Bounds for Estimation of Covariance Matrices From Heterogeneous Samples

Olivier Besson, *Senior Member, IEEE*,
Stéphanie Bidon, *Student Member, IEEE*, and
Jean-Yves Tourneret, *Member, IEEE*

***Abstract*—This correspondence derives lower bounds on the mean-square error (MSE) for the estimation of a covariance matrix $\boldsymbol{M}_p$, using samples $\boldsymbol{Z}_k$, $k=1,\ldots,K$, whose covariance matrices $\boldsymbol{M}_k$ are randomly distributed around $\boldsymbol{M}_p$. This framework can be encountered e.g., in a radar system operating in a nonhomogeneous environment, when it is desired to estimate the covariance matrix of a range cell under test, using training samples from adjacent cells, and the noise is nonhomogeneous between the cells. We consider two different assumptions for $\boldsymbol{M}_p$. First, we assume that $\boldsymbol{M}_p$ is a deterministic and unknown matrix, and we derive the Cramér–Rao bound for its estimation. In a second step, we assume that $\boldsymbol{M}_p$ is a random matrix, with some prior distribution, and we derive the Bayesian bound under this hypothesis.**

***Index Terms*—Bayesian bound, covariance matrix estimation, Cramér–Rao bound, heterogeneous environment.**

I. Problem Statement and Data Model

Estimating the covariance matrix of an observation vector is fundamental in many array processing applications, notably in adaptive radar detection where it is desired to estimate the noise statistics of a vector under test, so as to implement an adaptive detection scheme [1]. In an ideal situation, this task is performed using independent and identically distributed (i.i.d.) training samples, which share the same covariance matrix as the vector under test. In such a case, and under the assumption that all vectors are Gaussian, the sample covariance matrix (SCM) estimator is the maximum-likelihood estimator (MLE). However, heterogeneous environments are very frequently encountered [2], [3], and therefore the assumption of i.i.d. samples is often violated. More precisely, the training samples do not have the same covariance matrix as the vector under test, and they may even not share a common covariance matrix. In an attempt to take into account this fact, we proposed in [4] a model for heterogeneous environments; see also [5], where we discuss the rationale and relevance of such a model along with adaptive detection schemes related to it. More precisely, we assumed that the set of training samples can be divided in K groups. The kth group contains L_k snapshots $\{\boldsymbol{z}_{k,\ell}\}_{\ell=1}^{L_k}$ sharing the same covariance matrix $\boldsymbol{M}_k \neq \boldsymbol{M}_p$. When $K=1$, all training samples have a common covariance matrix, which is however different from $\boldsymbol{M}_p$. When $L_k=1$ for $k=1,\ldots,K$, all training samples have a different covariance matrix. The snapshots $z_{k,\ell}$ are assumed independent and Gaussian distributed, with covariance matrix $\boldsymbol{M}_k$, i.e., the distribution of $\boldsymbol{Z}_k = [\boldsymbol{z}_{k,1} \ \cdots \ \boldsymbol{z}_{k,L_k}]$, conditionally to $\boldsymbol{M}_k$ is

$$f(\boldsymbol{Z}_k \,|\, \boldsymbol{M}_k) = \pi^{-mL_k} |\boldsymbol{M}_k|^{-L_k} \operatorname{etr}\left\{-\boldsymbol{M}_k^{-1} \boldsymbol{Z}_k \boldsymbol{Z}_k^H\right\} \tag{1}$$

where $|\cdot|$ and $\operatorname{etr}\{\cdot\}$ stand for the determinant and the exponential of the trace of a matrix, respectively, and m is the size of the observation vector. The matrices $\boldsymbol{M}_k$ are assumed to be independent conditionally to $\boldsymbol{M}_p$, and distributed according to an inverse Wishart distribution with mean $\boldsymbol{M}_p$ and ν_k degrees of freedom, i.e., [6]

$$f(\boldsymbol{M}_k|\boldsymbol{M}_p) = \frac{|(\nu_k - m)\boldsymbol{M}_p|^{\nu_k} |\boldsymbol{M}_k|^{-(\nu_k+m)}}{\tilde{\Gamma}_m(\nu_k)} \times \operatorname{etr}\left\{-(\nu_k - m)\boldsymbol{M}_k^{-1}\boldsymbol{M}_p\right\} \tag{2}$$

where

$$\tilde{\Gamma}_m(p) = \pi^{m(m-1)/2} \prod_{k=1}^{m} \Gamma(p-k+1). \tag{3}$$

The scalar ν_k allows one to adjust the distance between $\boldsymbol{M}_k$ and $\boldsymbol{M}_p$: the larger ν_k, the closer $\boldsymbol{M}_k$ to $\boldsymbol{M}_p$ [6]. To summarize, the model for the training samples is given by

$$\boldsymbol{Z}_k \,|\, \boldsymbol{M}_k \sim \tilde{\mathcal{N}}_{m,L_k}(\boldsymbol{0}, \boldsymbol{M}_k, \boldsymbol{I}_{L_k}) \tag{4}$$

$$\boldsymbol{M}_k|\boldsymbol{M}_p \sim \tilde{\mathcal{W}}_m^{-1}((\nu_k - m)\boldsymbol{M}_p, \nu_k) \tag{5}$$

for $k = 1,\ldots,K$, where $\tilde{\mathcal{N}}_{m,L_k}(\boldsymbol{0}, \boldsymbol{M}_k, \boldsymbol{I}_{L_k})$ and $\tilde{\mathcal{W}}_m^{-1}((\nu_k - m)\boldsymbol{M}_p, \nu_k)$ denote the complex normal distribution and the complex inverse Wishart distribution, respectively. This correspondence considers two assumptions for $\boldsymbol{M}_p$, namely $\boldsymbol{M}_p$ is deterministic or $\boldsymbol{M}_p$ is a random matrix whose prior distribution is Wishart, with mean $\bar{\boldsymbol{M}}_p$ and μ degrees of freedom, i.e.,

$$f(\boldsymbol{M}_p) = \frac{1}{\tilde{\Gamma}_m(\mu)} |\mu^{-1}\bar{\boldsymbol{M}}_p|^{-\mu} |\boldsymbol{M}_p|^{\mu-m} \operatorname{etr}\left\{-\mu \boldsymbol{M}_p \bar{\boldsymbol{M}}_p^{-1}\right\}. \tag{6}$$

We denote this distribution as $\boldsymbol{M}_p \,|\, \bar{\boldsymbol{M}}_p \sim \tilde{\mathcal{W}}_m(\mu^{-1}\bar{\boldsymbol{M}}_p, \mu)$. Note that the distance between $\boldsymbol{M}_p$ and $\bar{\boldsymbol{M}}_p$ decreases as μ increases [6].

In [4], we proposed strategies for estimating $\boldsymbol{M}_p$ under this framework. The aim of this correspondence is to derive lower bounds for the MSE of $\boldsymbol{M}_p$. More precisely, we first assume that $\boldsymbol{M}_p$ is deterministic and derive its Cramér–Rao bound (CRB). Next, assuming that $\boldsymbol{M}_p$ is drawn from (6), we derive the Bayesian bound (BB) for its estimation. The counterpart of estimation, namely detection, is beyond the scope of the present correspondence and is not addressed here. Note also that, if the bounds enable one to measure the performance of estimators, they cannot always prejudge their performance in detection.

II. Cramér–Rao Bound (Deterministic $\boldsymbol{M}_p$)

We first derive the Cramér–Rao bound for estimation of $\boldsymbol{M}_p$, assuming the latter is a deterministic and unknown matrix. Let $\boldsymbol{Z} = [\boldsymbol{Z}_1 \ \ldots \ \boldsymbol{Z}_K]$ and first note that $f(\boldsymbol{Z} \,|\, \boldsymbol{M}_p)$ is given by

$$\begin{aligned} f(\boldsymbol{Z} \,|\, \boldsymbol{M}_p) &= \prod_{k=1}^{K} f(\boldsymbol{Z}_k \,|\, \boldsymbol{M}_p) \\ &= \prod_{k=1}^{K} \int f(\boldsymbol{Z}_k \,|\, \boldsymbol{M}_k) f(\boldsymbol{M}_k \,|\, \boldsymbol{M}_p) d\boldsymbol{M}_k \\ &= \prod_{k=1}^{K} \frac{|(\nu_k - m)\boldsymbol{M}_p|^{\nu_k}}{\pi^{mL_k}\tilde{\Gamma}_m(\nu_k)} \int |\boldsymbol{M}_k|^{-(\nu_k+L_k+m)} \\ &\quad \times \operatorname{etr}\left\{-\boldsymbol{M}_k^{-1}\left[(\nu_k - m)\boldsymbol{M}_p + \boldsymbol{Z}_k\boldsymbol{Z}_k^H\right]\right\} d\boldsymbol{M}_k \end{aligned}$$

Manuscript received July 16, 2007; revised December 13, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Petr Tichavsky. This work was supported by the Délégation Générale pour l'Armement (DGA) and by Thales Systèmes Aéroportés.

O. Besson and S. Bidon are with the Department of Electronics, Optronics and Signal, ISAE, University of Toulouse, , 31055 Toulouse, France (e-mail: besson@isae.fr; sbidon@isae.fr).

J.-Y. Tourneret is with IRIT/ENSEEIHT, 31071 Toulouse, France (e-mail: jean-yves.tourneret@enseeiht.fr).

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.2008.917341

$$
\begin{aligned}
&= \prod_{k=1}^{K} \frac{\tilde{\Gamma}_m(\nu_k + L_k)}{\pi^{mL_k}\tilde{\Gamma}_m(\nu_k)} |(\nu_k - m)\boldsymbol{M}_p|^{\nu_k} \\
&\quad \times |(\nu_k - m)\boldsymbol{M}_p + \boldsymbol{Z}_k\boldsymbol{Z}_k{}^H|^{-(\nu_k+L_k)} \\
&= \prod_{k=1}^{K} \frac{\tilde{\Gamma}_{L_k}(\nu_k + L_k)}{\pi^{mL_k}\tilde{\Gamma}_{L_k}(\nu_k + L_k - m)} |(\nu_k - m)\boldsymbol{M}_p|^{-L_k} \\
&\quad \times \left|\boldsymbol{I}_{L_k} + (\nu_k - m)^{-1}\boldsymbol{Z}_k{}^H\boldsymbol{M}_p{}^{-1}\boldsymbol{Z}_k\right|^{-(\nu_k+L_k)}. \qquad (7)
\end{aligned}
$$

We observe that $\boldsymbol{Z}_k$ is distributed according to a generalized complex multivariate t distribution, with $\nu_k + L_k - m$ degrees of freedom [7]. The log-likelihood function is thus

$$
\begin{aligned}
\Lambda(\boldsymbol{Z}\,|\,\boldsymbol{M}_p) = \text{const.} &+ \left(\sum_{k=1}^{K}\nu_k\right)\ln|\boldsymbol{M}_p| \\
&- \sum_{k=1}^{K}(\nu_k + L_k)\ln\left|(\nu_k - m)\boldsymbol{M}_p + \boldsymbol{Z}_k\boldsymbol{Z}_k{}^H\right|. \qquad (8)
\end{aligned}
$$

Let $\boldsymbol{m}_p = \text{vec}(\boldsymbol{M}_p)$ be the vector obtained by stacking the columns of $\boldsymbol{M}_p$ on top of each other. Accordingly, let $\tilde{\boldsymbol{m}}_p \in \mathbb{R}^{m^2\times 1}$ be the real-valued vector that consists of the elements along the diagonal of $\boldsymbol{M}_p$ and the real and imaginary parts of its elements under the diagonal. In order to obtain the CRB, we need to derive the Fisher information matrix (FIM) which is defined as [8]

$$
\tilde{\boldsymbol{F}}(\boldsymbol{M}_p) = \mathcal{E}_{\boldsymbol{Z}\,|\,\boldsymbol{M}_p}\left\{-\frac{\partial^2\Lambda(\boldsymbol{Z}\,|\,\boldsymbol{M}_p)}{\partial\tilde{\boldsymbol{m}}_p\partial\tilde{\boldsymbol{m}}_p^T}\right\}. \qquad (9)
$$

Observe that $\tilde{\boldsymbol{m}}_p = \boldsymbol{J}\boldsymbol{m}_p$ with $\boldsymbol{J}$ the (invertible) Jacobian matrix. It is straightforward to show that

$$
\boldsymbol{F}(\boldsymbol{M}_p) = \mathcal{E}_{\boldsymbol{Z}\,|\,\boldsymbol{M}_p}\left\{-\frac{\partial^2\Lambda(\boldsymbol{Z}\,|\,\boldsymbol{M}_p)}{\partial\boldsymbol{m}_p\,\partial\boldsymbol{m}_p{}^H}\right\} = \boldsymbol{J}^H\tilde{\boldsymbol{F}}(\boldsymbol{M}_p)\boldsymbol{J}. \qquad (10)
$$

For mathematical convenience, we will derive the matrix $\boldsymbol{F}(\boldsymbol{M}_p)$ in (10) and, with a slight abuse of language, refer to it as the FIM in the sequel. Herein, we define the derivative with respect to a complex scalar $x = x_R + ix_I$ as $\partial/\partial x \triangleq (1/2)[\partial/\partial x_R + i\partial/\partial x_I]$. Differentiating $\Lambda(\boldsymbol{Z}\,|\,\boldsymbol{M}_p)$ with respect to $\boldsymbol{M}_p$ yields the following result:

$$
\begin{aligned}
\frac{\partial\Lambda(\boldsymbol{Z}\,|\,\boldsymbol{M}_p)}{\partial\boldsymbol{M}_p} &= \left(\sum_{k=1}^{K}\nu_k\right)\boldsymbol{M}_p{}^{-1} \\
&- \sum_{k=1}^{K}(\nu_k + L_k)(\nu_k - m)\left[(\nu_k - m)\boldsymbol{M}_p + \boldsymbol{Z}_k\boldsymbol{Z}_k{}^H\right]^{-1}. \qquad (11)
\end{aligned}
$$

In order to differentiate (11), we use the fact that

$$
\frac{\partial\boldsymbol{M}_p{}^{-1}}{\partial\boldsymbol{M}_p{}^*(k,\ell)} = -\boldsymbol{M}_p{}^{-1}\frac{\partial\boldsymbol{M}_p}{\partial\boldsymbol{M}_p{}^*(k,\ell)}\boldsymbol{M}_p{}^{-1}.
$$

Accordingly, since $\boldsymbol{M}_p$ is Hermitian, for any two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$

$$
\begin{aligned}
&\left[\boldsymbol{A}\frac{\partial\boldsymbol{M}_p}{\partial\boldsymbol{M}_p{}^*(k,\ell)}\boldsymbol{B}\right]_{i,j} \\
&= \sum_{p,q=1}^{m}\boldsymbol{A}_{i,p}\left[\frac{\partial\boldsymbol{M}_p}{\partial\boldsymbol{M}_p{}^*(k,\ell)}\right]_{p,q}\boldsymbol{B}_{q,j} \\
&= \boldsymbol{A}_{i,k}\boldsymbol{B}_{\ell,j} = \left[\boldsymbol{B}^T\otimes A\right]_{i+(j-1)m,k+(\ell-1)m}
\end{aligned}
$$

where $\otimes$ stands for the Kronecker product [9]. Using these results, it is straightforward to show that

$$
\begin{aligned}
&\frac{\partial^2\Lambda(\boldsymbol{Z}\,|\,\boldsymbol{M}_p)}{\partial\boldsymbol{m}_p\,\partial\boldsymbol{m}_p{}^H} \\
&= -\left(\sum_{k=1}^{K}\nu_k\right)\boldsymbol{M}_p{}^{-T}\otimes\boldsymbol{M}_p{}^{-1} \\
&\quad + \sum_{k=1}^{K}(\nu_k + L_k)(\nu_k - m)^2 \\
&\quad \times\left[(\nu_k - m)\boldsymbol{M}_p + \boldsymbol{Z}_k\boldsymbol{Z}_k{}^H\right]^{-T} \\
&\quad \otimes\left[(\nu_k - m)\boldsymbol{M}_p + \boldsymbol{Z}_k\boldsymbol{Z}_k{}^H\right]^{-1}. \qquad (12)
\end{aligned}
$$

For the sake of notational convenience, let us introduce

$$
\tilde{\boldsymbol{Z}}_k = (\nu_k - m)^{-1/2}\boldsymbol{M}_p{}^{-1/2}\boldsymbol{Z}_k \qquad (13)
$$

$$
\tilde{\boldsymbol{B}}_k = \left(\boldsymbol{I}_m + \tilde{\boldsymbol{Z}}_k\tilde{\boldsymbol{Z}}_k^H\right)^{-1} \qquad (14)
$$

and note that

$$
\left[(\nu_k - m)\boldsymbol{M}_p + \boldsymbol{Z}_k\boldsymbol{Z}_k{}^H\right]^{-1} = (\nu_k - m)^{-1}\boldsymbol{M}_p{}^{-1/2}\tilde{\boldsymbol{B}}_k\boldsymbol{M}_p{}^{-1/2}. \qquad (15)
$$

Therefore, we can write

$$
\begin{aligned}
&\frac{\partial^2\Lambda(\boldsymbol{Z}\,|\,\boldsymbol{M}_p)}{\partial\boldsymbol{m}_p\,\partial\boldsymbol{m}_p{}^H} \\
&= -\left(\sum_{k=1}^{K}\nu_k\right)\boldsymbol{M}_p{}^{-T}\otimes\boldsymbol{M}_p{}^{-1} \\
&\quad + \sum_{k=1}^{K}(\nu_k + L_k)\left[\boldsymbol{M}_p{}^{-T/2}\otimes\boldsymbol{M}_p{}^{-1/2}\right] \\
&\quad \times\left[\tilde{\boldsymbol{B}}_k^T\otimes\tilde{\boldsymbol{B}}_k\right]\left[\boldsymbol{M}_p{}^{-T/2}\otimes\boldsymbol{M}_p{}^{-1/2}\right]. \qquad (16)
\end{aligned}
$$

In order to derive the FIM, we need to evaluate the statistical mean of $\tilde{\boldsymbol{B}}_k^T\otimes\tilde{\boldsymbol{B}}_k$. Towards this end, we first note that $\tilde{\boldsymbol{Z}}_k$ has a complex multivariate t distribution with $\nu_k + L_k - m$ degrees of freedom [7], i.e.,

$$
\begin{aligned}
f(\tilde{\boldsymbol{Z}}_k\,|\,\boldsymbol{M}_p) = &\frac{\tilde{\Gamma}_{L_k}(\nu_k + L_k)}{\pi^{mL_k}\tilde{\Gamma}_{L_k}(\nu_k + L_k - m)} \\
&\times|\boldsymbol{I}_{L_k} + \tilde{\boldsymbol{Z}}_k^H\tilde{\boldsymbol{Z}}_k|^{-(\nu_k+L_k)}. \qquad (17)
\end{aligned}
$$

It follows that $\tilde{\boldsymbol{B}}_k$, conditionally to $\boldsymbol{M}_p$, has a multivariate beta distribution, with (ν_k, L_k) degrees of freedom [7], [10]. i.e., $\tilde{\boldsymbol{B}}_k\,|\,\boldsymbol{M}_p \sim \tilde{\mathcal{B}}_m(\nu_k, L_k)$. Now, we make use of the following result. Let $\boldsymbol{B}$ be distributed as $\boldsymbol{B} \sim \tilde{\mathcal{B}}_r(p,q)$ with $p + q \geq r$. Then, for any matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ [11]–[13]

$$
\begin{aligned}
&\mathcal{E}\{\text{Tr}\{\boldsymbol{A}_1\boldsymbol{B}\boldsymbol{A}_2\boldsymbol{B}\}\} \\
&\quad = \frac{p}{p+q}\left[\frac{p(p+q)-1}{(p+q)^2-1}\text{Tr}\{\boldsymbol{A}_1\boldsymbol{A}_2\} + \frac{q}{(p+q)^2-1}\text{Tr}\{\boldsymbol{A}_1\}\text{Tr}\{\boldsymbol{A}_2\}\right]. \qquad (18)
\end{aligned}
$$

Let $\boldsymbol{e}_i$ denote the vector whose elements are all zero, except the ith element which equals 1. Accordingly, let us note $\boldsymbol{E}_{ij} = \boldsymbol{e}_i\boldsymbol{e}_j^T$. Then, using (18), one can obtain the $(i+(j-1)m, n+(\ell-1)m)$ element of $\tilde{\boldsymbol{B}}_k^T \otimes \tilde{\boldsymbol{B}}_k$ as

$$\begin{aligned}
&\mathcal{E}\{\tilde{\boldsymbol{B}}_k(\ell,j)\tilde{\boldsymbol{B}}_k(i,n)\}\\
&\quad= \mathcal{E}\left\{\boldsymbol{e}_\ell^T\tilde{\boldsymbol{B}}_k\boldsymbol{e}_j\boldsymbol{e}_i^T\tilde{\boldsymbol{B}}_k\boldsymbol{e}_n\right\}\\
&\quad= \mathcal{E}\left\{\mathrm{Tr}\{\boldsymbol{E}_{n\ell}\tilde{\boldsymbol{B}}_k\boldsymbol{E}_{ji}\tilde{\boldsymbol{B}}_k\}\right\}\\
&\quad= \frac{\nu_k}{\nu_k+L_k}\left[\frac{\nu_k(\nu_k+L_k)-1}{(\nu_k+L_k)^2-1}\mathrm{Tr}\{\boldsymbol{E}_{n\ell}\boldsymbol{E}_{ji}\}\right.\\
&\qquad\left.+\frac{L_k}{(\nu_k+L_k)^2-1}\mathrm{Tr}\{\boldsymbol{E}_{n\ell}\}\mathrm{Tr}\{\boldsymbol{E}_{ji}\}\right]\\
&\quad= \frac{\nu_k}{\nu_k+L_k}\left[\frac{\nu_k(\nu_k+L_k)-1}{(\nu_k+L_k)^2-1}\delta_{\ell,j}\delta_{i,n}+\frac{L_k}{(\nu_k+L_k)^2-1}\delta_{i,j}\delta_{\ell,n}\right].
\end{aligned} \tag{19}$$

It follows that

$$\mathcal{E}\left\{\tilde{\boldsymbol{B}}_k^T\otimes\tilde{\boldsymbol{B}}_k\right\} = \frac{\nu_k}{\nu_k+L_k}\left[\frac{\nu_k(\nu_k+L_k)-1}{(\nu_k+L_k)^2-1}\boldsymbol{I}_{m^2} + \frac{L_k}{(\nu_k+L_k)^2-1}\boldsymbol{e}\boldsymbol{e}^T\right] \tag{20}$$

where $\boldsymbol{e} = [\boldsymbol{e}_1^T \ \ \ldots \ \ \boldsymbol{e}_m^T]^T = \mathrm{vec}(\boldsymbol{I}_m)$. Consequently, the FIM can be expressed as

$$\boldsymbol{F}(\boldsymbol{M}_p) = \left[{\boldsymbol{M}_p}^{-T/2}\otimes{\boldsymbol{M}_p}^{-1/2}\right] \times\left[\alpha\boldsymbol{I}+\beta\boldsymbol{e}\boldsymbol{e}^T\right]\left[{\boldsymbol{M}_p}^{-T/2}\otimes{\boldsymbol{M}_p}^{-1/2}\right] \tag{21}$$

with

$$\begin{aligned}
\alpha &= \sum_{k=1}^K \nu_k - \nu_k\frac{\nu_k(\nu_k+L_k)-1}{(\nu_k+L_k)^2-1}\\
&= \sum_{k=1}^K \frac{\nu_k L_k(\nu_k+L_k)}{(\nu_k+L_k)^2-1}
\end{aligned} \tag{22}$$

$$\beta = -\sum_{k=1}^K \frac{\nu_k L_k}{(\nu_k+L_k)^2-1}. \tag{23}$$

It ensues that the Cramér–Rao bound can be written as

$$\begin{aligned}
\mathrm{CRB} &= \left[{\boldsymbol{M}_p}^{T/2}\otimes{\boldsymbol{M}_p}^{1/2}\right]\\
&\quad\times\left[\alpha\boldsymbol{I}+\beta\boldsymbol{e}\boldsymbol{e}^T\right]^{-1}\left[{\boldsymbol{M}_p}^{T/2}\otimes{\boldsymbol{M}_p}^{1/2}\right]\\
&= \alpha^{-1}\left[{\boldsymbol{M}_p}^{T/2}\otimes{\boldsymbol{M}_p}^{1/2}\right]\\
&\quad\times\left[\boldsymbol{I}-\frac{\beta\boldsymbol{e}\boldsymbol{e}^T}{\alpha+m\beta}\right]\left[{\boldsymbol{M}_p}^{T/2}\otimes{\boldsymbol{M}_p}^{1/2}\right]\\
&= \alpha^{-1}\left\{\left[{\boldsymbol{M}_p}^{T}\otimes\boldsymbol{M}_p\right]\right.\\
&\quad\left.-\frac{\beta}{\alpha+m\beta}\mathrm{vec}(\boldsymbol{M}_p)\mathrm{vec}\left({\boldsymbol{M}_p}^T\right)^T\right\}
\end{aligned} \tag{24}$$

where we have used the fact that $(\boldsymbol{A}\otimes\boldsymbol{B})\boldsymbol{e} = (\boldsymbol{A}\otimes\boldsymbol{B})\mathrm{vec}(\boldsymbol{I}) = \mathrm{vec}(\boldsymbol{B}\boldsymbol{A}^T)$ [9]. The MSE of any estimate $\hat{\boldsymbol{M}}_p$ of $\boldsymbol{M}_p$, $\mathcal{E}_{\boldsymbol{Z}|\boldsymbol{M}_p}\{\|\hat{\boldsymbol{M}}_p-\boldsymbol{M}_p\|^2\}$, is thus lower bounded by

$$\mathrm{Tr}\{\boldsymbol{F}(\boldsymbol{M}_p)^{-1}\} = \alpha^{-1}\left[\mathrm{Tr}\{\boldsymbol{M}_p\}^2 - \frac{\beta}{\alpha+m\beta}\mathrm{Tr}\{{\boldsymbol{M}_p}^2\}\right]. \tag{25}$$

Equation (25) provides a lower bound for the MSE of any estimator of $\boldsymbol{M}_p$, when $\boldsymbol{M}_p$ is a deterministic matrix. Some insights into the properties of the CRB can be gained by considering special cases.

1) Consider first the case $K=1$ and, for the sake of convenience, let us note $L=L_1$ and $\nu=\nu_1$. In this case, there are L snapshots, all sharing the same covariance matrix $\boldsymbol{M}_s=\boldsymbol{M}_1$, and the latter has an inverse Wishart prior, centered around $\boldsymbol{M}_p$, with ν degrees of freedom. Under this framework, it is straightforward to show that (25) reduces to

$$\begin{aligned}
\mathrm{Tr}\{\boldsymbol{F}(\boldsymbol{M}_p)^{-1}\} &= \frac{(\nu+L)^2-1}{\nu(\nu+L)L}\left[\mathrm{Tr}\{\boldsymbol{M}_p\}^2\right.\\
&\qquad\left.+\ (\nu+L-m)^{-1}\mathrm{Tr}\{{\boldsymbol{M}_p}^2\}\right]\\
&\simeq \frac{1}{\nu}\mathrm{Tr}\{\boldsymbol{M}_p\}^2 \text{ when } L\to\infty\\
&\simeq \frac{1}{L}\mathrm{Tr}\{\boldsymbol{M}_p\}^2 \text{ when } \nu\to\infty.
\end{aligned}$$

Two important observations can be made. First, note that, for finite ν, the lower bound does not go to zero but instead converges to $\nu^{-1}\mathrm{Tr}\{\boldsymbol{M}_p\}^2$. Therefore, consistent estimation of $\boldsymbol{M}_p$ is not possible within this framework. This phenomenon can be explained as follows. The snapshots $\boldsymbol{Z}$ provide information about $\boldsymbol{M}_s$, and we can expect them to provide accurate estimates of this matrix. However, $\boldsymbol{M}_s$ is randomly distributed "around" $\boldsymbol{M}_p$ and [6]

$$\begin{aligned}
&\mathcal{E}_{\boldsymbol{Z}|\boldsymbol{M}_p}\left\{\|\boldsymbol{M}_s-\boldsymbol{M}_p\|^2\right\}\\
&\quad= \frac{(\nu-m)\mathrm{Tr}\{\boldsymbol{M}_p\}^2+\mathrm{Tr}\{{\boldsymbol{M}_p}^2\}}{(\nu-m+1)(\nu-m-1)}\\
&\quad\simeq \frac{1}{\nu}\mathrm{Tr}\{\boldsymbol{M}_p\}^2\left[1+\frac{m(\nu-m)+1}{(\nu-m+1)(\nu-m-1)}\right].
\end{aligned}$$

Therefore $\nu^{-1}\mathrm{Tr}\{\boldsymbol{M}_p\}^2$ corresponds to the minimum distance between $\boldsymbol{M}_s$ and $\boldsymbol{M}_p$, and hence the "least" uncertainty that we can obtain when estimating $\boldsymbol{M}_p$ from $\boldsymbol{Z}$. The second point to be noted is that, when ν increases, the lower bound is inversely proportional to L. We recover here the well-known fact that, in a homogeneous environment, the CRB is inversely proportional to the number of snapshots.

2) Let us now consider the case of most interest to us, namely $L_k=1$, i.e., there are K snapshots with K different covariance matrices. For the sake of simplicity, let us assume that $\nu_k=\nu$, $\forall k=1,\ldots,K$. Then, the trace of the CRB becomes

$$\begin{aligned}
\mathrm{Tr}\{\boldsymbol{F}(\boldsymbol{M}_p)^{-1}\} &= \frac{\nu+2}{(\nu+1)K}\left[\mathrm{Tr}\{\boldsymbol{M}_p\}^2\right.\\
&\qquad\left.+\ (\nu+1-m)^{-1}\mathrm{Tr}\{{\boldsymbol{M}_p}^2\}\right]\\
&\xrightarrow[K\to\infty]{} 0\\
&\simeq \frac{1}{K}\mathrm{Tr}\{\boldsymbol{M}_p\}^2 \text{ when } \nu\to\infty.
\end{aligned}$$

An important observation follows from this result: in contrast to the preceding case, the CRB now goes to zero as the number of snapshots goes to infinity; therefore consistent estimation of $\boldsymbol{M}_p$ is possible, even for finite ν. This can be explained by the "diversity" effect. Indeed, when all snapshots have the same covariance matrix $\boldsymbol{M}_s$, they more or less provide the same "view" of $\boldsymbol{M}_p$ (we can think of $\boldsymbol{M}_s$ as a given point in the space of $m\times m$ Hermitian matrices, around $\boldsymbol{M}_p$). In contrast, when $L_k=1$, each snapshot provides a different point of view of $\boldsymbol{M}_p$, and this diversity can be exploited advantageously to yield consistent estimation of $\boldsymbol{M}_p$. Therefore, for a given number of snapshots, the case $L_k=1$ is a more favorable situation than the case $K = 1$. For large ν, however, the same CRB is obtained.

III. Bayesian Bound (Random $\boldsymbol{M}_p$)

We now assume that $\boldsymbol{M}_p$ is distributed according to a Wishart distribution with mean $\bar{\boldsymbol{M}}_p$ and μ degrees of freedom, see (6). The Bayesian bound is obtained as the inverse of the information matrix, which is given by [8]

$$\begin{aligned}\boldsymbol{F}_B &= \mathcal{E}_{\boldsymbol{Z},\boldsymbol{M}_p}\left\{-\frac{\partial^2\Lambda(\boldsymbol{Z},\boldsymbol{M}_p)}{\partial\boldsymbol{m}_p\,\partial\boldsymbol{m}_p{}^H}\right\}\\ &= \mathcal{E}_{\boldsymbol{Z},\boldsymbol{M}_p}\left\{-\frac{\partial^2\Lambda(\boldsymbol{Z}|\boldsymbol{M}_p)}{\partial\boldsymbol{m}_p\,\partial\boldsymbol{m}_p{}^H}-\frac{\partial^2\Lambda(\boldsymbol{M}_p)}{\partial\boldsymbol{m}_p\,\partial\boldsymbol{m}_p{}^H}\right\}\\ &= \mathcal{E}_{\boldsymbol{M}_p}\left\{\mathcal{E}_{\boldsymbol{Z}|\boldsymbol{M}_p}\left\{-\frac{\partial^2\Lambda(\boldsymbol{Z}\,|\,\boldsymbol{M}_p)}{\partial\boldsymbol{m}_p\,\partial\boldsymbol{m}_p{}^H}\right\}-\frac{\partial^2\Lambda(\boldsymbol{M}_p)}{\partial\boldsymbol{m}_p\,\partial\boldsymbol{m}_p{}^H}\right\}\\ &= \mathcal{E}_{\boldsymbol{M}_p}\left\{\boldsymbol{F}(\boldsymbol{M}_p)+(\mu-m)\boldsymbol{M}_p{}^{-T}\otimes\boldsymbol{M}_p{}^{-1}\right\}\end{aligned}\tag{26}$$

since, from (6), we have

$$\frac{\partial\Lambda(\boldsymbol{M}_p)}{\partial\boldsymbol{M}_p} = (\mu-m)\boldsymbol{M}_p{}^{-1}-\mu\bar{\boldsymbol{M}}_p^{-1}\tag{27a}$$

$$\frac{\partial^2\Lambda(\boldsymbol{M}_p)}{\partial\boldsymbol{m}_p\,\partial\boldsymbol{m}_p{}^H} = -(\mu-m)\boldsymbol{M}_p{}^{-T}\otimes\boldsymbol{M}_p{}^{-1}.\tag{27b}$$

The information matrix is thus the average value, with respect to the prior distribution $f(\boldsymbol{M}_p)$, of

$$\begin{aligned}\boldsymbol{F}'(\boldsymbol{M}_p) &= \boldsymbol{F}(\boldsymbol{M}_p)+(\mu-m)M_p{}^{-T}\otimes\boldsymbol{M}_p{}^{-1}\\ &= \left[\boldsymbol{M}_p{}^{-T/2}\otimes\boldsymbol{M}_p{}^{-1/2}\right]\\ &\quad\times\left[\alpha'\boldsymbol{I}+\beta\boldsymbol{e}\boldsymbol{e}^T\right]\left[\boldsymbol{M}_p{}^{-T/2}\otimes\boldsymbol{M}_p{}^{-1/2}\right]\\ &= \alpha'\left[\boldsymbol{M}_p{}^{-T}\otimes\boldsymbol{M}_p{}^{-1}\right]\\ &\quad+\beta\left[\boldsymbol{M}_p{}^{-T/2}\otimes\boldsymbol{M}_p{}^{-1/2}\right]\boldsymbol{e}\boldsymbol{e}^T\left[\boldsymbol{M}_p{}^{-T/2}\otimes\boldsymbol{M}_p{}^{-1/2}\right]\\ &= \alpha'\left[\boldsymbol{M}_p{}^{-T}\otimes\boldsymbol{M}_p{}^{-1}\right]\\ &\quad+\beta\mathrm{vec}\left(\boldsymbol{M}_p{}^{-1}\right)\mathrm{vec}\left(\boldsymbol{M}_p{}^{-T}\right)^T\end{aligned}\tag{28}$$

with $\alpha' = \alpha+\mu-m$. Let us now evaluate the average value of each term in the previous equation. The $(i+(j-1)m, k+(\ell-1)m)$ element of $\mathcal{E}_{\boldsymbol{M}_p}\{\boldsymbol{M}_p{}^{-T}\otimes\boldsymbol{M}_p{}^{-1}\}$ is [6]

$$\begin{aligned}&\mathcal{E}\left\{\boldsymbol{M}_p{}^{-1}(\ell,j)\boldsymbol{M}_p{}^{-1}(i,k)\right\}\\ &\quad= \mathcal{E}\left\{\mathrm{Tr}\left\{\boldsymbol{E}_{ji}\boldsymbol{M}_p{}^{-1}\boldsymbol{E}_{k\ell}\boldsymbol{M}_p{}^{-1}\right\}\right\}\\ &\quad= \mathrm{Tr}\left\{\boldsymbol{E}_{ji}\mathcal{E}\left\{\boldsymbol{M}_p{}^{-1}\boldsymbol{E}_{k\ell}\boldsymbol{M}_p{}^{-1}\right\}\right\}\\ &\quad= \frac{\mu^2(\mu-m)\mathrm{Tr}\left\{\boldsymbol{E}_{ji}\bar{\boldsymbol{M}}_p^{-1}\boldsymbol{E}_{k\ell}\bar{\boldsymbol{M}}_p^{-1}\right\}}{(\mu-m+1)(\mu-m)(\mu-m-1)}\\ &\qquad+\frac{\mu^2\mathrm{Tr}\left\{\boldsymbol{E}_{ji}\bar{\boldsymbol{M}}_p^{-1}\right\}\mathrm{Tr}\left\{\boldsymbol{E}_{k\ell}\bar{\boldsymbol{M}}_p^{-1}\right\}}{(\mu-m+1)(\mu-m)(\mu-m-1)}\\ &\quad= \frac{\mu^2(\mu-m)\bar{\boldsymbol{M}}_p^{-1}(i,k)\bar{\boldsymbol{M}}_p^{-1}(\ell,j)+\mu^2\bar{\boldsymbol{M}}_p^{-1}(i,j)\bar{\boldsymbol{M}}_p^{-1}(\ell,k)}{(\mu-m+1)(\mu-m)(\mu-m-1)}.\end{aligned}\tag{29}$$

Observing that the $(i+(j-1)m, k+(\ell-1)m)$ elements of $\boldsymbol{A}\otimes\boldsymbol{B}$ and $\mathrm{vec}(\boldsymbol{A})\mathrm{vec}(\boldsymbol{B})^T$ are $\boldsymbol{A}(j,\ell)\boldsymbol{B}(i,k)$ and $\boldsymbol{A}(i,j)\boldsymbol{B}(k,\ell)$, it follows that

$$\begin{aligned}&\mathcal{E}_{\boldsymbol{M}_p}\left\{\boldsymbol{M}_p{}^{-T}\otimes\boldsymbol{M}_p{}^{-1}\right\}\\ &\quad= \frac{\mu^2(\mu-m)\bar{\boldsymbol{M}}_p^{-T}\otimes\bar{\boldsymbol{M}}_p^{-1}+\mu^2\mathrm{vec}\left(\bar{\boldsymbol{M}}_p^{-1}\right)\mathrm{vec}\left(\bar{\boldsymbol{M}}_p^{-T}\right)^T}{(\mu-m+1)(\mu-m)(\mu-m-1)}\\ &\quad= \mu^2\frac{\left[\bar{\boldsymbol{M}}_p^{-T/2}\otimes\bar{\boldsymbol{M}}_p^{-1/2}\right]\left[(\mu-m)\boldsymbol{I}+\boldsymbol{e}\boldsymbol{e}^T\right]\left[\bar{\boldsymbol{M}}_p^{-T/2}\otimes\bar{\boldsymbol{M}}_p^{-1/2}\right]}{(\mu-m+1)(\mu-m)(\mu-m-1)}.\end{aligned}\tag{30}$$

Using similar arguments, it can be shown that

$$\begin{aligned}&\mathcal{E}_{\boldsymbol{M}_p}\left\{\mathrm{vec}\left(\boldsymbol{M}_p{}^{-1}\right)\mathrm{vec}\left(\boldsymbol{M}_p{}^{-T}\right)^T\right\}\\ &\quad= \mu^2\frac{\left[\bar{\boldsymbol{M}}_p^{-T/2}\otimes\bar{\boldsymbol{M}}_p^{-1/2}\right]\left[\boldsymbol{I}+(\mu-m)\boldsymbol{e}\boldsymbol{e}^T\right]\left[\bar{\boldsymbol{M}}_p^{-T/2}\otimes\bar{\boldsymbol{M}}_p^{-1/2}\right]}{(\mu-m+1)(\mu-m)(\mu-m-1)}.\end{aligned}\tag{31}$$

Gathering the previous results, we end up with the following expression:

$$\boldsymbol{F}_B = \left[\bar{\boldsymbol{M}}_p^{-T/2}\otimes\bar{\boldsymbol{M}}_p^{-1/2}\right]\left[\alpha''\boldsymbol{I}+\beta''\boldsymbol{e}\boldsymbol{e}^T\right]\left[\bar{\boldsymbol{M}}_p^{-T/2}\otimes\bar{\boldsymbol{M}}_p^{-1/2}\right]\tag{32}$$

where

$$\begin{aligned}\alpha'' &= \frac{\mu^2\left[\alpha'(\mu-m)+\beta\right]}{(\mu-m+1)(\mu-m)(\mu-m-1)}\\ \beta'' &= \frac{\mu^2\left[\alpha'+(\mu-m)\beta\right]}{(\mu-m+1)(\mu-m)(\mu-m-1)}.\end{aligned}\tag{33}$$

The Bayesian bound is obtained as the inverse of $\boldsymbol{F}_B$, which yields

$$BB = \alpha''^{-1}\left\{\left[\bar{\boldsymbol{M}}_p^T\otimes\bar{\boldsymbol{M}}_p\right]-\frac{\beta''}{\alpha''+m\beta''}\mathrm{vec}\left(\bar{\boldsymbol{M}}_p\right)\mathrm{vec}\left(\bar{\boldsymbol{M}}_p^T\right)^T\right\}.\tag{34}$$

Finally, under the assumption that $\boldsymbol{M}_p$ has a Wishart prior, the MSE of any estimator of $\boldsymbol{M}_p$ is lower-bounded by the following BB trace

$$\begin{aligned}&\mathrm{Tr}\left\{\boldsymbol{F}_B^{-1}\right\}\\ &\quad= \alpha''^{-1}\left[\mathrm{Tr}\{\bar{\boldsymbol{M}}_p\}^2-\frac{\beta''}{\alpha''+m\beta''}\mathrm{Tr}\left\{\bar{\boldsymbol{M}}_p^2\right\}\right]\\ &\quad= \frac{(\mu-m+1)(\mu-m)(\mu-m-1)}{\mu^2\left[(\mu-m)^2+\alpha(\mu-m)+\beta\right]}\left[\mathrm{Tr}\{\bar{\boldsymbol{M}}_p\}^2\right.\\ &\qquad\left.-\frac{\alpha+(\mu-m)(1+\beta)}{(\mu-m)^2+(\mu-m)\left[\alpha+m(1+\beta)\right]+m\alpha+\beta}\mathrm{Tr}\{\bar{\boldsymbol{M}}_p^2\}\right].\end{aligned}\tag{35}$$

The BB of $\boldsymbol{M}_p$ depends on μ and $\bar{\boldsymbol{M}}_p$, as expected. However, one can observe the similarity between (25) and (35). Note also that the lower bound in (35) depends on $\bar{\boldsymbol{M}}_p$ only through $\mathrm{Tr}\{\bar{\boldsymbol{M}}_p\}^2$ and $\mathrm{Tr}\{\bar{\boldsymbol{M}}_p^2\}$.

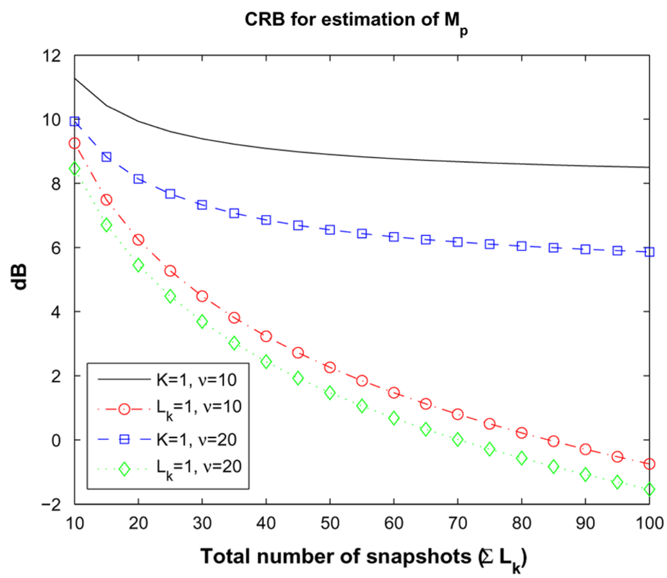

Fig. 1. Cramér-Rao bound versus number of snapshots.

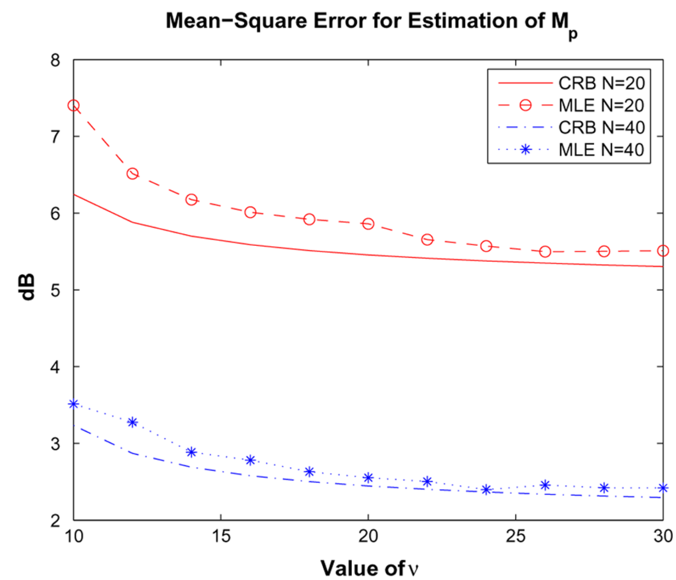

Fig. 3. Cramér–Rao bound and MSE of the MLE versus ν—$L_k = 1$.

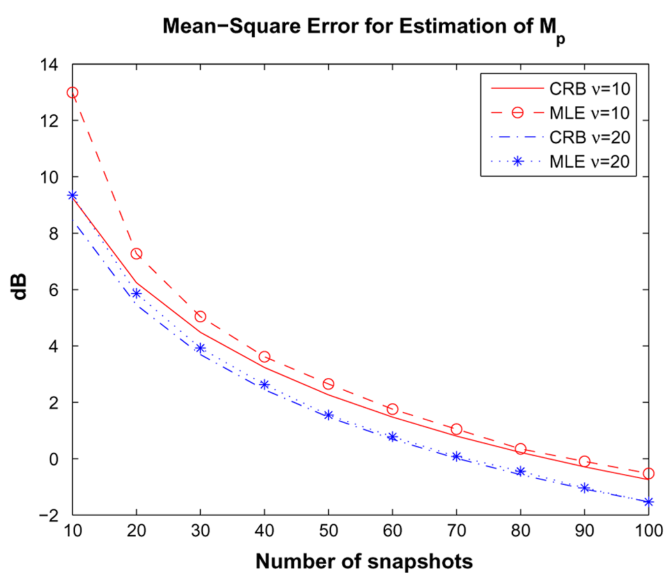

Fig. 2. Cramér–Rao bound and MSE of the MLE versus number of snapshots—$L_k = 1$.

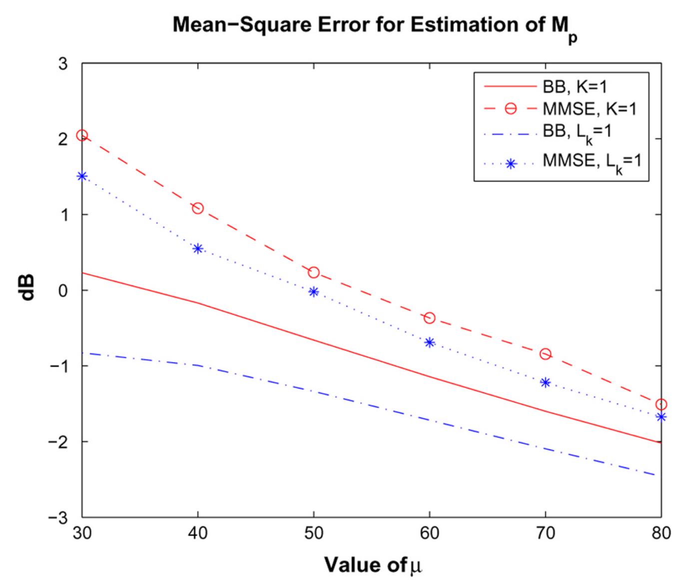

Fig. 4. Bayesian bound and MSE of the MMSE estimator versus μ—$N = 20$ and $\nu = 20$.

IV. Numerical Illustrations

In this section, we provide numerical illustrations of the CRB and BB properties. First, we contrast the behavior of the CRB in the two opposite cases, namely $K = 1$ and $L_k = 1$. For the sake of simplicity, when $L_k = 1$, we assume that all ν_k's are equal to a common value denoted as ν. Whatever the case, $N = \sum_{k=1}^{K} L_k$ denotes the total number of snapshots. In all simulations, the size of the observation space is $m = 8$. When considering the CRB the true covariance matrix is given by $\boldsymbol{M}_p(k,\ell) = 0.9^{|k-\ell|}$, while $\bar{\boldsymbol{M}}_p(k,\ell) = 0.9^{|k-\ell|}$ when $\boldsymbol{M}_p$ is assumed to be random. The matrices $\boldsymbol{M}_k$ were generated according to the inverse Wishart distribution of (2). In practice, the $\boldsymbol{M}_k$ are generated as $\boldsymbol{M}_k = (\boldsymbol{G}_k \boldsymbol{G}_k^H)^{-1}$ where $\boldsymbol{G}_k \in \mathbb{C}^{m \times \nu}$ is drawn from a zero-mean multivariate Gaussian distribution with covariance matrix $(\nu_k - m)^{-1} \boldsymbol{M}_p{}^{-1}$.

In Fig. 1, we display the CRB versus the total number of snapshots N, for two different values of ν, namely $\nu = 10$ and $\nu = 20$. This figure confirms the observations made previously. When $L_k = 1$, the CRB decreases nearly linearly with the number of snapshots, while for $K = 1$ we can observe a threshold effect, i.e., the CRB does no longer decrease when the number of snapshots increases. It can also be seen that the CRB decreases when ν increases, i.e., as the environment is more homogeneous. However, this improvement is more pronounced when $K = 1$ than when $L_k = 1$, which seems logical.

Next, we compare the performance of the MLE derived in [4] with the CRB, in the case $L_k = 1$. Figs. 2 and 3 consider the influence of the number of snapshots and ν, respectively. From inspection of these figures, it can be seen that the MLE has a performance quite close to the CRB. The difference between the two is smaller as either K or ν increases.

Finally, we provide illustrations of the BB properties. In Fig. 4 we contrast the trace of the BB for the two cases $K = 1$ and $L_k = 1$, and we study the influence of μ which rules the degree of *a priori* knowledge about $\bar{\boldsymbol{M}}_p$. In this figure, we also display the MSE of the MMSE estimator derived in [4]. The total number of snapshots is $N = 20$ and $\nu = 20$. As can be observed, for a given number of snapshots, the BB is smaller when $L_k = 1$ than when $K = 1$, which confirms the previous observations made on the CRB. Also, as could be expected, the BB decreases as μ increases, i.e., as the prior is more and more informative. Finally, we note that the MMSE estimator has a MSE close to the BB only for large values of μ.

V. Concluding Remarks

This correspondence derived lower bounds on the estimation of a covariance matrix $\boldsymbol{M}_p$ using heterogeneous samples $\boldsymbol{Z}_k$, $k = 1, \ldots, K$, which have covariance matrices $\boldsymbol{M}_k$ different from $\boldsymbol{M}_p$. When $\boldsymbol{M}_p$ is deterministic, we showed that consistent estimation of $\boldsymbol{M}_p$ is not feasible, when all samples share the same covariance matrix, i.e., when $K = 1$. Indeed, the CRB does not converge to zero as the number of training samples increases. In contrast, if all snapshots have different covariance matrices, randomly distributed around $\boldsymbol{M}_p$ (i.e., $L_k = 1$, for $k = 1, \ldots, K$), the CRB goes to zero when the number of training samples increases. The correspondence also derived the Bayesian bound associated to a random covariance matrix $\boldsymbol{M}_p$. The bounds derived herein enable one to quantify the degradation induced by heterogeneity, and can serve as references for any estimator of the covariance matrix $\boldsymbol{M}_p$.

Acknowledgment

The authors would like to thank Prof. G. Letac for enthusiastically sharing his expert knowledge on multivariate Wishart and beta distributions.

References

[1] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Reading, MA: Addison-Wesley, 1991.
[2] W. L. Melvin, “Space-time adaptive radar performance in heterogeneous clutter,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 621–633, Apr. 2000.
[3] W. L. Melvin, “A STAP overview,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1 , pt. 2, pp. 19–35, Jan. 2004.
[4] O. Besson, S. Bidon, and J.-Y. Tourneret, “Covariance matrix estimation with heterogeneous samples,” *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 909–920, Mar. 2008.
[5] S. Bidon, O. Besson, and J.-Y. Tourneret, “A Bayesian approach to adaptive detection in non-homogeneous environments,” *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 205–217, Jan. 2008.
[6] J. A. Tague and C. I. Caldwell, “Expectations of useful complex Wishart forms,” *Multidimen. Syst. Signal Process.*, vol. 5, pp. 263–279, 1994.
[7] C. G. Khatri and C. R. Rao, “Effects of estimated noise covariance matrix in optimal signal detection,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 5, pp. 671–679, May 1987.
[8] H. L. V. Trees, *Optimum Array Processing*. New York: Wiley, 2002.
[9] H. Lütkepohl, *Handbook of Matrices*. Chichester, U.K.: Wiley, 1996.
[10] C. G. Khatri, “Classical statistical analysis based on a certain multivariate complex Gaussian distribution,” *Ann. Math. Stat.*, vol. 36, no. 1, pp. 98–114, Feb. 1965.
[11] M. Capitaine and M. Casalis, “Asymptotic freeness by generalized moments for Gaussian and Wishart matrices,” *Indiana Univ. Math. J.*, vol. 53, no. 2, pp. 397–431, 2004.
[12] M. Capitaine and M. Casalis, “Cumulants for random matrices as convolutions on the symmetric group,” *Probab. Theory Relat. Fields*, vol. 136, no. 1, pp. 19–36, Sep. 2006.
[13] G. Letac, Expectation of $Z_{ij}Z_{kl}$ for a Matrix Beta Law 2007, private communication.

Cross Entropy Approximation of Structured Gaussian Covariance Matrices

Cheng-Yuan Liou and Bruce R. Musicus

***Abstract*—We apply two variations of the principle of minimum cross entropy (the Kullback information measure) to fit parameterized probability density models to observed data densities. For an array beamforming problem with P incident narrowband point sources, $N > P$ sensors, and colored noise, both approaches yield eigenvector fitting methods similar to that of the MUSIC algorithm and of the oblique transformation in factor analysis. Furthermore, the corresponding cross entropies (CE) are related to the MDL model order selection criterion .**

***Index Terms*—Array beamforming, eigenvector methods, factor analysis, generalized principle component analysis, Kullback information measure, minimum cross entropy (CE), oblique transformation, stochastic estimation, structured covariance.**

I. Introduction

Many existing high resolution methods for spectral analysis and for optimal beamforming utilize covariance matrices estimated from observed data. Often, an underlying structure for the covariance matrix is known in advance, and our goal is to estimate the covariance matrix with this structure which best fits the observed data. Previous literature has suggested a variety of methods of optimally estimating structured covariance matrices from data [1]–[5]. In this correspondence, we will apply the minimum cross entropy (CE) and minimum reverse cross-entropy (RCE) [6] principles to estimate the covariance matrix. These principles have proved to be quite powerful in a wide variety of signal processing applications, such as complex independent component analysis [7], [8], encoding mechanism [9]. They have been justified as being “optimal” under suitable assumptions. In Section II, we apply the CE and RCE procedures to the problem of estimating structured covariance matrices, and in Section III we demonstrate the utility of the idea for a beamforming application.

II. Problem Statement

Let $\underline{x}$ be an N-dimensional real or complex random vector. Assume that a Gaussian probability density for $\underline{x}$ is either known *a priori* or has been estimated by some procedure from observed data

$$p(\underline{x}) = N(\underline{m}, R) \tag{1}$$

where $\underline{m}$ is the expected value of $\underline{x}$, and R is the covariance matrix, $R = E[(\underline{x} - \underline{m})(\underline{x} - \underline{m})^H]$, and where $\underline{x}^H$ is the Hermitian (complex conjugate transpose) of $\underline{x}$. Suppose we wish to approximate this $p(\underline{x})$ with a parameterized probability density function (pdf)

$$q_\theta(\underline{x}) = N(\underline{m}_\theta, R_\theta) \tag{2}$$

where $\underline{\theta}$ denotes the unknown parameters in the model $q_\theta(\underline{x})$ which are to be estimated. Conceptually, we wish to choose $\underline{\theta}$ to make $q_\theta(\underline{x})$ op-

Manuscript received March 25, 2007; revised January 8, 2008. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sven Nordebo.

C.-Y. Liou is with the Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, R.O.C. (e-mail: cyliou@csie.ntu.edu.tw).

B. R. Musicus resides in Boston, MA 02421 USA (e-mail: bmusicus@rcn.com).

Digital Object Identifier 10.1109/TSP.2008.917878