Estimation Performance for the Bayesian Hierarchical Linear Model

Mohammed Nabil El Korso, Remy Boyer, Pascal Larzabal, Bernard-Henri Fleury

▶ **To cite this version:**

Mohammed Nabil El Korso, Remy Boyer, Pascal Larzabal, Bernard-Henri Fleury. Estimation Performance for the Bayesian Hierarchical Linear Model. IEEE Signal Processing Letters, 2016, 23 (4), pp.488-492. 10.1109/LSP.2016.2528579 . hal-01264666

HAL Id: hal-01264666

https://hal.science/hal-01264666

Submitted on 10 Apr 2016

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

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

Estimation Performance for the Bayesian Hierarchical Linear Model

Mohammed Nabil El Korso, Rémy Boyer, Pascal Larzabal and Bernard-Henri Fleury

***Abstract*—Bayesian hierarchical modelling is a well established branch of Bayesian inference. In this letter, we derive and study the estimation performance for the Bayesian hierarchical linear model. Specifically, we consider a linear model with hierarchical priors for the involved amplitude and noise vectors. We provide closed-form expressions of the Bayesian Cramér-Rao bound (BCRB) for the following settings: (i) an arbitrary prior and hyper-prior and (ii) a Gaussian-$\mathcal{Y}$ prior for the amplitudes, while the prior of noise is a Gaussian-$\mathcal{X}$ in both cases. Gaussian-$\mathcal{X}$ means that the conditional prior given the hyper-parameter is Gaussian and $\mathcal{X}$ is the hyper-prior. For the hierarchical distribution associated with spherical invariant random variables, the BCRB has a compact closed-form expression and enjoys several interesting properties that are discussed. Finally, we provide a theoretical analysis of the statistical efficiency of the linear minimum mean square error estimator in the low and high noise variance regimes when the hyper-parameters are stochastically dominant.**

I. INTRODUCTION

Bayesian hierarchical linear modeling [8]–[11] is one of the major success stories of modern Bayesian inference [1]–[3]. A wide range of applications have emerged, for instance in frame-based theory [12], compressed sensing [4,5,13,14], source localization [15], astronomical data analysis [16], neuro-imaging [7] and wireless communications [6,17].

In the literature, we can find a plethora of estimators of the parameters of the Bayesian hierarchical linear model (BHLM), among them the linear minimum-mean-square error (LMMSE) estimator that will be discussed later. In order to benchmark the performance of such estimators we present analytical expressions of Bayesian lower bounds on their mean square error (MSE) [18,19]. Bayesian lower bounds are of practical interest since, given an estimator, they indicate how close to optimum this estimator performs and indirectly whether some improvement is still possible. They also allow to design the linear model in such a way to obtain the best achievable accuracy [20,21].

Among the large choice of the possible lower bounds, we focus our attention on Van Trees' Bayesian Cramér-Rao bound (BCRB) [18] for the estimation of the amplitude and noise parameters/hyper-parameters for the BHLM in the two following contexts:

(i) Unspecified arbitrary prior and hyper-prior for the amplitudes and Gaussian-$\mathcal{X}$ noise prior.

(ii) Gaussian-$\mathcal{Y}$ amplitude prior and Gaussian-$\mathcal{X}$ noise prior.

With Gaussian-$\mathcal{X}$ we mean that the conditional prior given the hyper-parameter is Gaussian and $\mathcal{X}$ is the hyper-prior distribution. The proposed approach has two main advantages: (i) the closed-form expressions can provide useful information on the nature of the parametric estimation problem in the context of the BHLM and (ii) the computational complexity of the proposed bounds is low, whereas the complexity of computing the MMSE is often prohibitive due to the posterior mean that it involves.

II. HIERARCHICAL BAYESIAN LINEAR MODEL

To specify the BHLM, we consider the linear model [4]–[7]

$$\mathbf{y} = \mathbf{H}\boldsymbol{\alpha} + \mathbf{n} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^N$ is the observation vector, $\mathbf{H}$ denotes a known deterministic real $N \times K$ matrix and $\boldsymbol{\alpha} = [\alpha_1 \dots \alpha_K]^T \in \mathbb{R}^K$ is the vector of amplitudes. We assume a two-layer hierarchical prior for the parameter (vector) of interest $\boldsymbol{\alpha}$. Specifically, the prior of $\boldsymbol{\alpha}$ depends on an hyperparameter $\boldsymbol{\gamma} = [\gamma_1 \dots \gamma_K]^T$ with a specified prior. With this assumption the joint probability density function (pdf) of $[\boldsymbol{\alpha}\,\boldsymbol{\gamma}]$ is of the form $p(\boldsymbol{\alpha}, \boldsymbol{\gamma}) = p(\boldsymbol{\alpha}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})$. The noise vector $\mathbf{n}$ follows a Gaussian-$\mathcal{X}$ distribution, meaning that $\mathbf{n}|\lambda \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{I}_N)$, where the noise precision λ follows a given distribution $\mathcal{X}$ with pdf $p(\lambda)$. The unknown parameter vector, $\boldsymbol{\theta}$, consists of the vector of amplitudes, their hyper-parameters and the noise precision, i.e., $\boldsymbol{\theta} = [\boldsymbol{\psi}^T\ \lambda]^T$ with $\boldsymbol{\psi} = [\boldsymbol{\alpha}^T\ \boldsymbol{\gamma}^T]^T$. The joint pdf of $(\mathbf{y}, \boldsymbol{\theta})$ reads

$$p(\mathbf{y}, \boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\alpha}, \lambda)p(\boldsymbol{\alpha}|\boldsymbol{\gamma})p(\boldsymbol{\gamma})p(\lambda). \tag{2}$$

In writing (2) we make the additional assumption that the hyper-parameters $\boldsymbol{\gamma}$ and λ are independent. This pdf together with (1) specify the BHLM.

III. BCRB FOR THE BHLM WITH GAUSSIAN-$\mathcal{X}$ NOISE

Let $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{y})$ be an estimator of $\boldsymbol{\theta}$. Its MSE fulfils the inequality $\mathbb{E}\left\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T\right\} \geq \mathbf{C} = \mathbf{J}^{-1}$, where $\mathbf{C}$ denotes the BCRB matrix and $\mathbf{J}$ is the Bayesian information matrix (BIM) for $\boldsymbol{\theta}$ [18].

Under weak regularity assumptions on $p(\mathbf{y}, \boldsymbol{\theta})$ in (2), we have the following result.

Result III.1. *The BCRB matrix for the BHLM with Gaussian-$\mathcal{X}$ noise distribution reads*

$$\mathbf{C} = \mathbf{J}^{-1} = \text{bdiag}\left\{\mathbf{C}_{\boldsymbol{\psi},\boldsymbol{\psi}}, \mathrm{C}_{\lambda,\lambda}\right\} \tag{3}$$

where bdiag *denotes the block diagonal operator and*

$$\mathbf{C}_{\boldsymbol{\psi},\boldsymbol{\psi}} = \begin{bmatrix} \mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\alpha}} & \mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\gamma}} \\ \mathbf{C}_{\boldsymbol{\gamma},\boldsymbol{\alpha}} & \mathbf{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}} \end{bmatrix}. \tag{4}$$

$$\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\alpha}} = \left[\mathbb{E}\{\lambda\}\mathbf{H}^T\mathbf{H} + \mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\alpha})}\right\} - \mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\gamma})}\right\}\left(\mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\right\} + \mathbf{J}_{HP}^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\right)^{-1}\mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\gamma})}\right\}^T\right]^{-1}, \tag{5}$$

$$\mathbf{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}} = \left[\mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\right\} + \mathbf{J}_{HP}^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\right]^{-1}\left(\mathbf{I} + \mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\gamma})}\right\}^T\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\alpha}}\,\mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\gamma})}\right\}\left(\mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\right\} + \mathbf{J}_{HP}^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\right)^{-1}\right), \tag{6}$$

$$\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\gamma}} = -\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\alpha}}\,\mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\gamma})}\right\}\left[\mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\right\} + \mathbf{J}_{HP}^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\right]^{-1}, \quad \mathrm{C}_{\lambda,\lambda} = \left(\mathbb{E}\left\{\frac{-\partial^2\log p(\lambda)}{\partial\lambda^2}\right\} + \frac{N}{2}\mathbb{E}\left\{\lambda^{-2}\right\}\right)^{-1} \tag{7}$$

The terms $\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\alpha}}$, $\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\gamma}}$, $\mathbf{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}}$, and $\mathrm{C}_{\lambda,\lambda}$ are given in (5)-(7), with the expressions for $\mathbf{J}_P^{(.,.)}$ and $\mathbf{J}_{HP}^{(.,.)}$ reported in Appendix VIII-A.

Proof. See Appendix VIII-A. □

Remark III.1. *Due to the block structure of the BCRB matrix (3), the optimal estimation error of the noise hyperparameter and the optimal estimation errors of the amplitudes and their hyper-parameters are uncorrelated.*

IV. BCRB FOR THE BHLM WITH GAUSSIAN-$\mathcal{Y}$ AMPLITUDE AND GAUSSIAN-$\mathcal{X}$ NOISE PRIORS

In this section, we consider case (ii) where the amplitudes follow a Gaussian-$\mathcal{Y}$ distribution, while noise is assumed to be Gaussian-$\mathcal{X}$ distributed. This means that the conditional pdf of the amplitudes and the noise are Gaussian given the hyper-parameters $\boldsymbol{\gamma}$ and λ, respectively [2,9,23]. In this case, Result III.1 particularizes as follows:

Result IV.1. *Let assume that $\boldsymbol{\alpha}|\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}\{\boldsymbol{\gamma}\})$, with* diag *denoting the diagonal operator, and that the hyper-parameters γ_i, $i \in \{1,\dots,K\}$ are independently drawn from a distribution $\mathcal{X}$ with pdf $p(\lambda)$. Then, the diagonal elements of the matrices in (5)-(7) have the closed-form expressions given in (10)-(11) for $i \in \{1,\dots,K\}$. In these expressions, $\mathbf{H}_i = [\mathbf{h}_j : j \in \{1,\dots,K\}\backslash\{i\}]$, with $\mathbf{h}_k$, $k \in \{1,\dots,K\}$, denoting the k-th column of $\mathbf{H}$, and $\mathbf{D}_i$ is the $(K-1)\times(K-1)$ diagonal matrix*

$$\begin{aligned}\mathbf{D}_i &= \mathrm{diag}\Big\{[\mathbf{D}]_{jj} : j \in \{1,\dots,K\}\backslash\{i\}\Big\} &(8)\\ &= \mathrm{diag}\big\{\mathbb{E}\{\gamma_j^{-1}\} : j \in \{1,\dots,K\}\backslash\{i\}\big\} &(9)\end{aligned}$$

with $\mathbf{D} = \mathbb{E}\left\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\alpha})}\right\}$ given in Appendix VIII-B.

Proof. See Appendix VIII-B. □

Remark IV.1. (i) *The Gaussian zero-mean amplitude prior assumption implies that the optimal estimation errors of the parameters and hyper-parameters are uncorrelated.*
(ii) *The optimal estimation error of the hyper-parameters is no longer a function of their corresponding parameters.*

Remark IV.2.
- In the low noise variance regime the BCRB *for α_i has the simple expression*

$$\mathrm{BCRB}(\alpha_i) \approx \frac{1}{\mathbb{E}\{\lambda\}\left\|\mathbf{P}_{\mathbf{H}_i}^{\perp}\mathbf{h}_i\right\|^2} \tag{12}$$

where $\mathbf{P}_{\mathbf{H}_i}^{\perp}$ denotes the orthogonal projector on $\langle\mathbf{H}_i\rangle^{\perp}$.
- The optimal estimation performance of the hyper-parameters is independent of noise.

A. BCRB *for spherically invariant random amplitudes*

In this section, we focus on the derivation of closed-form expressions of the BCRB matrix when the amplitude vector is a spherically invariant random variable (SIRV) [26]. This choice is motivated by the popularity of this model in engineering and signal processing applications [4]–[7,27]–[32]. The modelling of a SIRV (also called Gaussian scale mixture in the literature, e.g., [17]) is done by considering a two-scale compound-Gaussian process containing two components, which are commonly referred to as the texture and the speckle terms, such that $p(\boldsymbol{\alpha}|\boldsymbol{\gamma}) = \prod_{i=1}^{K} p(\alpha_i|\gamma_i)$ with $\alpha_i = \sqrt{\gamma_i}x_i$, $i \in \{1,\dots,K\}$. In our context the speckle term, x_i, is assumed to follow a Gaussian distribution with zero mean and unit variance, whereas, the positive texture term, i.e. $\gamma_i > 0$, follows a given a priori distribution $\mathcal{Y}$ with pdf $p(\gamma_i)$. From $p(\boldsymbol{\alpha}|\boldsymbol{\gamma})$, notice that this coincides with the BHLM with a Gaussian-$\mathcal{Y}$ amplitude distribution.

In the following, we consider the so-called generalized inverse Gaussian (GIG) distribution, denoted by $\mathcal{GIG}(\rho,\phi,\omega)$ with parameters ρ,ϕ and ω defined in [35] as hierarchical prior for the amplitudes. The GIG distribution encompasses the well-known gamma and inverse-gamma distributions, which lead to, respectively, a multi-variate K-distributed and multi-variate student-t amplitude vector. Both models are known to be well adapted to practical scenario [3]. The next result gives a closed-form expression of the BCRB in the aforementioned case, see Appendix VIII-C for the proof.

Result IV.2. *The entries of the* BCRB *matrix associated with the BHLM (1) with Gaussian-GIG amplitudes, i.e., $\gamma_i \sim \mathcal{GIG}(\rho_i,\phi_i,\omega_i)$, and Gaussian-$\mathcal{X}$ noise precision are given by (10) for the amplitudes where $\mathbb{E}\{\gamma_i^{-1}\} = M_i = \sqrt{\frac{\rho_i}{\phi_i}\frac{K_{\omega_i-1}(\sqrt{\rho_i\phi_i})}{K_{\omega_i}(\sqrt{\rho_i\phi_i})}}$, $[\mathbf{D}_i]_{jj} = M_j$ and*

$$\mathrm{BCRB}(\gamma_i) = \frac{\phi_i K_{\omega_i}(\sqrt{\rho_i\phi_i})}{\rho_i\left(\left(\omega_i - \frac{1}{2}\right)K_{\omega_i-2}(\sqrt{\rho_i\phi_i}) + \sqrt{\rho_i\phi_i}K_{\omega_i-3}(\sqrt{\rho_i\phi_i})\right)} \tag{13}$$

where $K_\mu(.)$ is the modified Bessel function of the second kind of order μ.

In the following, we use Result IV.2 to derive the BCRB for the amplitudes for some special cases of the GIG prior distribution listed in [34]. Note that we merely need to derive

$$\mathrm{C}_{\alpha_i,\alpha_i} \triangleq \mathrm{BCRB}(\alpha_i) = \frac{\mathbb{E}\{\lambda\}^{-1}}{\mathbf{h}_i^T\left(\mathbf{I}-\mathbf{H}_i\left[\mathbf{H}_i^T\mathbf{H}_i+\mathbb{E}\{\lambda\}^{-1}\mathbf{D}_i\right]^{-1}\mathbf{H}_i^T\right)\mathbf{h}_i+\mathbb{E}\{\lambda\}^{-1}\mathbb{E}\{\gamma_i^{-1}\}}, \tag{10}$$

$$\mathrm{C}_{\gamma_i,\gamma_i} \triangleq \mathrm{BCRB}(\gamma_i) = \left(\frac{1}{2}\mathbb{E}\{\gamma_i^{-2}\} - \mathbb{E}\left\{\frac{\partial^2 \log p(\gamma_i)}{\partial^2\gamma_i}\right\}\right)^{-1} \quad \text{and} \quad \mathrm{C}_{\alpha_i,\gamma_i} \triangleq \mathrm{BCRB}(\alpha_i,\gamma_i) = 0 \tag{11}$$

M_i, $i = 1,\ldots,K$. To compute this parameter we shall rely on some asymptotic approximations of the modified Bessel function of the second kind. Specifically, from [33, Eq.9.6.8, Eq.9.6.9] we have for $\omega_i < 0$

$$K_0(z) \approx -\ln(z) \quad \text{and} \quad K_{\omega_i}(z) \approx 2^{\omega_i-1}\gamma(\omega_i)z^{-\omega_i}, \quad z \text{ small.} \tag{14}$$

From [33, Eq.9.6.6] and (14) we obtain for $\omega_i < 0$

$$K_{\omega_i}(z) \approx \gamma(-\omega_i)2^{-\omega_i-1}z^{\omega_i}, \quad z \text{ small.} \tag{15}$$

Making use of (14), (15), and Result IV.2, we obtain analytical expressions for $\mathrm{BCRB}(\gamma_i)$ for different settings of the parameters ρ_i, ϕ_i, and ω_i of the GIG distribution [35]:

• $\omega_i = 0$: The hyper-prior $\mathcal{GIG}(\rho_i,\phi_i,\omega_i)$ coincides with the Barndroff-Nielsen hyperbolic distribution $\mathcal{HBN}(\phi_i,\rho_i)$ [36]. In this case, $M_i = \frac{\sqrt{\frac{\rho_i}{\phi_i}}K_{-1}(\sqrt{\rho_i\phi_i})}{K_0(\sqrt{\rho_i\phi_i})}$ and

$$\mathrm{BCRB}(\gamma_i) = \frac{\sqrt{\phi_i}}{\rho_i^{3/2}}\frac{K_0(\sqrt{\rho_i\phi_i})}{K_{-3}(\sqrt{\rho_i\phi_i})-\frac{1}{2}K_{-2}(\sqrt{\rho_i\phi_i})}. \tag{16}$$

• $\omega_i = -\frac{1}{2}$: $\mathcal{GIG}(\rho_i,\phi_i,\omega_i)$ reduces to the Wald distribution, also known as the inverse-Gaussian distribution, $\mathcal{IG}(\sqrt{\frac{\rho_i}{\phi_i}},\phi_i)$ [34]. Making use of [37, A1.5] we obtain $M_i = \rho_i/\phi_i$ and

$$\mathrm{BCRB}(\gamma_i) = \frac{\phi_i^{\frac{3}{2}}}{3\sqrt{\rho_i}(1+\sqrt{\rho_i\phi_i})}. \tag{17}$$

• $\omega_i < 0$ and $\rho_i \to 0$: $\mathcal{GIG}(\rho_i,\phi_i,\omega_i)$ converges to the inverse-gamma distribution Inv-Gamma$(-\omega_i,\frac{\phi_i}{2})$ with shape ω_i and scale $\frac{\phi_i}{2}$. Using (14)-(15) we have $M_i \approx \frac{\gamma(\omega_i-1)}{\phi_i\gamma\omega_i}$ and

$$\mathrm{BCRB}(\gamma_i) \approx \frac{\phi_i^2}{4}\frac{\gamma(-\omega_i)}{(\omega_i-\frac{1}{2})\gamma(2-\omega_i)+2\gamma(3-\omega_i)}. \tag{18}$$

• $\omega_i > 0$ and $\phi_i \to 0$: $\mathcal{GIG}(\rho_i,\phi_i,\omega_i)$ converges to the gamma distribution Gamma$(\omega_i,\frac{\rho_i}{2})$ with shape ω_i and rate $\frac{\rho_i}{2}$. Plugging (14) and (15) into M_i and assuming ϕ_i is small we obtain $M_i \approx \frac{\rho_i}{2}\frac{\gamma(\omega_i-1)}{\gamma(\omega_i)}$ for $\omega_i > 1$ and

$$\mathrm{BCRB}(\gamma_i) \approx \frac{4}{\rho_i^2}\frac{\gamma(\omega_i)}{(\omega_i-\frac{1}{2})\gamma(\omega_i-2)} \quad \text{for} \quad \omega_i > 2. \tag{19}$$

V. Analysis of the LMMSE estimator

A. MSE expression

Generalizing the Gauss-Markov Theorem [40] to the BHLM, the MSE of the LMMSE estimator is given by

$$\begin{aligned}\mathrm{MSE_L}(\hat{\boldsymbol{\alpha}}) &= \mathbb{E}\left\{\mathbb{E}\{||\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}}||^2|\boldsymbol{\gamma},\lambda\}\right\}\\ &= \mathrm{Tr}\mathbb{E}\left[\lambda^{-1}\left(\mathbf{H}^T\mathbf{H}+\mathrm{diag}\left\{(\lambda\gamma_1)^{-1},\ldots,(\lambda\gamma_K)^{-1}\right\}\right)^{-1}\right],\end{aligned}$$

where Tr is the trace operator.

B. Stochastic dominance scenario

We consider the context of stochastic dominance[1].

1) Low noise variance regime: In this regime, we obtain for the GIG prior distribution: $\mathrm{MSE_L}(\hat{\boldsymbol{\alpha}}) \simeq \sum_{k=1}^K\left(\frac{\phi_k}{\rho_k}\right)^{\frac{1}{2}}\frac{K_{\omega_k+1}(\sqrt{\rho_k\phi_k})}{K_{\omega_k}(\sqrt{\rho_k\phi_k})}$ and $\mathrm{Tr}\left(\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\alpha}}\right) \simeq \sum_{k=1}^K\left(\frac{\phi_k}{\rho_k}\right)^{\frac{1}{2}}\frac{K_{\omega_k}(\sqrt{\rho_k\phi_k})}{K_{\omega_k-1}(\sqrt{\rho_k\phi_k})}$. Thus, the variance of the LMMSE estimator approaches the BCRB when $|\omega_i| \gg 1$. For example, among the GIG distribution family, the Gaussian-gamma and Gaussian-inverse-gamma distributions fulfill this condition provided the shape parameter is sufficiently large.

2) High noise variance regime: In this regime, we have $\mathrm{MSE_L}(\hat{\boldsymbol{\alpha}}) \simeq \mathbb{E}\left\{\lambda^{-1}\right\}\mathrm{Tr}\left[\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\right]$, whereas $\mathrm{Tr}\left(\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\alpha}}\right) \simeq \mathbb{E}\left\{\lambda\right\}^{-1}\mathrm{Tr}\left[\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\right]$. From Jensen's inequality [41], $\mathbb{E}\left\{\lambda\right\}^{-1} \leq \mathbb{E}\left\{\lambda^{-1}\right\}$. Thus the MSE of the LMMSE estimator is higher than the BCRB. The MSE may be close to the BCRB for distributions that fulfil $\mathbb{E}\left\{\lambda\right\}^{-1} \simeq \mathbb{E}\left\{\lambda^{-1}\right\}$, among them the gamma and inverse-gamma distributions.

VI. Numerical investigations

Due to space limitation, we confine the numerical investigations to i) compare the BCRBs for two different selections of the hyper-prior distributions of the amplitudes, i.e. the gamma and inverse-gamma distributions with different settings of the shape parameters, and (ii) quantify how close the MSE of the LMMSE estimator is to these bounds. Fig. 1 depicts the numerical results.

First of all, Fig. 1 reveals that the BCRB and the MSE computed using the inverse-gamma prior distribution are lower than those resulting from using the gamma prior distribution. As expected, the MSE achieves the BCRB in the asymptotic region (low noise variance) and in the so-called non-informative region (high noise variance). Nevertheless, a certain gap between the MSE and the BCRB remains in the transition region. This gap is mainly due to the presence of outliers in this region, of which the effect is not taken into account in the BRCB [18].

Fig. 1 also shows that at low SNR the BCRB and the MSE increase (decrease) as the shape parameter of the gamma (inverse-gamma) distribution increases. For the gamma distribution this can be explained by the fact that as the shape

[1] i.e., $P(\lambda^{-1} \geq \zeta) \geq P(\gamma_k \geq \zeta), \forall\zeta$, and $\exists\zeta : P(\lambda^{-1} \geq \zeta) > P(\gamma_k \geq \zeta)$. See [24,25] for some practical applications.

parameter decreases, the prior distribution of the amplitudes becomes more heavy-tailed. Note that the Gaussian distribution is the limit of the gamma distribution when the shape tends to infinity. Thus, the selection of the Gaussian distribution in the SIRV family leads to the largest BCRB at low SNR and to the worst performance of the LMSEE estimator in this regime. This behaviour is expected, as it was already noticed in the deterministic case, e.g., in [38]. As also expected, when the SNR is above 0dB the value of the shape parameter is irrelevant.

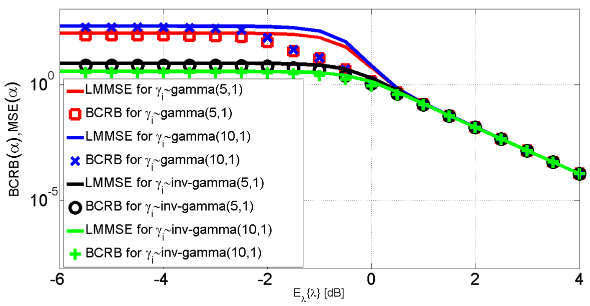

Fig. 1: BCRB and MSE of the LMMSE estimator versus the mean noise precision for different selections of the hyper-prior distribution for the amplitudes.

VII. Conclusions

In this paper, we derived closed-form expressions of the Bayesian Cramér-Rao bounds in the context of Bayesian hierarchical linear models. We gave special attention to the commonly used case of Gaussian-$\mathcal{Y}$ distributed amplitudes and Gaussian-$\mathcal{X}$ distributed noise precision. We derived a general form of the Bayesian Cramér-Rao bounds for SIRV distributed amplitudes and we obtained closed-form expressions for the generalized inverse Gaussian prior distribution. Furthermore, theoretical and numerical analyses revealed interesting properties that were discussed in the paper. Finally, we assessed how close the derived BCRB is to the MSE of the LMMSE estimator.

Acknowledgement

Part of this work was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM♯ (Grant agreement no. 318306), the iCODE institute (ANR-11-IDEX-0003-02), the TITAN project and the MAGELLAN project(ANR-14-CE23-0004-01).

VIII. Appendix

A. Proof of Result III.1

We assume that the BHLM is identifiable. Under weak regularity conditions [18] the $(2K+1)\times(2K+1)$ BIM reads

$$\mathbf{J} = \mathbb{E}\Big\{\mathbf{J}_S^{(\boldsymbol{\theta},\boldsymbol{\theta})}\Big\} + \mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\theta},\boldsymbol{\theta})}\Big\} + \mathbf{J}_{HP}^{(\boldsymbol{\theta},\boldsymbol{\theta})} + \mathbf{J}_N^{(\boldsymbol{\theta},\boldsymbol{\theta})}. \tag{20}$$

The four involved matrices are defined as follows: $[\mathbf{J}_S^{(\boldsymbol{\theta},\boldsymbol{\theta})}]_{ij} = \mathbb{E}\Big\{-\frac{\partial^2 \log p(\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\gamma},\lambda)}{\partial\theta_i\partial\theta_j}|\boldsymbol{\alpha},\boldsymbol{\gamma},\lambda\Big\}$, $[\mathbf{J}_P^{(\boldsymbol{\theta},\boldsymbol{\theta})}]_{ij} = \mathbb{E}\Big\{-\frac{\partial^2 \log p(\boldsymbol{\alpha}|\boldsymbol{\gamma})}{\partial\theta_i\partial\theta_j}|\boldsymbol{\gamma}\Big\}$, $[\mathbf{J}_{HP}^{(\boldsymbol{\theta},\boldsymbol{\theta})}]_{ij} = \mathbb{E}\Big\{-\frac{\partial^2 \log p(\boldsymbol{\gamma})}{\partial\theta_i\partial\theta_j}\Big\}$, and $[\mathbf{J}_N^{(\boldsymbol{\theta},\boldsymbol{\theta})}]_{ij} = \mathbb{E}\Big\{-\frac{\partial^2 \log p(\lambda)}{\partial\theta_i\partial\theta_j}\Big\}$, $(i,j)\in\{1,\ldots,2K+1\}^2$. As the noise is assumed to be Gaussian-$\mathcal{X}$, we have $\mathbf{y}|\boldsymbol{\alpha},\boldsymbol{\gamma},\lambda \sim \mathcal{N}(\boldsymbol{\mu},\mathbf{R})$ with $\boldsymbol{\mu} = \mathbf{H}\boldsymbol{\alpha}$ and $\mathbf{R} = \lambda^{-1}\mathbf{I}_N$. Thus, the pdf of observation $\mathbf{y}$ conditioned on $\boldsymbol{\theta}$ is solely a function of the amplitudes and the noise precision. By using the Slepian-Bang formula [22] we obtain $\mathbb{E}\Big\{\mathbf{J}_S^{(\boldsymbol{\theta},\boldsymbol{\theta})}\Big\} = \mathrm{bdiag}\Big\{\mathbb{E}\Big\{\mathbf{J}_S^{(\boldsymbol{\alpha},\boldsymbol{\alpha})}\Big\},\mathbf{0},\mathbb{E}\Big\{J_S^{(\lambda,\lambda)}\Big\}\Big\}$ with $\mathrm{bdiag}\{.\}$ denoting the block diagonal operator. Moreover, $\mathbb{E}\Big\{\mathbf{J}_S^{(\boldsymbol{\alpha},\boldsymbol{\alpha})}\Big\} = \mathbb{E}\{\lambda\}\,\mathbf{H}^T\mathbf{H}$ and $\mathbb{E}\Big\{J_S^{(\lambda,\lambda)}\Big\} = \frac{N}{2}\mathbb{E}\{\lambda^{-2}\}$. Since $p(\boldsymbol{\alpha}|\boldsymbol{\gamma})$ is not function of λ, we have $\mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\theta},\boldsymbol{\theta})}\Big\} = \mathrm{bdiag}\left\{\begin{bmatrix}\mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\alpha})}\Big\} & \mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\gamma})}\Big\}\\ \mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\gamma},\boldsymbol{\alpha})}\Big\} & \mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\Big\}\end{bmatrix},\mathbf{0}\right\}$. Because $p(\boldsymbol{\gamma})$ is not a function of $\boldsymbol{\alpha}$ and λ, $\mathbf{J}_{HP}^{(\boldsymbol{\theta},\boldsymbol{\theta})} = \mathrm{bdiag}\Big\{\mathbf{0},\mathbf{J}_{HP}^{(\boldsymbol{\gamma},\boldsymbol{\gamma})},\mathbf{0}\Big\}$ and $\mathbf{J}_N^{(\boldsymbol{\theta},\boldsymbol{\theta})} = \mathrm{bdiag}\Big\{\mathbf{0},\mathbf{0},J_N^{(\lambda,\lambda)}\Big\}$ with $J_N^{(\lambda,\lambda)} = \mathbb{E}\left\{-\frac{\partial^2\log p(\lambda)}{\partial\lambda^2}\right\}$.

B. Proof of Result IV.1

After some algebraic manipulations, we obtain $[\mathbf{D}]_{ij} = \Big[\mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\alpha})}\Big\}\Big]_{ij} = \mathbb{E}\big\{\gamma_i^{-1}\big\}\,\delta_{i-j}$, $\Big[\mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\Big\}\Big]_{ij} = \mathbb{E}\big\{\alpha_i^2\gamma_i^{-3}\big\}\,\delta_{i-j} - \frac{1}{2}\mathbb{E}\big\{\gamma_i^{-2}\big\}\,\delta_{i-j}$ and $\Big[\mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\gamma})}\Big\}\Big]_{ij} = -\mathbb{E}\big\{\alpha_i\gamma_i^{-2}\big\}\,\delta_{i-j}$, with δ denoting the Kronecker delta symbol. From

$$\mathbb{E}\big\{\alpha_i^k\gamma_i^{-k-1}\big\} = \begin{cases} 0 & \text{for } k=1,\\ \mathbb{E}\big\{\gamma_i^{-2}\big\} & \text{for } k=2,\end{cases} \tag{21}$$

we have $\mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\gamma})}\Big\} = \mathbf{0}$ and $\Big[\mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\Big\}\Big]_{ij} = \frac{1}{2}\mathbb{E}\big\{\gamma_i^{-2}\big\}\delta_{i-j}$. Using these last expressions in (5)-(7) yields $\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\gamma}} = \mathbf{0}$ and

$$\begin{aligned}\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\alpha}} &= \left[\mathbb{E}\{\lambda\}\mathbf{H}^T\mathbf{H} + \mathrm{diag}\left\{\mathbb{E}\big\{\gamma_1^{-1}\big\}\ldots\mathbb{E}\big\{\gamma_K^{-1}\big\}\right\}\right]^{-1},\\ \mathbf{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}} &= \left[\frac{1}{2}\mathrm{diag}\big\{\mathbb{E}\big\{\gamma_1^{-2}\big\},\ldots,\mathbb{E}\big\{\gamma_K^{-2}\big\}\big\} + \mathbf{J}_{HP}^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\right]^{-1}\end{aligned}$$

with

$$\left[\mathbf{J}_{HP}^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\right]_{ij} = -\mathbb{E}\left\{\frac{\partial^2\log p(\gamma_i)}{\partial\gamma_i^2}\right\}\delta_{i-j}. \tag{22}$$

Consider the rearrangement matrix $\mathbf{P}_i$ that operates on $\boldsymbol{\alpha}$ as follows: $\mathbf{P}_i\boldsymbol{\alpha} = [\alpha_i\alpha_1\ldots\alpha_{i-1}\alpha_{i+1}\ldots\alpha_K]^T$. Then we have $\mathbf{P}_i\mathbf{H}\mathbf{P}_i^T = [\mathbf{h}_i\quad\mathbf{H}_i]$. Since $[\mathbf{C}_{\mathbf{P}_i\boldsymbol{\alpha},\mathbf{P}_i\boldsymbol{\alpha}}]_{11} = [\mathbf{C}_{\boldsymbol{\alpha},\boldsymbol{\alpha}}]_{ii}$, by following the same methodology as applied in [39] we obtain (10) after some tedious but straightforward manipulations.

C. Closed-form expression of the BCRB

By making use of the i.i.d. assumption we show readily that $\frac{\partial^2\log(p(\boldsymbol{\gamma}|\boldsymbol{\rho},\boldsymbol{\phi},\boldsymbol{\omega}))}{\partial\gamma_i^2} = (1-\omega_i)\,\gamma_i^{-2} - \phi_i\gamma_i^{-3}$ when $\gamma_i \sim \mathcal{GIG}(\rho_i,\phi_i,\omega_i)$. Inserting in (22) yields $\Big[\mathbb{E}\Big\{\mathbf{J}_{HP}^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\Big\}\Big]_{ii} = \mathbb{E}\big\{(\omega_i-1)\,\gamma_i^{-2} + \phi_i\gamma_i^{-3}\big\}$, $i\in\{1,\ldots K\}$. Recall that we obtained in Appendix VIII-B $\Big[\mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\alpha},\boldsymbol{\alpha})}\Big\}\Big]_{ii} = \mathbb{E}\big\{\gamma_i^{-1}\big\}$, $\Big[\mathbb{E}\Big\{\mathbf{J}_P^{(\boldsymbol{\gamma},\boldsymbol{\gamma})}\Big\}\Big]_{ii} = \frac{1}{2}\mathbb{E}\big\{\gamma_i^{-2}\big\}$, $i\in\{1,\ldots K\}$. Result IV.2 follows then from $\mathbb{E}\{\gamma_i^n\} = \left(\frac{\phi_i}{\rho_i}\right)^{\frac{n}{2}}\frac{K_{\omega_i+n}(\sqrt{\rho_i\phi_i})}{K_{\omega_i}(\sqrt{\rho_i\phi_i})}$.

REFERENCES

[1] A. Gelman, "Prior distributions for variance parameters in hierarchical models", International Society for Bayesian Analysis, vol. 1, No. 3, pp. 515-534, 2006.

[2] A. Lee, F. Caron, A. Doucet and C. Holmes, "A hierarchical Bayesian framework for constructing sparsity-inducing priors", ArXiv:1009.1914, 2010.

[3] A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin "Bayesian Data Analysis", Second Edition. Boca Raton, FL: Chapman and Hall/CRC. ISBN 1-58488-388-X, 2003.

[4] S. D. Babacan, S. Nakajima, and M. N. Do, "Bayesian Group-Sparse Modeling and Variational Inference", IEEE Trans. on Signal Processing, Vol. 62, No. 11, June, pp.2906-2921, 2014.

[5] A. Lee, F. Caron, A. Doucet, and C. Holmes, "A hierarchical Bayesian framework for constructing sparsity-inducing priors", ArXiv:1009.1914, Sept. 2010.

[6] N. L. Pedersen, C. N. Manchon, D. Shutin, B. H. Fleury, "Application of Bayesian hierarchical prior modeling to sparse channel estimation", in Proc. IEEE Int. Communications Conf. (ICC) (2012) 34873492.

[7] 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. 641655, 2010.

[8] S. Adrian, "A general Bayesian linear model", Journal of the Royal Statistical Society. Series B (Methodological), pp. 67-75, 1973.

[9] M. West and M. D. Escobar, "Hierarchical priors and mixture models, with application in regression and density estimation", Institute of Statistics and Decision Sciences, Duke University, 1993.

[10] V. Ollier, R. Boyer, M. N. El Korso, and P. Larzabal, "Performance bounds for hierarchical priors in the large random matrix framework", submitted to, Proc. of IEEE Workshop on Sensor Array and Multi-channel Processing (SAM-16).

[11] L. Dennis and S. Adrian, "Bayes estimates for the linear model", Journal of the Royal Statistical Society. Series B (Methodological), pp. 1-41, 1972.

[12] L. Chaeri, J. C. Pesquet, J. Y. Tourneret, P. Ciuciu and A. Benazza-Benyahia, "A Hierarchical Bayesian Model for Frame Representation", IEEE Trans. on Signal Processing, vol. 58, no. 11, 2010.

[13] X. Tan and J. Li, "Computationally Efficient Sparse Bayesian Learning via Belief Propagation", IEEE Trans. on Signal Processing, vol. 58, no. 4, 2010.

[14] S. Ji, D. Dunson and L. Carin, "Multitask Compressive Sensing", IEEE Trans. on Signal Processing, vol. 57, no. 1, 2009.

[15] F. Asano, H. Asoh and K. Nakadai, "Sound Source Localization Using Joint Bayesian Estimation With a Hierarchical Noise Model", IEEE Trans. on Audio Speech and Language Processing, vol. 21, no. 9, 2013.

[16] N. Dobigeon, J. Y. Tourneret and J. D. Scargle, "Joint Segmentation of Multivariate Astronomical Time Series: Bayesian Sampling With a Hierarchical Model", IEEE Trans. on Signal Processing, vol. 55, no. 2, 2007.

[17] N. L. Pedersen, C. N. Manchon, M.-A. Badiu, D. Shutin and B. H. Fleury "Sparse Estimation Using Bayesian Hierarchical Prior Modeling for Real and Complex Linear Models", Signal processing, 115 (2015), http://dx.doi.org/10.1016/j.sigpro.2015.03.013.

[18] H. L. Van Trees and K. L. Bell, "Bayesian Bounds for Parameter Estimation and Nonlinear Filtering-Tracking", John Wiley and Son, Inc., 2007.

[19] A. Renaux, P. Forster, P. Larzabal, C. Richmond, and A. Nehorai, "A fresh look at the bayesian bounds of the Weiss-Weinstein family", IEEE Trans. on Signal Processing, vol. 56, no. 11, pp. 5334-5352, 2008.

[20] H. Gazzah and S. Marcos, "Cramér Rao Bounds for Antenna Array Design", IEEE Trans. on Signal Processing, vol. 54, no. 1, pp. 336-345, 2006.

[21] J. Li, L. Xu, P. Stoica, K. W. Forsythe, W. Keith and D. W. Bliss, "Range compression and waveform optimization for MIMO radar: a Cramér Rao bound based study", IEEE Trans. on Signal Processing, vol. 56, no. 1, pp. 218-232, 2008.

[22] P. Stoica and R. L. Moses, "Spectral analysis of signals", Pearson/Prentice Hall Upper Saddle River, NJ, 2005.

[23] C. Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.

[24] M. Shaked and J. G. Shanthikumar, Stochastic orders and their applications, 1st ed. Springer, 1994

[25] C. Tepedelenlioglu, A. Rajan and Y. Zhang, "Applications of Stochastic Ordering to Wireless Communications", IEEE Trans. on Wireless Communication, vol. 10, no. 12, pp. 4249-4257, 2011.

[26] K. Yao, "A representation theorem and its applications to spherically invariant random processes", IEEE Trans. on Information Theory, vol. 19, no. 5, pp. 600-608, 1973.

[27] X. Zhang, M. N. El Korso and M. Pesavento, "MIMO radar target localization and performance evaluation under SIRP clutter", Submitted to, Elsevier Signal Processing Journal.

[28] E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor, "Complex elliptically symmetric distributions: survey, new results and applications", IEEE Trans. on Signal Processing, vol. 60, no. 11, pp. 5597-5625, 2012.

[29] X. Zhang, M. N. El Korso and M. Pesavento, "Maximum likelihood and maximum a posteriori direction-of-arrival estimation in the presence of SIRP noise", in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), Shaingai, China, March 2016.

[30] D. Middleton, "Man-made noise in urban environments and transportation systems: Models and measurements", IEEE Trans. on Wireless Communication, vol. 21, no. 11, pp. 12321241, 1973.

[31] F. Pascal, Y. Chitour, J. P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis", IEEE Trans. on Signal Processing, vol. 56, no. 1, pp. 3448, 2008.

[32] M. Rangaswamy, "Spherically invariant random processes for modeling non-Gaussian radar clutter", in Proc. 27th Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, vol. 2, pp. 1106-1110, 1993.

[33] M. Abramovitch and L. A. Stegun, "Handbook of Mathematical Tables". National Bureau of Standards, Washington, DC., 1970.

[34] N. L. Johnson, S. Kotz, and N. Balakrishnan, "Continuous Multivariate Distributions", volume 1, Models and Applications. New York: John Wiley & Sons, 2002.

[35] E. Eberlein and E. A. Hammerstein, "Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes", Seminar on Stochastic Analysis, Random Fields and Applications IV, pp. 221-264, Springer, 2004.

[36] O. E. Barndorf-Nielsen, "Normal Inverse Gaussian Distributions and stochastic volatility modelling", Scand. J. Statist. 24, pp. 1-13, 1997.

[37] O. E. Barndorff-Nielsen and P. Blaesild, "Hyperbolic distributions and ramifications: contributions to the theory and applications", in: C. Taillie, G. Patil and B. Baldessari, Eds., Statistical Distributions in Scientific Work (Reidel, Dordrecht), vol. 4, pp. 1944, 1981.

[38] M. N. El Korso, A. Renaux and P. Forster, "CRLB under K-distributed observation with parameterized mean", in Proc. of IEEE International Sensor Array and Multichannel Signal Processing Workshop, A Coruna, Spain, 2014.

[39] L. L. Scharf and L. T. Mac Worther, "Geometry of the Cramér-Rao bound", Elsevier, Signal Processing, vol. 31, no. 3, pp. 301-311, Apr. 1993.

[40] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, NJ: Prentice Hall, 1993.

[41] S. G. Krantz, "Jensen's Inequality", Handbook of Complex Variables. Boston, MA: Birkhauser, 1999.