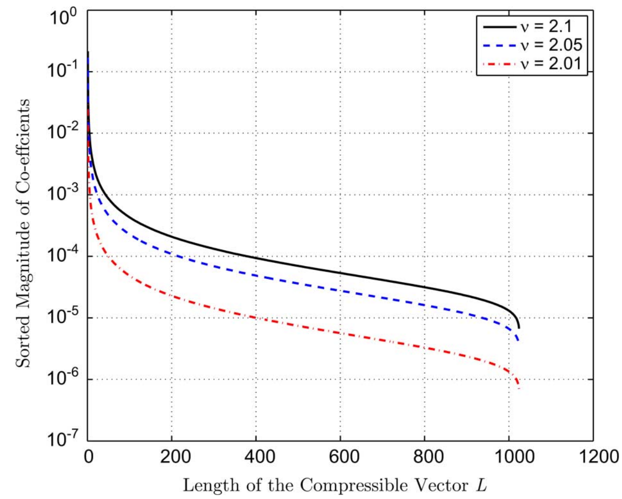

Fig. 5. Decay profile of the sorted magnitudes of *i.i.d.* samples drawn from a Student-t distribution.

VI. SIMULATIONS AND DISCUSSION

The vector estimation problem in the SBL framework typically involves the joint estimation of the hyperparameter and the unknown compressible vector $\mathbf{x}$. Since the hyperparameter estimation problem cannot be solved in closed form, iterative estimators are employed [4]. In this section, we consider the iterative updates based on the EM algorithm first proposed in [4]. We also consider the algorithm proposed in [6] based on the Automatic Relevance Determination (ARD) framework. We plot the MSE performance in estimating $\mathbf{x}$, $\boldsymbol{\gamma}$ and ξ with the linear model in (1) and (2), for the EM algorithm, labeled $\texttt{EM}$, and the ARD based Reweighted ℓ_1 algorithm, labeled $\texttt{ARD} - \texttt{SBL}$. We compare the performance of the estimators against the derived lower bounds.

We simulate the lower bounds for a random underdetermined $(N < L)$ measurement matrix $\boldsymbol{\Phi}$, whose entries are *i.i.d.* and standard Bernoulli $(\{+1, -1\})$ distributed. A compressible signal of dimension L is generated by sampling from a Student-t distribution with the value of ν ranging from 2.01 to 2.05, which is the range in which the signal is "*somewhat*" compressible, for high dimensional signals [3]. Fig. 5 shows the decay profile of the sorted magnitudes of $L = 1024$ *i.i.d.* samples drawn from a Student-t distribution for different ν and with the value of $\mathbb{E}(x_i^2)$ fixed at 10^{-3}.

A. Lower Bounds on the MSE Performance of $\hat{\mathbf{x}}(\mathbf{y})$

In this subsection, we compare the MSE performance of the ARD-SBL estimator and the EM based estimator $\hat{\mathbf{x}}(\mathbf{y})$. Fig. 6 depicts the MSE performance of $\hat{\mathbf{x}}(\mathbf{y})$ for different SNRs and $N = 750$ and 1000, with $\nu = 2.01$. We compare it with the HCRB/BCRB derived in (8), which is obtained by assuming the knowledge of the realization of the hyperparameters $\boldsymbol{\gamma}$. We see

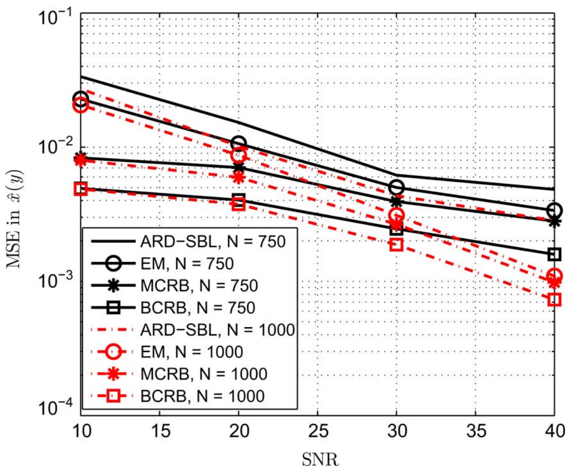

Fig. 6. The MSE performance of $\hat{\mathbf{x}}(\mathbf{y})$ and the corresponding MCRB and BCRB, as a function of SNR, with $\nu = 2.01$.

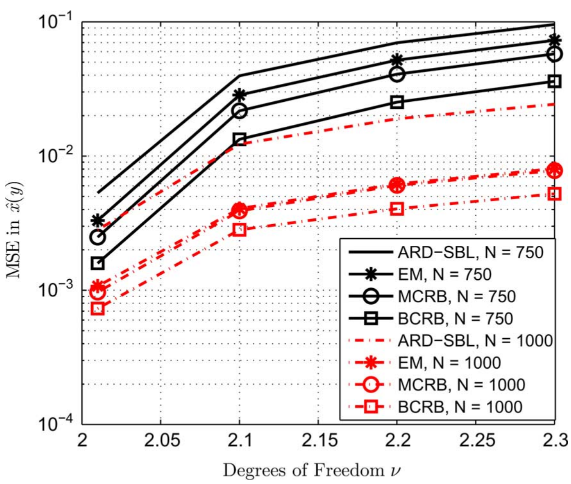

Fig. 7. The MSE performance of $\hat{\mathbf{x}}(\mathbf{y})$ and the corresponding MCRB and BCRB, as a function of ν, with $\mathrm{SNR} = 40$ dB.

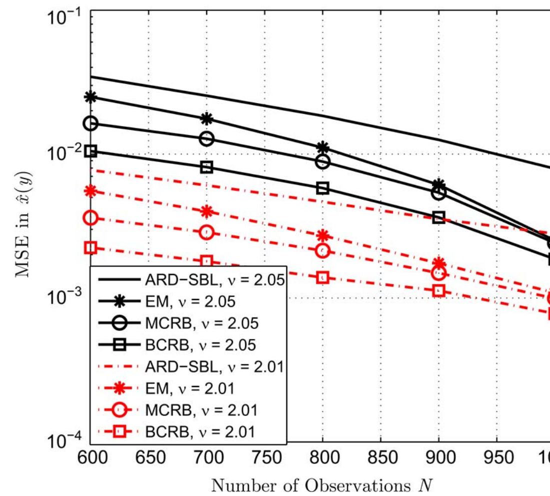

Fig. 8. The MSE performance of $\hat{\mathbf{x}}(\mathbf{y})$ and the corresponding MCRB and BCRB, as a function of N, with $\mathrm{SNR} = 40$ dB.

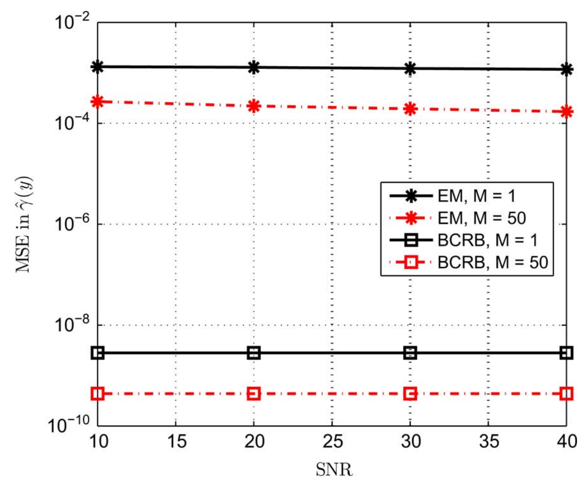

Fig. 9. The MSE performance of $\hat{\boldsymbol{\gamma}}(\mathbf{y})$ and the corresponding HCRB, as a function of SNR, with $N = 1000$.

that the MCRB derived in (14) is a tight lower bound on the MSE performance at high SNR and N.

Fig. 7 shows the comparative MSE performance of the ARD-SBL estimator and EM based estimator as a function of varying degrees of freedom ν, at an SNR of 40 dB and $N = 1000$ and 750. As expected, the MSE performance of the algorithms is better at low values of ν since the signal is more compressible, and the MCRB and BCRB also reflect this behavior. The MCRB is a tight lower bound, especially for high values of N. Fig. 8 shows the MSE performance of the ARD-SBL estimator and EM based estimator as a function of N, at an SNR of 40 dB and for two different values of ν. The MSE performance of the EM algorithm converges to that of the MCRB at higher N.

B. Lower Bounds on the MSE Performance of $\hat{\boldsymbol{\gamma}}(\mathbf{y})$

In this subsection, we compare the different lower bounds for the MSE of the estimator $\hat{\boldsymbol{\gamma}}(\mathbf{y})$ for the SMV and MMV-SBL system model. Fig. 9 shows the MSE performance of $\hat{\boldsymbol{\gamma}}(\mathbf{y})$ as a function of SNR and M, when γ is a random parameter, $N = 1000$ and $\nu = 2.01$. In this case, it turns out that there is a large gap between the performance of the EM based estimate and the lower bound.

When $\boldsymbol{\gamma}$ is deterministic, we first note that the EM based ML estimator for $\boldsymbol{\gamma}$ is asymptotically optimal and the lower bounds are practical for large data samples [19]. The results are listed in Table II. We see that for $L = 2048$ and $N = 1500$, the MCRB and BCRB are tight lower bounds, with MCRB being marginally tighter than the BCRB. However, as M increases, the gap between the MSE and the lower bounds increases.

C. Lower Bounds on the MSE Performance of $\hat{\xi}(\mathbf{y})$

In Fig. 10, we compare the lower bounds on the MSE of the estimator $\hat{\xi}(\mathbf{y})$ in the SMV and MMV-SBL settings, for different values of N and M. Here, ξ is sampled from the IG pdf (18), with parameters $c = 3$ and $d = 0.2$.

When ξ is deterministic, the EM based ML estimator for $\boldsymbol{\gamma}$ is asymptotically optimal and the lower bounds are practical for large data samples [19]. Table III lists the MSE values of $\hat{\xi}(\mathbf{y})$, the corresponding HCRB and MCRB for deterministic but unknown noise variance, while the true noise variance is fixed at 10^{-3}. We see that for $L = 2048$ and $N = 1500$, the MCRB

TABLE II
VALUES OF THE MSE OF THE ESTIMATOR $\hat{\gamma}(\mathbf{y})$, THE MCRB AND THE BCRB, FOR $\boldsymbol{\theta}_d = [\gamma]$ AS A FUNCTION OF SNR, FOR $N = 1500$.

SNR(dB)		10	20	30	40
$M = 1$	MSE	0.054	0.053	0.051	0.050
	MCRB	0.052	0.051	0.050	0.049
	BCRB	0.049	0.049	0.049	0.049
$M = 50$	MSE	0.0450	0.039	0.035	0.030
	MCRB $\times 10^{-2}$	0.12	0.11	0.10	0.09
	BCRB$\times 10^{-3}$	0.977	0.977	0.977	0.977

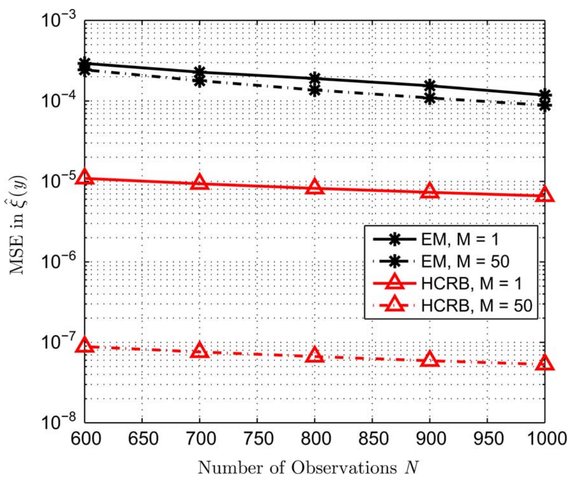

Fig. 10. The MSE performance of $\hat{\xi}(\mathbf{y})$ and its HCRB, as a function of N.

TABLE III
VALUES OF THE MSE OF THE ESTIMATOR $\hat{\xi}(\mathbf{y})$, THE MCRB AND THE HCRB FOR $\theta_d = [\xi]$, AS A FUNCTION OF N.

	N	1500	1600	1700	1800
$M = 1$	MSE $\times 10^{-8}$	0.736	0.663	0.636	0.592
	MCRB$\times 10^{-8}$	0.380	0.340	0.307	0.279
	HCRB$\times 10^{-8}$	0.133	0.125	0.118	0.111
$M = 50$	MSE $\times 10^{-9}$	0.930	0.892	0.866	0.847
	MCRB$\times 10^{-10}$	0.680	0.652	0.614	0.573
	HCRB$\times 10^{-10}$	0.267	0.250	0.235	0.222

is marginally tighter than the HCRB. However, when the noise variance is random, we see from Fig. 10 that there is a large gap between the MSE performance and the HCRB.

VII. CONCLUSION

In this work, we derived Cramér-Rao type lower bounds on the MSE, namely, the HCRB, BCRB and MCRB, for the SMV-SBL and the MMV-SBL problem of estimating compressible signals. We used a hierarchical model for the compressible priors to obtain the bounds under various assumptions on the unknown parameters. The bounds derived by assuming a hyperprior distribution on the hyperparameters themselves provided key insights into the MSE performance of SBL and the values of the parameters that govern these hyperpriors. We derived the MCRB for the generalized compressible prior distribution, of which the Student-t and Generalized Pareto prior distribution are special cases. We showed that the MCRB is tighter than the BCRB. We compared the lower bounds with the MSE performance of the ARD-SBL and the EM algorithm using Monte Carlo simulations. The numerical results illustrated the near-optimality of EM based updates for SBL, which makes it attractive for practical implementations.

APPENDIX

Proof of Proposition 1: Using the graphical model of Fig. 1 in (5),

$$\begin{aligned}\mathbf{H}^{\boldsymbol{\theta}}(\mathbf{x}) &\triangleq -\mathbb{E}_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}\left[\nabla^2_{\mathbf{x}} \log p_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}(\mathbf{y},\mathbf{x};\boldsymbol{\gamma})\right] \\ &= -\mathbb{E}_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}\left[\nabla_{\mathbf{x}}\left(\frac{\boldsymbol{\Phi}^T(\mathbf{y}-\boldsymbol{\Phi}\mathbf{x})}{\sigma^2} - \boldsymbol{\Upsilon}^{-1}\mathbf{x}\right)\right] \\ &= \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \boldsymbol{\Upsilon}^{-1}. \end{aligned} \tag{23}$$

Similarly, it is straightforward to show that $\nabla_{\mathbf{x}}\nabla_{\boldsymbol{\gamma}} \log p_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}(\mathbf{y},\mathbf{x};\boldsymbol{\gamma}) = \operatorname{diag}\left(\frac{x_1}{\gamma_1^2}, \frac{x_2}{\gamma_2^2}, \ldots, \frac{x_L}{\gamma_L^2}\right)$. Since x_i are zero mean random variables,

$$\begin{aligned}\mathbf{H}^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\mathbf{x}) &= -\mathbb{E}_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}\left[\nabla_{\boldsymbol{\gamma}}\nabla_{\mathbf{x}} \log p_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}}(\mathbf{y},\mathbf{x};\boldsymbol{\gamma})\right] \\ &= \mathbf{0}_{L\times L}, \\ \mathbf{H}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) &= -\mathbb{E}_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma}} \\ &\quad \times \left[\nabla^2_{\boldsymbol{\gamma}}(\log p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) + \log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\gamma}))\right].\end{aligned}$$

Now, since $\log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\gamma}) = \sum_{i=1}^{L} \log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x}_i;\boldsymbol{\gamma}_i)$, we get,

$$\frac{\partial^2 \log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\gamma})}{\partial\gamma_i\partial\gamma_j} = \begin{cases}\frac{1}{2\gamma_i^2} - \frac{x_i^2}{\gamma_i^3} & \text{if } i=j \\ 0 & \text{if } i\neq j.\end{cases} \tag{24}$$

Taking $-\mathbb{E}_{\mathbf{X};\boldsymbol{\gamma}}(\cdot)$ on both sides of the above equation and noting that $\mathbb{E}_{\mathbf{X};\boldsymbol{\gamma}}(x_i^2) = \gamma_i$, we obtain

$$\begin{aligned}\mathbf{H}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) &= \operatorname{diag}\left(-\mathbb{E}_{\mathbf{X};\boldsymbol{\gamma}}\left[\frac{\partial^2 \log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\gamma})}{\partial\gamma_i^2}\right]\right) \\ &= \operatorname{diag}\left(\left[\frac{1}{2\gamma_1^2}, \ldots, \frac{1}{2\gamma_L^2}\right]\right).\end{aligned} \tag{25}$$

This completes the proof.

Proof of Proposition 2: Using the graphical model of Fig. 1 in (5),

$$\begin{aligned}\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}) &\triangleq -\mathbb{E}_{\mathbf{Y},\mathbf{X},\Gamma}\left[\nabla^2_{\mathbf{x}} \log p_{\mathbf{Y},\mathbf{X},\Gamma}(\mathbf{y},\mathbf{x},\boldsymbol{\gamma})\right] \\ &= -\mathbb{E}_{\mathbf{Y},\mathbf{X},\Gamma}\left[\nabla_{\mathbf{x}}\left(\frac{\boldsymbol{\Phi}^T(\mathbf{y}-\boldsymbol{\Phi}\mathbf{x})}{\sigma^2} - \boldsymbol{\Upsilon}^{-1}\mathbf{x}\right)\right] \\ &= \mathbb{E}_{\Gamma}\left[\frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \boldsymbol{\Upsilon}^{-1}\right] \qquad (26) \\ &= \frac{\boldsymbol{\Phi}^T\boldsymbol{\Phi}}{\sigma^2} + \mathbb{E}_{\Gamma}\left[\boldsymbol{\Upsilon}^{-1}\right]. \qquad (27)\end{aligned}$$

The expression for $\mathbb{E}_{\Gamma}\left[\boldsymbol{\Upsilon}^{-1}\right]$ w.r.t. γ_i is given by,

$$\begin{aligned}\mathbb{E}_{\Gamma}\left[\frac{1}{\gamma_i}\right] &= K_\gamma \int_{\gamma_i=0}^{\infty} \gamma_i^{\left(-\frac{\nu}{2}-2\right)} \exp\left\{-\frac{\nu}{2\lambda\gamma_i}\right\} d\gamma_i \qquad (28) \\ &= K_\gamma \frac{\Gamma\left(\frac{\nu}{2}+1\right)}{\left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}+1}} \underbrace{\int_{\gamma_i=0}^{\infty} \mathcal{IG}\left(\frac{\nu}{2}+1, \frac{\nu}{2\lambda}\right) d\gamma_i}_{=1} \\ &= \lambda, \qquad (29)\end{aligned}$$

since $K_\gamma = \left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}}\left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1}$. Hence, the overall bound is given by

$$\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x}) = \frac{\mathbf{\Phi}^T\mathbf{\Phi}}{\sigma^2} + \lambda \mathbf{I}_{L\times L}. \quad (30)$$

Using the graphical model of Fig. 1 in (5), for $\boldsymbol{\theta} = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$, $\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ is defined as

$$\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma}) \triangleq -\mathbb{E}_{\mathbf{Y},\mathbf{X},\Gamma}\left[\nabla^2_{\boldsymbol{\gamma}}\left(\log p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) + \log p_{\mathbf{X}|\Gamma}(\mathbf{x}|\boldsymbol{\gamma}) + \log p_\Gamma(\boldsymbol{\gamma})\right)\right]. \quad (31)$$

Since the expressions for $\log p_{\mathbf{X}|\Gamma}(\mathbf{x}|\boldsymbol{\gamma})$ and $\log p_\Gamma(\boldsymbol{\gamma})$ are separable and symmetric w.r.t. γ_i, the off-diagonal terms of $\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma})$ are zero, and it is sufficient to evaluate the diagonal terms $-\mathbb{E}_{\mathbf{Y},\mathbf{X},\Gamma}\left(\frac{\partial^2(\log p_{\mathbf{X}|\Gamma}(\mathbf{x}|\boldsymbol{\gamma})+\log p_\Gamma(\boldsymbol{\gamma}))}{\partial\gamma_i^2}\right)$. Differentiating the expression w.r.t. γ_i twice,

$$\frac{\partial^2\left(\log p_{\mathbf{X}|\Gamma}(\mathbf{x}|\boldsymbol{\gamma}) + \log p_\Gamma(\boldsymbol{\gamma})\right)}{\partial\gamma_i^2} = -\frac{(\nu+1)}{2\gamma_i^2} + \frac{\nu}{\lambda\gamma_i^3}. \quad (32)$$

The expression for $-\mathbb{E}_\Gamma\left[-\frac{(\nu+1)}{2\gamma_i^2} + \frac{\nu}{\lambda\gamma_i^3}\right]$ is given by

$$\mathbb{E}_\Gamma\left[\frac{(\nu+1)}{2\gamma_i^2} - \frac{\nu}{\lambda\gamma_i^3}\right] = K_\gamma \int_{\gamma_i=0}^{\infty}\left[\frac{(\nu+1)\gamma_i^{-2}}{2} - \frac{\nu\gamma_i^{-3}}{\lambda}\right]\gamma_i^{\left(-\frac{\nu}{2}-1\right)}\exp\left\{-\frac{\nu}{2\lambda\gamma_i}\right\}\mathrm{d}\gamma_i, \quad (33)$$

where $K_\gamma = \left(\frac{\nu}{2\lambda}\right)^{\frac{\nu}{2}}\left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1}$. After some manipulation, it can be shown that the above integral reduces to

$$-\mathbb{E}_\Gamma\left[-\frac{(\nu+1)}{2\gamma_i^2} + \frac{\nu}{\lambda\gamma_i^3}\right] = \frac{\lambda^2(\nu+2)(\nu+7)}{2\nu}. \quad (34)$$

Thus, the $(ij)^{\text{th}}$ component of $\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\mathbf{x})$ is given by

$$(\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\mathbf{x}))_{ij} = \frac{\partial^2 \log p_{\mathbf{X}|\Gamma}(\mathbf{x}|\boldsymbol{\gamma})}{\partial\gamma_i\partial x_i} = -\frac{x_i}{\gamma_i^2}, \quad (35)$$

and $\mathbf{B}^{\boldsymbol{\theta}}(\mathbf{x},\boldsymbol{\gamma}) = (\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\mathbf{x}))^T$. Since $\mathbb{E}_{\mathbf{X}|\Gamma}(x_i) = 0$, $\mathbf{B}^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\mathbf{x}) = \mathbf{0}_{L\times L}$. This completes the proof.

Proof of Theorem 1: To establish the regularity condition, the first order derivative of the log likelihood $\log p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})$ is required. This, in turn, requires the evaluation of $\frac{\partial \log|\mathbf{\Sigma}_y|}{\partial\gamma_j}$ and $\frac{\partial \mathbf{y}^T\mathbf{\Sigma}_y^{-1}\mathbf{y}}{\partial\gamma_j}$. Using the chain rule for differentiation [26], we have

$$\frac{\partial\log|\mathbf{\Sigma}_y|}{\partial\gamma_j} = \mathrm{Tr}\left\{\left(\frac{\partial\log|\mathbf{\Sigma}_y|}{\partial\mathbf{\Sigma}_y}\right)^T\frac{\partial\mathbf{\Sigma}_y}{\partial\gamma_j}\right\} = \mathrm{Tr}\left\{(\mathbf{\Sigma}_y^{-1})^T\Phi_j\Phi_j^T\right\} = \Phi_j^T\mathbf{\Sigma}_y^{-1}\Phi_j. \quad (36)$$

Here, we have used the identity $\nabla_X \log|X| = X^{-1}$ [26] and results from vector calculus [26] to obtain $\frac{\partial\mathbf{\Sigma}_y}{\partial\gamma_j} = \Phi_j\Phi_j^T$, where Φ_j is the j^{th} column of $\mathbf{\Phi}$. Similarly, the derivative of $\mathbf{y}^T\mathbf{\Sigma}_y^{-1}\mathbf{y}$ can be obtained as

$$\frac{\partial\mathbf{y}^T\mathbf{\Sigma}_y^{-1}\mathbf{y}}{\partial\gamma_j} = \mathrm{Tr}\left\{\left(\frac{\partial\mathbf{y}^T\mathbf{\Sigma}_y^{-1}\mathbf{y}}{\partial\mathbf{\Sigma}_y^{-1}}\right)^T\frac{\partial\mathbf{\Sigma}_y^{-1}}{\partial\gamma_j}\right\} = -\Phi_j^T\mathbf{\Sigma}_y^{-1}\mathbf{y}\mathbf{y}^T\mathbf{\Sigma}_y^{-1}\Phi_j, \quad (37)$$

and hence,

$$\frac{\partial}{\partial\gamma_j}\log p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma}) = \frac{\Phi_j^T\mathbf{\Sigma}_y^{-1}\mathbf{y}\mathbf{y}^T\mathbf{\Sigma}_y^{-1}\Phi_j - \Phi_j^T\mathbf{\Sigma}_y^{-1}\Phi_j}{2}. \quad (38)$$

Taking $\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma}}(\cdot)$ on both the sides of the above equation,

$$\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma}}\left[\frac{\partial}{\partial\gamma_j}\log p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})\right] = \frac{\Phi_j^T\mathbf{\Sigma}_y^{-1}\left\{\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y}\mathbf{y}^T)\right\}\mathbf{\Sigma}_y^{-1}\Phi_j - \Phi_j^T\mathbf{\Sigma}_y^{-1}\Phi_j}{2} = 0, \quad (39)$$

since $\mathbb{E}_{\mathbf{Y}}(\mathbf{y}\mathbf{y}^T) = \mathbf{\Sigma}_y$. Hence, the pdf satisfies the required regularity constraint.

Now, the MCRB for $\boldsymbol{\theta} = [\boldsymbol{\gamma}]$ is obtained by computing the second derivative of the log likelihood, as follows:

$$\begin{aligned} -&\frac{\partial^2}{\partial\gamma_i\partial\gamma_j}\log p_{\mathbf{Y},\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma}) \\ &= \frac{1}{2}\frac{\partial}{\partial\gamma_i}(\Phi_j^T\mathbf{\Sigma}_y^{-1}\Phi_j - (\Phi_j^T\mathbf{\Sigma}_y^{-1}\mathbf{y})^2) \\ &= \frac{1}{2}\mathrm{Tr}\left\{\Phi_j\Phi_j^T(-\mathbf{\Sigma}_y^{-1}\Phi_i\Phi_i^T\mathbf{\Sigma}_y^{-1})\right\} \\ &\quad - (\Phi_j^T\mathbf{\Sigma}_y^{-1}\mathbf{y})\mathrm{Tr}\left\{\left(\frac{\partial(\Phi_j^T\mathbf{\Sigma}_y^{-1}\mathbf{y})}{\partial\mathbf{\Sigma}_y^{-1}}\right)^T\frac{\partial\mathbf{\Sigma}_y^{-1}}{\partial\gamma_i}\right\} \\ &= -\frac{1}{2}\left(\Phi_j^T\mathbf{\Sigma}_y^{-1}\Phi_i\right)\left(\Phi_i^T\mathbf{\Sigma}_y^{-1}\Phi_j\right) \\ &\quad + \left(\Phi_j^T\mathbf{\Sigma}_y^{-1}\mathbf{y}\right)\left(\mathbf{y}^T\mathbf{\Sigma}_y^{-1}\Phi_i\right)\left(\Phi_i^T\mathbf{\Sigma}_y^{-1}\Phi_j\right). \end{aligned} \quad (40)$$

Taking $-\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma}}(\cdot)$ on both the sides of the above expression,

$$(\mathbf{M}^{\boldsymbol{\gamma}})_{ij} \triangleq -\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma}}\left[\frac{\partial^2\log p_{\mathbf{Y};\boldsymbol{\gamma}}(\mathbf{y};\boldsymbol{\gamma})}{\partial\gamma_i\partial\gamma_j}\right] = \frac{(\Phi_j^T\mathbf{\Sigma}_y^{-1}\Phi_i)^2}{2}, \quad (41)$$

as stated in (11). This completes the proof.

Proof of Theorem 2: The proof follows from the proof for Theorem 3 in Appendix E by substituting $\tau = 2$.

Proof of Theorem 3: The MCRB for estimation of the compressible random vector with $\boldsymbol{\theta} = [\mathbf{x}]$ is given by

$$\begin{aligned} \mathbf{M}^{\mathbf{x}} &= -\mathbb{E}_{\mathbf{Y},\mathbf{X}}[\nabla^2_{\mathbf{x}}\log p_{\mathbf{Y},\mathbf{X}}(\mathbf{y},\mathbf{x})] \\ &= -\mathbb{E}_{\mathbf{Y},\mathbf{X}}[\nabla^2_{\mathbf{x}}\log p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) + \nabla^2_{\mathbf{x}}\log p_{\mathbf{X}}(\mathbf{x})]. \end{aligned} \quad (42)$$

The first term above is given by

$$\begin{aligned} -\mathbb{E}_{\mathbf{Y},\mathbf{X}}\left[\nabla^2_{\mathbf{x}}\log p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})\right] &= -\mathbb{E}_{\mathbf{Y},\mathbf{X}}\left[\nabla_{\mathbf{x}}\frac{\mathbf{\Phi}^T(\mathbf{y}-\mathbf{\Phi}\mathbf{x})}{\sigma^2}\right] \\ &= -\mathbb{E}_{\mathbf{Y},\mathbf{X}}\left[\frac{-\mathbf{\Phi}^T\mathbf{\Phi}}{\sigma^2}\right] = \frac{\mathbf{\Phi}^T\mathbf{\Phi}}{\sigma^2}. \end{aligned} \quad (43)$$

Note that $p_{\mathbf{X}}(\mathbf{x})$ is not differentiable if any of its components $x_i = 0$. However, the measure of $x_i = 0$ is zero since the distribution is continuous, and hence, this condition can be safely ignored. Now,

$$\frac{\partial\log p_{\mathbf{X}}(\mathbf{x})}{\partial x_i} = \begin{cases} -\frac{(\nu+1)\lambda x_i^{\tau-1}}{(\nu+\lambda x_i^\tau)} & \text{if } x_i > 0 \\ (-1)^\tau\frac{(\nu+1)\lambda x_i^{\tau-1}}{(\nu+(-1)^\tau\lambda x_i^\tau)} & \text{if } x_i < 0. \end{cases}$$

First, we consider the case of $x_i > 0$. Differentiating the above w.r.t. x_i again, we obtain

$$\frac{\partial^2}{\partial x_i^2}\log p_{\mathbf{X}}(\mathbf{x}) = \frac{-(\nu+1)\lambda(\tau-1)x_i^{\tau-2}}{(\nu+\lambda x_i^\tau)} + \frac{\lambda^2\tau(\nu+1)x_i^{2\tau-2}}{(\nu+\lambda x_i^\tau)^2}. \quad (44)$$

Taking $-\mathbb{E}_{\mathbf{X}}(\cdot)$ on both sides of the above equation, we get

$$-\mathbb{E}_{\mathbf{X}}\left(\frac{\partial^2}{\partial x_i^2}\log p_{\mathbf{X}}(\mathbf{x})\right) = \frac{K(\nu+1)\lambda}{\nu} \int_0^\infty \left(\frac{(\tau-1)x_i^{\tau-2}}{\left(1+\frac{\lambda x_i^\tau}{\nu}\right)^{\frac{\nu+\tau+1}{\tau}}} - \frac{\lambda\tau x_i^{2\tau-2}}{\nu\left(1+\frac{\lambda x_i^\tau}{\nu}\right)^{\frac{\nu+2\tau+1}{\tau}}}\right)\mathrm{d}x_i. \quad (45)$$

The above can be simplified using the transformation $t_i = \frac{\lambda x_i^\tau}{\nu}$ and using $\int_0^\infty \frac{t^{u-1}}{(1+t)^{u+v}}\mathrm{d}t = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$, we get

$$-\mathbb{E}_{\mathbf{X}}\left(\frac{\partial^2}{\partial x_i^2}\log p_{\mathbf{X}}(\mathbf{x})\right) = \frac{K(\nu+1)(\tau-1)}{\tau}\left(\frac{\lambda}{\nu}\right)^{\frac{1}{\tau}} \Gamma\left(1-\frac{1}{\tau}\right)\left\{\frac{\Gamma\left(\frac{\nu+\tau+2}{\tau}\right)-\frac{1}{\tau}\Gamma\left(\frac{\nu+2}{\tau}\right)}{\Gamma\left(\frac{v+2\tau+1}{\tau}\right)}\right\} \quad \text{for} \quad x_i > 0. \quad (46)$$

For the case of $x_i < 0$ also, the expression reduces to the integral given in (45). Hence, we have

$$-\mathbb{E}_{\mathbf{X}}\left(\frac{\partial^2}{\partial x_i^2}\log p_{\mathbf{X}}(\mathbf{x})\right) = \frac{K(\nu+1)^2(\tau-1)}{\tau(\nu+\tau+1)}\left(\frac{\lambda}{\nu}\right)^{\frac{1}{\tau}} \left(\frac{\Gamma\left(\frac{\tau-1}{\tau}\right)\Gamma\left(\frac{\nu+2}{\tau}\right)}{\Gamma\left(\frac{v+\tau+1}{\tau}\right)}\right). \quad (47)$$

Substituting the expression for K in the above, we get

$$-\mathbb{E}_{\mathbf{X}}\left(\frac{\partial^2}{\partial x_i^2}\log p_{\mathbf{X}}(\mathbf{x})\right) = \frac{\tau^2(\nu+1)}{(\nu+\tau+1)}\left(\frac{\lambda}{\nu}\right)^{\frac{2}{\tau}} \frac{\Gamma\left(\frac{\nu+2}{\tau}\right)\Gamma\left(2-\frac{1}{\tau}\right)}{\Gamma\left(\frac{1}{\tau}\right)\Gamma\left(\frac{v}{\tau}\right)}. \quad (48)$$

Combining the expression above and (43), we obtain the MCRB in (17).

Proof of Proposition 3: In this case, we define $\boldsymbol{\theta}' = [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$ and hence, $\boldsymbol{\theta} = [\boldsymbol{\theta}'^T, \xi]^T$. In order to compute the HCRB, we need to find $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\xi)$, $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\theta}')$ and $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\theta}', \xi)$. We have $\log p_{\mathbf{Y},\mathbf{X};\boldsymbol{\gamma},\xi}(\mathbf{y},\mathbf{x};\boldsymbol{\gamma},\xi) = \log p_{\mathbf{Y}|\mathbf{X};\xi}(\mathbf{y}|\mathbf{x};\xi) + \log p_{\mathbf{X};\boldsymbol{\gamma}}(\mathbf{x};\boldsymbol{\gamma})$, where $\xi = \sigma^2$. Using (5), the submatrix $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\theta}') = \mathbf{H}^{\boldsymbol{\theta}'}$, i.e., the same as computed earlier in (8) when $\boldsymbol{\gamma}$ is unknown deterministic and by (10) when $\boldsymbol{\gamma}$ is random. Hence, we focus on the block matrices that occur due to the additional parameter ξ. First, $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\xi)$ is computed as in Sec. 3.6 in [19], from which, $-\mathbb{E}_{\mathbf{Y},\mathbf{X};\xi}\left[-\frac{N}{2\xi^2}\right] = \frac{N}{2\xi^2}$.

From Lemma 1, it directly follows that $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\xi) = \mathbf{0}_{L\times 1}$. Using (5), we compute $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\mathbf{x},\xi)$ as follows:

$$\mathbf{H}_\xi^{\boldsymbol{\theta}}(\mathbf{x},\xi) = \mathbb{E}_{\mathbf{X}}(\mathbb{E}_{\mathbf{Y}|\mathbf{X};\xi}(\boldsymbol{\Phi}^T\mathbf{y} - \boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{x})). \quad (49)$$

Since $\mathbb{E}_{\mathbf{Y}|\mathbf{X};\xi}(\mathbf{y}) = \boldsymbol{\Phi}\mathbf{x}$, $\mathbb{E}_{\mathbf{X}}(\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\mathbf{x}) - \boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{x}) = \mathbf{0}_{L\times 1}$. This completes the proof.

Proof of Proposition 4: In this case, we define $\boldsymbol{\theta} \triangleq [\boldsymbol{\theta}'^T, \xi]$ and $\boldsymbol{\theta}' \triangleq [\mathbf{x}^T, \boldsymbol{\gamma}^T]^T$. In order to compute the HCRB, we need to find $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\xi)$, $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\theta}')$ and $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\theta}',\xi)$. Using (5), the expression for $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\theta}')$ is the same as computed earlier in (8) when $\boldsymbol{\gamma}$ is unknown deterministic and by (10) when $\boldsymbol{\gamma}$ is random. Since ξ is random, the expectation has to be taken over the distribution of ξ also, and hence,

$$\mathbf{H}_\xi^{\boldsymbol{\theta}}(\xi) = -\mathbb{E}_{\mathbf{Y},\mathbf{X},\Xi}\left[\frac{\partial^2}{\partial\xi^2}(\log p_{\mathbf{Y}|\mathbf{X},\Xi}(\mathbf{y}|\mathbf{x},\xi) + \log p_\Xi(\xi))\right] = \mathbb{E}_\Xi\left(\frac{\frac{N}{2}-c-1}{\xi^2} + \frac{2d}{\xi^3}\right). \quad (50)$$

The above expectation is evaluated as

$$\mathbf{H}_\xi^{\boldsymbol{\theta}}(\xi) = \frac{\left(\frac{N}{2}-c-1\right)d^c}{\Gamma(c)}\int_{\xi=0}^{\infty}\xi^{-2}\xi^{(-c-1)}\exp\left\{-\frac{d}{\xi}\right\}\mathrm{d}\xi + \frac{2d^{(c+1)}}{\Gamma(c)}\int_{\xi=0}^{\infty}\xi^{-3}\xi^{(-c-1)}\exp\left\{-\frac{d}{\xi}\right\}\mathrm{d}\xi = \frac{c(c+1)\left(\frac{N}{2}+c+3\right)}{d^2}. \quad (51)$$

To find the other components of the matrix, we compute $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\theta}',\xi) = (\mathbf{H}_\xi^{\boldsymbol{\theta}}(\xi,\boldsymbol{\theta}'))^T$, which consists of $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\xi)$ and $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\mathbf{x},\xi)$. From Lemma 1, $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma},\xi) = \mathbf{0}_{L\times 1}$. Using the definition of $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\mathbf{x},\xi)$, from (49) and since $p_\Xi(\xi)$ is not a function of x_i, we see that $\mathbf{H}_\xi^{\boldsymbol{\theta}}(\mathbf{x},\xi) = (\mathbf{H}_\xi^{\boldsymbol{\theta}}(\xi,\mathbf{x}))^T = \mathbf{0}_{L\times 1}$. Thus, we obtain the FIM given by (21).

Proof of Theorem 4: First, we show that the log likelihood $\log(p_{\mathbf{Y};\boldsymbol{\gamma},\xi}(\mathbf{y};\boldsymbol{\gamma},\xi))$ in (3) satisfies the regularity condition w.r.t. ξ. Differentiating the log likelihood w.r.t. ξ and taking $-\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma},\xi}(\cdot)$ on both the sides of the equation,

$$\frac{\partial}{\partial\xi}\log(p_{\mathbf{Y};\boldsymbol{\gamma},\xi}(\mathbf{y},\boldsymbol{\gamma},\xi)) = \frac{1}{2}\frac{\partial}{\partial\xi}(-\log|\boldsymbol{\Sigma}_y| - \mathbf{y}^T\boldsymbol{\Sigma}_y^{-1}\mathbf{y}) = -\frac{1}{2}\left[\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-1}) - \mathrm{Tr}(\mathbf{y}\mathbf{y}^T(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_y^{-1}))\right], \quad (52)$$

$$\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma},\xi}\left[\mathrm{Tr}(-\frac{1}{2}\boldsymbol{\Sigma}_y^{-1}) + \frac{1}{2}\mathrm{Tr}(\mathbf{y}\mathbf{y}^T(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_y^{-1}))\right] = \frac{1}{2}\left[\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-1}) - \mathrm{Tr}(\boldsymbol{\Sigma}_y^{-1})\right] = 0. \quad (53)$$

Hence, the regularity condition is satisfied. From (41), we have $(\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma}))_{ij} = -\frac{(\Phi_j^T\boldsymbol{\Sigma}_y^{-1}\Phi_i)^2}{2}$. To obtain $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi)$, we differentiate (52) w.r.t. ξ to obtain

$$\frac{\partial^2}{\partial\xi^2}(\log p_{\mathbf{Y};\boldsymbol{\gamma},\xi}(\mathbf{y};\boldsymbol{\gamma},\xi)) = \frac{1}{2}\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-2}) - \mathrm{Tr}(\mathbf{y}\mathbf{y}^T(\boldsymbol{\Sigma}_y^{-3})). \quad (54)$$

Taking $-\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma},\xi}(\cdot)$ on both sides of the above equation,

$$\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi) = -\mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma},\xi}\left[\frac{1}{2}\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-2}) - \mathrm{Tr}(\mathbf{y}\mathbf{y}^T\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-3}))\right] = \mathrm{Tr}(\boldsymbol{\Sigma}_y^{-2}) - \frac{1}{2}\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-2}) = \frac{1}{2}\mathrm{Tr}(\boldsymbol{\Sigma}_y^{-2}). \quad (55)$$

The vector $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma}, \xi)$ is found by differentiating (38) w.r.t. ξ and taking the negative expectation:

$$\begin{aligned}(\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma}, \xi))_i &= \mathbb{E}_{\mathbf{Y};\boldsymbol{\gamma},\xi}\left[\frac{\partial}{\partial \xi}\left(\frac{\Phi_i^T \boldsymbol{\Sigma}_y^{-1}\Phi_i - \Phi_i^T \boldsymbol{\Sigma}_y^{-1}\mathbf{y}\mathbf{y}^T\boldsymbol{\Sigma}_y^{-1}\Phi_i}{2}\right)\right] \\ &= \frac{1}{2}\Phi_i^T \boldsymbol{\Sigma}_y^{-2}\Phi_i. \end{aligned} \tag{56}$$

Since $\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi, \boldsymbol{\gamma}) = (\mathbf{M}_\xi^{\boldsymbol{\theta}}(\boldsymbol{\gamma}, \xi))^T$, the i^{th} term of $(\mathbf{M}_\xi^{\boldsymbol{\theta}}(\xi, \boldsymbol{\gamma}))_i = \frac{1}{2}\Phi_i^T \boldsymbol{\Sigma}_y^{-2}\Phi_i$. The MCRB $\mathbf{M}_\xi^{\boldsymbol{\theta}}$ can now be obtained by combining the expressions in (41), (55) and (56); this completes the proof.

REFERENCES

[1] D. Needell and J. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.
[2] V. Cevher, "Learning with compressible priors," in *Proc. Adv. NIPS*, , pp. 7–12.
[3] R. Gribonval, V. Cevher, and M. Davies, "Compressible priors for high-dimensional statistics," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5016–5034, 2012.
[4] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–214, 2001.
[5] M. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proc. 9th Int. Workshop Artif. Intell. Statist.*, 2003, vol. 1, pp. 3–6.
[6] D. Wipf and S. Nagarajan, "A new view of automatic relevance determination," *Adv. NIPS*, vol. 20, pp. 1625–1632, 2008.
[7] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *Proc. Signal Process. Adapt. Sparse Structured Representations (SPARS)*, 2009.
[8] O. Williams, A. Blake, and R. Cipolla, "Sparse Bayesian learning for efficient visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1292–1304, 2005.
[9] O. Williams, A. Blake, and R. Cipolla, "A sparse probabilistic learning algorithm for real-time tracking," in *Proc. 9th Int. Conf. Comput. Vis.*, 2003, pp. 353–360.
[10] D. Wipf, J. Owen, H. Attias, K. Sekihara, and S. Nagarajan, "Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG," *NeuroImage*, vol. 49, no. 1, pp. 641–655, 2010.
[11] D. Wipf and S. Nagarajan, "Beamforming using the relevance vector machine," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 1023–1030.
[12] R. Prasad and C. R. Murthy, "Bayesian learning for joint sparse OFDM channel estimation and data detection," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2010, pp. 1–6.
[13] B. Babadi, N. Kalouptsidis, and V. Tarokh, "Asymptotic achievability of the Cramér-Rao bound for noisy compressive sampling," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1233–1236, 2009.
[14] Z. Ben-Haim and Y. Eldar, "The Cramér-Rao bound for estimating a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3384–3389, 2010.
[15] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "On the achievability of Cramér-Rao bound in noisy compressed sensing," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 518–526, 2012.
[16] J. Helferty and D. Mmiscudgett, "Optimal observer trajectories for bearings only tracking by minimizing the trace of the Cramér-Rao lower bound," in *Proc. 32nd Conf. Decision Control*, 1993, pp. 936–939.
[17] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, 2004.
[18] D. Wipf and B. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, 2007.
[19] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[20] H. V. Trees, *Detection, Estimation, and Modulation Theory—Part I*. New York: Wiley, 1968, vol. 2.
[21] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locations-Part I: Far-field sources," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 3, pp. 286–299, 1987.
[22] J. Dauwels, "on graphical models for communications and machine learning: Algorithms, bounds, and analog implementation," Ph.D. dissertation, Swiss Federal Inst. Technol., Zürich, Switzerland, 2005.
[23] D. Wipf, J. Palmer, and B. Rao, "Perspectives on sparse Bayesian learning," *Adv. NIPS*, vol. 16, p. 249, 2004.
[24] A. Armagan, D. Dunson, and J. Lee, "Generalized double Pareto shrinkage," *Arxiv preprint arXiv:1104.0861*, Aug. 2011 [Online]. Available: http://arxiv.org/abs/1104.0861
[25] S. Balakrishnan and D. Madigan, "Priors on the variance in sparse Bayesian learning; The Demi-Bayesian Lasso," 2009 [Online]. Available: http://www.stat.columbia.edu/~madigan/PAPERS/recent-Papers.html
[26] K. Petersen and M. Pedersen, The Matrix Cookbook, 2008 [Online]. Available: http://matrixcookbook.com

Ranjitha Prasad received the B.E. degree in electronics and communication engineering from the National Institute of Engineering, Mysore, India, in 2004, and the M.S. degree in electrical engineering from the Indian Institute of Technology Madras, Chennai, India, in 2009.

From July 2004 to 2006, she worked as a Senior Design Engineer at Tata Elxsi, Bangalore, India. She is currently working towards the Ph.D. degree at the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. Her research interests include signal processing for communications, adaptive filter theory, sparse Bayesian learning, and compressive sensing.

Chandra R. Murthy (S'03–M'06–SM'11) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology Madras, Chennai, India, in 1998 and the M.S. degree from Purdue University, West Lafayette, IN, in 2000 and the Ph.D. degree from the University of California, San Diego, in 2006, both in electrical and computer engineering.

From 2000 to 2002, he was an Engineer for Qualcomm Inc., San Jose, CA, where he worked on WCDMA baseband transceiver design and 802.11b baseband receivers. From August 2006 to August 2007, he worked as a Staff Engineer at Beceem Communications Inc., Bangalore, India, on advanced receiver architectures for the 802.16e Mobile WiMAX standard. Since September 2007, he has been an Assistant Professor at the Department of Electrical Communication Engineering at the Indian Institute of Science, Bangalore, India. His research interests are in the areas of cognitive radio, energy harvesting wireless sensors and MIMO systems with channel-state feedback.

Dr. Murthy is currently serving as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.