



HAL
open science

Covariance matrix estimation with heterogeneous samples

Olivier Besson, Stéphanie Bidon, Jean-Yves Tournet

► **To cite this version:**

Olivier Besson, Stéphanie Bidon, Jean-Yves Tournet. Covariance matrix estimation with heterogeneous samples. *IEEE Transactions on Signal Processing*, 2008, 5 (3), pp.909-920. 10.1109/TSP.2007.908995 . hal-03610264

HAL Id: hal-03610264

<https://hal.science/hal-03610264v1>

Submitted on 16 Mar 2022

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.

Covariance Matrix Estimation With Heterogeneous Samples

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

Abstract—We consider the problem of estimating the covariance matrix \mathbf{M}_p of an observation vector, using heterogeneous training samples, i.e., samples whose covariance matrices are not exactly \mathbf{M}_p . More precisely, we assume that the training samples can be clustered into K groups, each one containing L_k snapshots sharing the same covariance matrix \mathbf{M}_k . Furthermore, a Bayesian approach is proposed in which the matrices \mathbf{M}_k are assumed to be random with some prior distribution. We consider two different assumptions for \mathbf{M}_p . In a fully Bayesian framework, \mathbf{M}_p is assumed to be random with a given prior distribution. Under this assumption, we derive the minimum mean-square error (MMSE) estimator of \mathbf{M}_p which is implemented using a Gibbs-sampling strategy. Moreover, a simpler scheme based on a weighted sample covariance matrix (SCM) is also considered. The weights minimizing the mean square error (MSE) of the estimated covariance matrix are derived. Furthermore, we consider estimators based on colored or diagonal loading of the weighted SCM, and we determine theoretically the optimal level of loading. Finally, in order to relax the *a priori* assumptions about the covariance matrix \mathbf{M}_p , the second part of the paper assumes that this matrix is deterministic and derives its maximum-likelihood estimator. Numerical simulations are presented to illustrate the performance of the different estimation schemes.

Index Terms—Covariance matrices, estimation, heterogeneous environment, maximum-likelihood estimation, minimum mean-square error (MMSE) estimation, Monte Carlo methods.

I. INTRODUCTION

ESTIMATION of covariance matrices is a fundamental issue in most adaptive detection problems, where a known signal (up to a scaling factor) embedded in noise (whose statistics are unknown) has to be detected in a cell under test (CUT). When the noise is Gaussian distributed and its covariance matrix \mathbf{M}_p is known, the optimal detector consists of a whitening step followed by matched filtering. When the covariance matrix \mathbf{M}_p is unknown, it is natural to estimate it and to substitute $\hat{\mathbf{M}}_p$ for its estimate in the optimal detector [1].

This work was supported in part by the Délégation Générale pour l'Armement (DGA) and by Thales Systèmes Aéroportés. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Erik G. Larsson.

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

J.-Y. Tournet is with IRIT/ENSEEIH, 31071 Toulouse, France (e-mail: jean-yves.tournet@enseeiht.fr).

Consequently, estimation of \mathbf{M}_p is of utmost importance. In space-time adaptive processing (STAP) for radar applications, one usually resorts to training samples obtained from range cells adjacent to the CUT, with the underlying assumption that these samples will more or less share the same statistical properties as the noise in the CUT. In an ideal case, where all training samples are independent, Gaussian distributed with the same covariance matrix \mathbf{M}_p , using the SCM as an estimate results in an adaptive filter whose performance is independent of \mathbf{M}_p and only depends on the number of training samples and the number of array elements [1]–[3]. Indeed, Reed, Mallet, and Brennan have shown that the output signal-to-noise ratio (SNR) of this adaptive filter, normalized to the optimum SNR, has a beta distribution. Moreover, the SNR loss is less than 3 dB as soon as $2m$ training samples are available, where m stands for the dimension of the observation space. In STAP applications, m corresponds to the number of array elements times the number of pulses in a coherent processing interval, and can, thus, be quite large. It then becomes complicated to have m training samples which are completely homogeneous with the CUT.

It is widely recognized that clutter environments are most often heterogeneous [4], [5]. Heterogeneity can stem from the terrain itself, e.g., complex scattering environments (such as urban areas), clutter transitions (e.g., land-sea), presence of discretized in some range cells, etc. It can also be a direct consequence of the array geometry. For example, in the case of a forward-looking radar, it is known that the clutter is distributed along an ellipse in the angle-Doppler plane, and that this ellipse varies with distance, especially at short ranges [6]. Heterogeneous environments adversely affect detection performance, and using the sample covariance matrix of the training samples results in significant performance degradation, if no further processing is used. In order to mitigate the deleterious effects of heterogeneous environments, knowledge-aided (KA) processing has recently gained significant attention [7], [8]. KA processing consists of improving conventional adaptive detection schemes via some *a priori* information, obtained, e.g., from digital elevation maps, synthetic aperture radar imagery. In most studies so far, this *a priori* information is embedded in the covariance matrix of the noise in the CUT. In other words, some approximate value $\bar{\mathbf{M}}_p$ of \mathbf{M}_p is assumed to be available, and then used to design a KA adaptive filter, see, e.g., [9]–[12] and references therein. However, these approaches do not necessarily use a statistical model that would relate the training samples to this *a priori* covariance matrix, with a view to derive an optimal estimator. In such a case, a Bayesian model seems to

be appropriate as it allows one to formulate in a theoretically sound manner such a relation. Covariance matrix estimation in a Bayesian framework is an old topic. However, the application of these results to array processing has recently received much attention [13], [14]. These references present a very comprehensive discussion about the Bayesian framework and solutions for estimating the covariance matrix in homogeneous environments. In [15], see also [16], we proposed a Bayesian approach to model nonhomogeneous environments. More precisely, we assumed that both \mathbf{M}_p and the common covariance matrix of the training samples, say \mathbf{M}_s , were random with some appropriate joint distribution. Additionally, and following KA processing ideas, we assumed that \mathbf{M}_p had a prior distribution with known mean $\bar{\mathbf{M}}_p$. This framework established a clear statistical relation between the training samples and \mathbf{M}_p , which was used to derive optimal estimators of \mathbf{M}_p . Moreover, both the degree of heterogeneity and the level of *a priori* knowledge were tuned through scalar variables, resulting in a very flexible model. This paper extends the results of [15] by considering that the training samples are not homogeneous between themselves, i.e., they do not necessarily share the same covariance matrix. More precisely, we assume that the secondary data can be clustered into K groups characterized by different random covariance matrices \mathbf{M}_k , $k = 1, \dots, K$. These random matrices have prior distributions denoted as $f(\mathbf{M}_k|\mathbf{M}_p)$. Similarly to [15], we first assume that \mathbf{M}_p has a prior distribution with known mean $\bar{\mathbf{M}}_p$. We then extend the results of [15] and propose an extended Gibbs-sampler to obtain the minimum mean-square error (MMSE) estimate of \mathbf{M}_p . Additionally, we derive an optimal SCM-type estimator, with possibly diagonal or colored loading, see below for details. A second approach is also proposed relaxing the previous assumptions on \mathbf{M}_p , and assuming that \mathbf{M}_p is an unknown deterministic matrix. The maximum-likelihood estimator (MLE) of \mathbf{M}_p is then naturally studied.

II. DATA MODEL

As explained previously, we assume that the secondary data can be clustered into K groups, each of them containing L_k samples and sharing the same covariance matrix \mathbf{M}_k . The particular case $K = 1$ corresponds to the scenario considered in [15]. The extreme case $L_k = 1$ for $k = 1, \dots, K$ would correspond to a situation where all training samples would have a different covariance matrix. In the case where heterogeneity is due to the terrain, the snapshots could be segmented according to some *a priori* geographic information, provided, e.g., by digital elevation data or digital land classification data [8], [17]. Let $\{\mathbf{z}_{k,\ell}\}_{\ell=1}^{L_k}$ denote the snapshots of the k th group. We assume that the $\mathbf{z}_{k,\ell}$ are independent and Gaussian distributed, with covariance matrix \mathbf{M}_k , i.e., the distribution of $\mathbf{Z}_k = [\mathbf{z}_{k,1} \cdots \mathbf{z}_{k,L_k}]$, conditionally to \mathbf{M}_k is

$$f(\mathbf{Z}_k|\mathbf{M}_k) = \pi^{-mL_k} |\mathbf{M}_k|^{-L_k} \text{etr} \left\{ -\mathbf{M}_k^{-1} \mathbf{Z}_k \mathbf{Z}_k^H \right\} \quad (1)$$

where \cdot^H denotes the Hermitian transpose, $|\cdot|$ and $\text{etr}\{\cdot\}$ stand for the determinant and the exponential of the trace of the matrix between braces, respectively. We denote this distribution as $\mathbf{Z}_k|\mathbf{M}_k \sim \mathcal{CN}_{m,L_k}(\mathbf{0}, \mathbf{M}_k, \mathbf{I}_{L_k})$ as in [3] to emphasize that the columns of \mathbf{Z}_k are independent. The matrices \mathbf{M}_k are assumed

to be independent conditionally to \mathbf{M}_p . Moreover, they are distributed according to an inverse Wishart distribution with mean \mathbf{M}_p and ν_k degrees of freedom, i.e.,

$$\mathbf{M}_k|\mathbf{M}_p \sim \mathcal{CW}_m^{-1}((\nu_k - m)\mathbf{M}_p, \nu_k). \quad (2)$$

The distribution of \mathbf{M}_k , conditionally to \mathbf{M}_p , is then given by [18]

$$f(\mathbf{M}_k|\mathbf{M}_p) \propto |\mathbf{M}_p|^{\nu_k} |\mathbf{M}_k|^{-(\nu_k+m)} \text{etr} \left\{ -(\nu_k - m)\mathbf{M}_k^{-1} \mathbf{M}_p \right\} \quad (3)$$

where \propto means proportional to. The scalar ν_k allows one to adjust the distance between \mathbf{M}_k and \mathbf{M}_p : the larger ν_k , the closer \mathbf{M}_k to \mathbf{M}_p , see [15] for details and [18] for general results on Wishart and inverse Wishart distributions. In a radar scenario, one could choose increasing ν_k s as the cells corresponding to \mathbf{Z}_k are closer to the CUT. Note that (2) is the conjugate prior of (1), a common choice enabling one to obtain analytical expressions for the posterior distributions. We refer the reader to [14] for a comprehensive discussion about the choice of a prior.

This paper considers two different approaches to model \mathbf{M}_p :

- 1) a fully Bayesian approach in which \mathbf{M}_p is a random matrix whose prior distribution is Wishart, with mean $\bar{\mathbf{M}}_p$ and μ degrees of freedom, i.e.,

$$f(\mathbf{M}_p) \propto |\mathbf{M}_p|^{\mu-m} \text{etr} \left\{ -\mu \mathbf{M}_p \bar{\mathbf{M}}_p^{-1} \right\}. \quad (4)$$

We denote this distribution as $\mathbf{M}_p|\bar{\mathbf{M}}_p \sim \mathcal{CW}_m(\mu^{-1}\bar{\mathbf{M}}_p, \mu)$ and we assume that $\bar{\mathbf{M}}_p$ is known. Observe that the statistical mean of \mathbf{M}_p is $\mathcal{E}\{\mathbf{M}_p\} = \bar{\mathbf{M}}_p$, where $\mathcal{E}\{\cdot\}$ stands for the mathematical expectation, and that the distance between \mathbf{M}_p and $\bar{\mathbf{M}}_p$ decreases as μ increases [18]. However, these two matrices will be different with probability one. The matrix $\bar{\mathbf{M}}_p$ can be obtained from various sources of information, including digital terrain maps or digital land classification data [8], [17]. Accordingly, the clutter covariance matrix model derived by Ward [19] could be used to obtain $\bar{\mathbf{M}}_p$, see [10], [12] for such an approach. Within this framework, we extend the results of [15] and derive the MMSE estimate of \mathbf{M}_p using a Gibbs sampler. Furthermore, we consider a simpler scheme, namely a SCM-type estimator where each matrix $\mathbf{Z}_k \mathbf{Z}_k^H$ is weighted by a possibly different scalar w_k . We then derive the values of w_k that result in a MMSE. Accordingly, we consider estimators based on diagonal or colored loading (with $\bar{\mathbf{M}}_p$ as the loading matrix) of the weighted SCM, and we derive the optimal values of the loading levels.

- 2) a mixed approach relaxing the previous assumptions and assuming that \mathbf{M}_p is a deterministic and unknown matrix.

In this case, the MLE of \mathbf{M}_p is derived.

In the sequel we let $\mathbf{Z} = [\mathbf{Z}_1 \dots \mathbf{Z}_K]$ denote the whole secondary data matrix.

III. BAYESIAN APPROACHES WITH RANDOM \mathbf{M}_p

In this section, we assume that \mathbf{M}_p is a random matrix, whose prior distribution is defined in (4). We successively consider the

MMSE estimator and weighted SCM (with possibly loading) estimators.

A. MMSE Estimation

The MMSE estimator of \mathbf{M}_p is computed from the posterior distribution $f(\mathbf{M}_p|\mathbf{Z})$. We start with the joint posterior distribution of $\mathbf{M}_p, \{\mathbf{M}_k\}$, conditionally to \mathbf{Z} . Under the stated hypotheses, one can write

$$\begin{aligned} f(\mathbf{M}_p, \{\mathbf{M}_k\}|\mathbf{Z}) &\propto f(\mathbf{Z}|\mathbf{M}_p, \{\mathbf{M}_k\}) \left[\prod_{k=1}^K f(\mathbf{M}_k|\mathbf{M}_p) \right] \\ &\quad \times f(\mathbf{M}_p|\bar{\mathbf{M}}_p) \\ &\propto \prod_{k=1}^K |\mathbf{M}_k|^{-L_k} \text{etr} \left\{ -\mathbf{M}_k^{-1} \mathbf{Z}_k \mathbf{Z}_k^H \right\} \\ &\quad \times \prod_{k=1}^K |\mathbf{M}_p|^{\nu_k} |\mathbf{M}_k|^{-(\nu_k+m)} \\ &\quad \times \text{etr} \left\{ -(\nu_k - m) \mathbf{M}_p \mathbf{M}_k^{-1} \right\} |\mathbf{M}_p|^{\mu-m} \\ &\quad \times \text{etr} \left\{ -\mu \mathbf{M}_p \bar{\mathbf{M}}_p^{-1} \right\}. \end{aligned} \quad (5)$$

The posterior distribution $f(\mathbf{M}_p|\mathbf{Z})$ can be obtained by integrating (5) over the $\{\mathbf{M}_k\}$, which yields

$$\begin{aligned} f(\mathbf{M}_p|\mathbf{Z}) &= \int \dots \int f(\mathbf{M}_p, \{\mathbf{M}_k\}|\mathbf{Z}) d\mathbf{M}_1 \dots d\mathbf{M}_K \\ &\propto |\mathbf{M}_p|^{\mu-m+\sum_{k=1}^K \nu_k} \text{etr} \left\{ -\mu \mathbf{M}_p \bar{\mathbf{M}}_p^{-1} \right\} \\ &\quad \times \prod_{k=1}^K \int |\mathbf{M}_k|^{-(\nu_k+L_k+m)} \\ &\quad \times \text{etr} \left\{ -\mathbf{M}_k^{-1} \left[(\nu_k - m) \mathbf{M}_p + \mathbf{Z}_k \mathbf{Z}_k^H \right] \right\} d\mathbf{M}_k \\ &\propto |\mathbf{M}_p|^{\mu-m+\sum_{k=1}^K \nu_k} \text{etr} \left\{ -\mu \mathbf{M}_p \bar{\mathbf{M}}_p^{-1} \right\} \\ &\quad \times \prod_{k=1}^K \left| (\nu_k - m) \mathbf{M}_p + \mathbf{Z}_k \mathbf{Z}_k^H \right|^{-(\nu_k+L_k)} \end{aligned} \quad (6)$$

where, to obtain the last equality of the equation, we used the fact that the integral in the middle of the equation is proportional to the integral of an inverse Wishart distribution with parameter matrix $(\nu_k - m) \mathbf{M}_p + \mathbf{Z}_k \mathbf{Z}_k^H$ and $\nu_k + L_k$ degrees of freedom. The MMSE estimate can in theory be obtained by averaging (6), which amounts to calculating the integral

$$\int \mathbf{M}_p f(\mathbf{M}_p|\mathbf{Z}) d\mathbf{M}_p. \quad (7)$$

Unfortunately, there does not exist any analytical expression for this integral, and, hence, it must be approximated somehow. As we discussed in [15], a natural approximation can be obtained by generating matrices distributed according to $f(\mathbf{M}_p|\mathbf{Z})$ using Markov chain Monte Carlo (MCMC) methods. These matrices are then averaged to approximate (7). The posterior distribution $f(\mathbf{M}_p|\mathbf{Z})$ defined in (6) does not belong to a classical family of distributions. This paper proposes to implement a completion Gibbs sampler (see [20, p. 374] for more details)

which generates matrices \mathbf{M}_p and $\{\mathbf{M}_k\}$ distributed according to the joint posterior distribution $f(\mathbf{M}_p, \{\mathbf{M}_k\}|\mathbf{Z})$. The Gibbs-sampler will successively generate random matrices \mathbf{M}_p and $\{\mathbf{M}_k\}$ distributed according to their full conditional distributions. More precisely, assuming that $\mathbf{M}_p^{(i)}$ and $\mathbf{M}_k^{(i)}$ are available at the i th iteration, the generation of $\mathbf{M}_k^{(i+1)}$ and $\mathbf{M}_p^{(i+1)}$ is achieved as follows:

- for $k = 1 : K$
generate $\mathbf{M}_k^{(i+1)}$ according to $f(\mathbf{M}_k|\mathbf{M}_p^{(i)}, \{\mathbf{M}_\ell^{(i)}\}_{\ell \neq k}, \mathbf{Z})$
end
 - generate $\mathbf{M}_p^{(i+1)}$ according to $f(\mathbf{M}_p|\{\mathbf{M}_k^{(i)}\}, \mathbf{Z})$.
- Using (5), we have

$$\begin{aligned} f(\mathbf{M}_p|\{\mathbf{M}_k\}, \mathbf{Z}) &\propto |\mathbf{M}_p|^{\mu-m+\sum_{k=1}^K \nu_k} \\ &\quad \times \text{etr} \left\{ - \left[\mu \bar{\mathbf{M}}_p^{-1} + \sum_{k=1}^K (\nu_k - m) \mathbf{M}_k^{-1} \right] \mathbf{M}_p \right\} \end{aligned} \quad (8)$$

$$\begin{aligned} f(\mathbf{M}_k|\mathbf{M}_p, \{\mathbf{M}_\ell\}_{\ell \neq k}, \mathbf{Z}) &\propto |\mathbf{M}_k|^{-(\nu_k+L_k+m)} \\ &\quad \times \text{etr} \left\{ - \left[(\nu_k - m) \mathbf{M}_p + \mathbf{Z}_k \mathbf{Z}_k^H \right] \mathbf{M}_k^{-1} \right\} \end{aligned} \quad (9)$$

and, hence, the conditional distributions can be expressed as

$$\begin{aligned} \mathbf{M}_p|\{\mathbf{M}_k\}, \mathbf{Z} &\sim \mathcal{CW}_m \left(\left[\mu \bar{\mathbf{M}}_p^{-1} + \sum_{k=1}^K (\nu_k - m) \mathbf{M}_k^{-1} \right]^{-1}, \right. \\ &\quad \left. \mu + \sum_{k=1}^K \nu_k \right) \end{aligned} \quad (10)$$

$$\begin{aligned} \mathbf{M}_k|\mathbf{M}_p, \{\mathbf{M}_\ell\}_{\ell \neq k}, \mathbf{Z} &\sim \mathcal{CW}_m^{-1} \left((\nu_k - m) \mathbf{M}_p + \mathbf{Z}_k \mathbf{Z}_k^H, \nu_k + L_k \right). \end{aligned} \quad (11)$$

The Gibbs sampler will, thus, iterate between (10) and (11). Observe that since \mathbf{M}_p [respectively, \mathbf{M}_k] is generated from a Wishart [respectively, inverse Wishart] distribution, these matrices are guaranteed to be Hermitian positive definite along the iterations. As usual, the first N_{bi} generated matrices (belonging to the so-called burn-in period) are not used for the estimation of \mathbf{M}_p , and only the N_r last matrices are used, yielding the following MMSE estimate:

$$\hat{\mathbf{M}}_p^{\text{MMSE}} = \frac{1}{N_r} \sum_{i=N_{bi}+1}^{N_{bi}+N_r} \mathbf{M}_p^{(i)}. \quad (12)$$

The proposed sampling strategy can be viewed as K Gibbs steps similar to those derived in [15]. This strategy naturally reduces to the algorithm proposed in [15] for $K = 1$.

B. Optimal SCM-Type Estimation

This section focuses on a class of estimators that are simpler to implement than the previous MMSE estimator, for the sake of

computational savings. More precisely, we consider estimators of \mathbf{M}_p defined by

$$\tilde{\mathbf{S}} = \sum_{k=1}^K w_k \mathbf{Z}_k \mathbf{Z}_k^H = \sum_{k=1}^K w_k \mathbf{S}_k \quad (13)$$

where the w_k 's are real-valued parameters. This class of estimators thus consists of a weighted sum of the sample covariance matrices of each group. Note that the classical sample covariance matrix estimator

$$\left(\sum_{k=1}^K L_k \right)^{-1} \sum_{k=1}^K \mathbf{Z}_k \mathbf{Z}_k^H \quad (14)$$

corresponds to uniform weighting, i.e., $w_k = (\sum_{k=1}^K L_k)^{-1}$. Under the stated assumptions, we derive the MSE of $\tilde{\mathbf{S}}$, i.e., the squared Frobenius norm of $\tilde{\mathbf{S}} - \mathbf{M}_p$, and show that there exists an optimal value of $\mathbf{w} = [w_1 \cdots w_K]^T$ which results in MMSE (herein \cdot^T denoting the transposition operator). More precisely, in the Appendix we prove the following result.

Proposition 1: The conditional and unconditional MSE of $\tilde{\mathbf{S}}$ are given by

$$\begin{aligned} & \mathcal{E} \left\{ \|\tilde{\mathbf{S}} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} \\ &= \left[\left(\sum_{k=1}^K w_k L_k \right) - 1 \right]^2 \text{Tr} \left\{ \mathbf{M}_p^2 \right\} \\ &+ \sum_{k=1}^K w_k^2 L_k^2 \frac{\text{Tr} \left\{ \mathbf{M}_p^2 \right\} + (\nu_k - m) \text{Tr} \left\{ \mathbf{M}_p \right\}^2}{(\nu_k - m + 1)(\nu_k - m - 1)} \\ &+ \sum_{k=1}^K w_k^2 L_k \frac{(\nu_k - m) \text{Tr} \left\{ \mathbf{M}_p^2 \right\} + (\nu_k - m)^2 \text{Tr} \left\{ \mathbf{M}_p \right\}^2}{(\nu_k - m + 1)(\nu_k - m - 1)} \end{aligned} \quad (15)$$

and

$$\begin{aligned} & \mathcal{E} \left\{ \|\tilde{\mathbf{S}} - \mathbf{M}_p\|^2 \right\} \\ &= \left[\left(\sum_{k=1}^K w_k L_k \right) - 1 \right]^2 \left[\text{Tr} \left\{ \bar{\mathbf{M}}_p^2 \right\} + \mu^{-1} \text{Tr} \left\{ \bar{\mathbf{M}}_p \right\}^2 \right] \\ &+ \sum_{k=1}^K w_k^2 \frac{(\nu_k + L_k - m) L_k}{(\nu_k - m + 1)(\nu_k - m - 1)} \\ &\times \left[(\nu_k - m + \mu^{-1}) \text{Tr} \left\{ \bar{\mathbf{M}}_p \right\}^2 \right. \\ &\quad \left. + [\mu^{-1}(\nu_k - m) + 1] \text{Tr} \left\{ \bar{\mathbf{M}}_p^2 \right\} \right] \\ &= c [\mathbf{w}^T \mathbf{L} - 1]^2 + \mathbf{w}^T \mathbf{\Gamma} \mathbf{w} \end{aligned} \quad (16)$$

where $\mathbf{L} = [L_1 \cdots L_K]^T$, and the definitions of the constant c and the diagonal matrix $\mathbf{\Gamma}$ follow immediately from (16).

Observe that c is a scalar and $\mathbf{\Gamma}$ is a diagonal matrix which depend on $\bar{\mathbf{M}}_p$, μ and ν_k , L_k , for $k = 1, \dots, K$. Since these quantities are known, it is possible to find the value of \mathbf{w} that yields the MMSE. Two routes can be taken to achieve this goal:

- the first solution consists of minimizing (16) with respect to \mathbf{w} , without any constraint on \mathbf{w} . It is straightforward to show that the minimizing argument of the MSE in (16) is

$$\mathbf{w}_o = c [\mathbf{\Gamma} + c \mathbf{L} \mathbf{L}^T]^{-1} \mathbf{L} = \frac{c \mathbf{\Gamma}^{-1} \mathbf{L}}{1 + c \mathbf{L}^T \mathbf{\Gamma}^{-1} \mathbf{L}}. \quad (17)$$

- as an alternative, the MSE can be minimized with the constraint that $\mathbf{w}^T \mathbf{L} = 1$. This amounts to enforce that $\tilde{\mathbf{S}}$ is an unbiased estimate of \mathbf{M}_p , whatever \mathbf{M}_p , see (35). In this case, $\mathbf{w}^T \mathbf{\Gamma} \mathbf{w}$ should be minimized under the constraint that $\mathbf{w}^T \mathbf{L} = 1$. The solution to this constrained optimization problem is well-known and is given by

$$\mathbf{w}_u = \frac{\mathbf{\Gamma}^{-1} \mathbf{L}}{\mathbf{L}^T \mathbf{\Gamma}^{-1} \mathbf{L}}. \quad (18)$$

The weight vectors (17) and (18) can, thus, be calculated and used in (13) to yield the weighted SCM estimators. It should be observed that the optimal weights in (17), (18) depend on $\bar{\mathbf{M}}_p$ only through $\text{Tr} \left\{ \bar{\mathbf{M}}_p^2 \right\}$ and $\text{Tr} \left\{ \bar{\mathbf{M}}_p \right\}^2$. Note also that the weighted SCM estimators will differ from the SCM estimator provided that the parameters ν_k are different; indeed, if all ν_k and L_k are equal, the three estimators are identical. In contrast, if there are differences among the ν_k , the weights are likely to be different, giving more importance to the groups of snapshots that are most homogeneous with the CUT.

Remark 1: The conditional MSE of $\tilde{\mathbf{S}}$ given \mathbf{M}_p , is of the form $c' [\mathbf{w}^T \mathbf{L} - 1]^2 + \mathbf{w}^T \mathbf{\Gamma}' \mathbf{w}$, where c' and $\mathbf{\Gamma}'$ depend on \mathbf{M}_p , μ and ν_k , L_k , for $k = 1, \dots, K$, see (15). Therefore, it is possible to obtain the vector \mathbf{w} which minimizes $\mathcal{E} \left\{ \|\tilde{\mathbf{S}} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\}$. However, this vector depends on \mathbf{M}_p , which is unknown. Thus, the practical implementation of such an estimator is not feasible.

C. Estimation Based on Loading of the Weighted SCM

In adaptive beamforming and detection, diagonal loading [21], [22] is recognized as an efficient means to compensate for various types of errors in covariance matrix estimation. Diagonal loading consists of adding a scaled identity matrix to the SCM, and can be viewed as a regularization of the SCM, resulting in enhanced performance with limited number of samples. Accordingly, colored loading, which consists of adding a colored covariance matrix to the SCM, has recently been proposed in the context of KA-STAP [10], [12]. In these references, it was shown that adding a matrix proportional to $\bar{\mathbf{M}}_p$ to the SCM enables one to significantly improve the performance of adaptive detectors. Following this route, we now consider estimates of \mathbf{M}_p of the form

$$\hat{\mathbf{M}}_p = w_0 \bar{\mathbf{M}}_p + \sum_{k=1}^K w_k \mathbf{Z}_k \mathbf{Z}_k^H = w_0 \bar{\mathbf{M}}_p + \tilde{\mathbf{S}}. \quad (19)$$

Our goal is to obtain the values of w_0 and \mathbf{w} that result in MMSE. Using (15) along with (35), we can write

$$\begin{aligned} \mathcal{E} \left\{ \|\hat{\mathbf{M}}_p - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} &= \mathcal{E} \left\{ \|w_0 \bar{\mathbf{M}}_p + \tilde{\mathbf{S}} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} \\ &= \mathcal{E} \left\{ \|\tilde{\mathbf{S}} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} + w_0^2 \|\bar{\mathbf{M}}_p\|^2 \\ &\quad + 2w_0 \text{Tr} \left\{ \mathcal{E} \left\{ (\tilde{\mathbf{S}} - \mathbf{M}_p) \bar{\mathbf{M}}_p | \mathbf{M}_p \right\} \right\} \\ &= \mathcal{E} \left\{ \|\tilde{\mathbf{S}} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} \\ &\quad + w_0^2 \|\bar{\mathbf{M}}_p\|^2 + 2w_0 \text{Tr} \left\{ (\mathbf{w}^T \mathbf{L} - 1) \mathbf{M}_p \bar{\mathbf{M}}_p \right\}. \end{aligned} \quad (20)$$

Next, we average the previous equation with respect to the prior distribution of \mathbf{M}_p to obtain

$$\begin{aligned} \mathcal{E} \left\{ \|\hat{\mathbf{M}}_p - \mathbf{M}_p\|^2 \right\} &= c[\mathbf{w}^T \mathbf{L} - 1]^2 + \mathbf{w}^T \mathbf{\Gamma} \mathbf{w} \\ &\quad + w_0^2 \|\bar{\mathbf{M}}_p\|^2 + 2w_0 (\mathbf{w}^T \mathbf{L} - 1) \text{Tr} \left\{ \bar{\mathbf{M}}_p^2 \right\}. \end{aligned} \quad (21)$$

Since $\mathcal{E} \{ \hat{\mathbf{M}}_p | \mathbf{M}_p \} = w_0 \bar{\mathbf{M}}_p + (\mathbf{w}^T \mathbf{L}) \mathbf{M}_p$, it is no longer possible to have an unbiased estimate of \mathbf{M}_p , for all \mathbf{M}_p , and, hence, we minimize the MSE in (21) without any constraint on w_0 and \mathbf{w} . Differentiating (21) with respect to w_0 and \mathbf{w} , and equating the result to zero, we obtain the following set of equations:

$$\mathbf{\Gamma} \mathbf{w} + c(\mathbf{L} \mathbf{L}^T \mathbf{w} - \mathbf{L}) + w_0 \mathbf{L} \text{Tr} \left\{ \bar{\mathbf{M}}_p^2 \right\} = \mathbf{0} \quad (22a)$$

$$w_0 + \mathbf{w}^T \mathbf{L} = 1. \quad (22b)$$

It follows that the optimal values of w_0 and \mathbf{w} are given by

$$\begin{pmatrix} \mathbf{w} \\ w_0 \end{pmatrix} = \begin{pmatrix} \mathbf{\Gamma} + c\mathbf{L}\mathbf{L}^T & \text{Tr} \left\{ \bar{\mathbf{M}}_p^2 \right\} \mathbf{L} \\ \mathbf{L}^T & 1 \end{pmatrix}^{-1} \begin{pmatrix} c\mathbf{L} \\ 1 \end{pmatrix}. \quad (23)$$

Equation (23) provides the optimal loading level w_0 , as well as the optimal weights to be applied to each matrix $\mathbf{Z}_k \mathbf{Z}_k^H$ in order to achieve a MMSE. Similarly to the SCM-type estimators, we observe that these weights depend on $\bar{\mathbf{M}}_p$ only through $\text{Tr} \{ \bar{\mathbf{M}}_p^2 \}$ and $\text{Tr} \{ \bar{\mathbf{M}}_p \}$.

A similar procedure can be applied to obtain the best estimator based on diagonal loading of the weighted SCM. Toward this end, let us now consider the following estimator:

$$\hat{\mathbf{M}}_p = w_0 \mathbf{I} + \sum_{k=1}^K w_k \mathbf{Z}_k \mathbf{Z}_k^H = w_0 \mathbf{I} + \tilde{\mathbf{S}}. \quad (24)$$

The conditional MSE of such an estimate is given by

$$\begin{aligned} \mathcal{E} \left\{ \|\hat{\mathbf{M}}_p - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} &= \mathcal{E} \left\{ \|w_0 \mathbf{I} + \tilde{\mathbf{S}} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} \\ &= \mathcal{E} \left\{ \|\tilde{\mathbf{S}} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} + mw_0^2 \\ &\quad + 2w_0 \text{Tr} \left\{ \mathcal{E} \left\{ (\tilde{\mathbf{S}} - \mathbf{M}_p) | \mathbf{M}_p \right\} \right\} \\ &= \mathcal{E} \left\{ \|\tilde{\mathbf{S}} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} + mw_0^2 \\ &\quad + 2w_0 (\mathbf{w}^T \mathbf{L} - 1) \text{Tr} \left\{ \mathbf{M}_p \right\}. \end{aligned} \quad (25)$$

Taking the expectation with respect to $f(\mathbf{M}_p)$, we end up with the following expression for the unconditional MSE:

$$\begin{aligned} \mathcal{E} \left\{ \|\hat{\mathbf{M}}_p - \mathbf{M}_p\|^2 \right\} &= c[\mathbf{w}^T \mathbf{L} - 1]^2 + \mathbf{w}^T \mathbf{\Gamma} \mathbf{w} + mw_0^2 \\ &\quad + 2w_0 (\mathbf{w}^T \mathbf{L} - 1) \text{Tr} \left\{ \bar{\mathbf{M}}_p \right\}. \end{aligned} \quad (26)$$

In order to minimize this MSE, we equate to zero the derivative of (26) with respect to w_0 and \mathbf{w} , yielding

$$\mathbf{\Gamma} \mathbf{w} + c(\mathbf{L} \mathbf{L}^T \mathbf{w} - \mathbf{L}) + w_0 \mathbf{L} \text{Tr} \left\{ \bar{\mathbf{M}}_p \right\} = \mathbf{0} \quad (27a)$$

$$mw_0 + \mathbf{w}^T \mathbf{L} \text{Tr} \left\{ \bar{\mathbf{M}}_p \right\} = \text{Tr} \left\{ \bar{\mathbf{M}}_p \right\}. \quad (27b)$$

The optimal values of w_0 and \mathbf{w} are, hence

$$\begin{pmatrix} \mathbf{w} \\ w_0 \end{pmatrix} = \begin{pmatrix} \mathbf{\Gamma} + c\mathbf{L}\mathbf{L}^T & \text{Tr} \left\{ \bar{\mathbf{M}}_p \right\} \mathbf{L} \\ \text{Tr} \left\{ \bar{\mathbf{M}}_p \right\} \mathbf{L}^T & m \end{pmatrix}^{-1} \begin{pmatrix} c\mathbf{L} \\ \text{Tr} \left\{ \bar{\mathbf{M}}_p \right\} \end{pmatrix} \quad (28)$$

which provides the optimal (diagonal) loading level and SCM weights.

Remark 2: Conventional diagonal loading would amount to consider estimates of the form

$$\hat{\mathbf{M}}_p = w_0 \mathbf{I} + L^{-1} \sum_{k=1}^K \mathbf{Z}_k \mathbf{Z}_k^H = w_0 \mathbf{I} + L^{-1} \mathbf{S}$$

where $L = \sum_{k=1}^K L_k$. Doing so, the conditional MSE becomes

$$\begin{aligned} \mathcal{E} \left\{ \|\hat{\mathbf{M}}_p - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} &= \mathcal{E} \left\{ \|w_0 \mathbf{I} + L^{-1} \mathbf{S} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} \\ &= \mathcal{E} \left\{ \|L^{-1} \mathbf{S} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} + mw_0^2 \\ &\quad + 2w_0 \text{Tr} \left\{ \mathcal{E} \left\{ (L^{-1} \mathbf{S} - \mathbf{M}_p) | \mathbf{M}_p \right\} \right\} \\ &= \mathcal{E} \left\{ \|L^{-1} \mathbf{S} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} + mw_0^2 \\ &\geq \mathcal{E} \left\{ \|L^{-1} \mathbf{S} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} \end{aligned}$$

as $\mathcal{E} \{ (L^{-1} \mathbf{S} - \mathbf{M}_p) | \mathbf{M}_p \} = \mathbf{0}$, see (35). In other words, as far as covariance matrix estimation is concerned, the MSE cannot be decreased compared to that of the conventional SCM estimator. This is why it is necessary to consider a weighted SCM in the case of diagonal loading.

Remark 3: In what precedes, the MSE, i.e., the square of the Frobenius norm of the estimation error, was used to assess the performance of the various estimators. Although this is the common choice in estimation theory, it should be observed that one is concerned here with the estimation of a covariance matrix, which is Hermitian positive definite. Consequently, this matrix belongs to the quotient space $P_m = \mathbf{GL}(m, \mathbb{C}) / \mathbf{U}(m)$, where $\mathbf{GL}(m, \mathbb{C})$ denotes the general linear group and $\mathbf{U}(m)$ is the Lie group of unitary matrices [23]. As advocated in [23], the natural metric on P_m , which involves the generalized eigenvalues of the covariance matrix and its estimate, could be used instead of the Frobenius norm. Our experience is that it does not fundamentally modify the hierarchy between the estimators compared to the MSE criterion, except for the diagonal or colored loading estimators. The latter perform rather well with regards to the MSE criterion, see the next section. In contrast, with the metric of [23], they do not provide a significant improvement compared to the SCM-type estimators.

IV. MAXIMUM-LIKELIHOOD ESTIMATION WITH A DETERMINISTIC \mathbf{M}_p

The fully Bayesian approaches described previously rely on the assumption that \mathbf{M}_p has a Wishart distribution with known mean $\bar{\mathbf{M}}_p$. In order to relax such hypothesis, we now consider that \mathbf{M}_p is deterministic and thus we do not make any specific assumption about it. Since \mathbf{M}_p is assumed to be deterministic and unknown, we turn to the derivation of the maximum-likelihood estimate of \mathbf{M}_p . Towards this end, we first need to derive an expression for $f(\mathbf{Z}|\mathbf{M}_p)$. Under the stated assumptions, we have

$$\begin{aligned}
f(\mathbf{Z}|\mathbf{M}_p) &= \int \cdots \int f(\mathbf{Z}|\{\mathbf{M}_k\}) f(\{\mathbf{M}_k\}|\mathbf{M}_p) d\mathbf{M}_1 \cdots d\mathbf{M}_K \\
&= \int \cdots \int \prod_{k=1}^K f(\mathbf{Z}_k|\mathbf{M}_k) f(\mathbf{M}_k|\mathbf{M}_p) d\mathbf{M}_1 \cdots d\mathbf{M}_K \\
&\propto \prod_{k=1}^K |\mathbf{M}_p|^{\nu_k} \int |\mathbf{M}_k|^{-(\nu_k+L_k+m)} \\
&\quad \times \text{etr} \left\{ -\mathbf{M}_k^{-1} \left[(\nu_k-m)\mathbf{M}_p + \mathbf{Z}_k \mathbf{Z}_k^H \right] \right\} d\mathbf{M}_k \\
&\propto \prod_{k=1}^K |\mathbf{M}_p|^{\nu_k} \left| (\nu_k-m)\mathbf{M}_p + \mathbf{Z}_k \mathbf{Z}_k^H \right|^{-(\nu_k+L_k)}. \quad (29)
\end{aligned}$$

The log-likelihood function is, thus

$$\begin{aligned}
\Lambda(\mathbf{Z}|\mathbf{M}_p) &= \text{const.} + \left(\sum_{k=1}^K \nu_k \right) \ln |\mathbf{M}_p| \\
&\quad - \sum_{k=1}^K (\nu_k + L_k) \ln \left| (\nu_k-m)\mathbf{M}_p + \mathbf{Z}_k \mathbf{Z}_k^H \right|. \quad (30)
\end{aligned}$$

Differentiating $\Lambda(\mathbf{Z}|\mathbf{M}_p)$ with respect to \mathbf{M}_p gives the following result:

$$\begin{aligned}
\frac{\partial \Lambda(\mathbf{Z}|\mathbf{M}_p)}{\partial \mathbf{M}_p} &= \left(\sum_{k=1}^K \nu_k \right) \mathbf{M}_p^{-1} \\
&\quad - \sum_{k=1}^K (\nu_k + L_k) \left[\mathbf{M}_p + (\nu_k-m)^{-1} \mathbf{Z}_k \mathbf{Z}_k^H \right]^{-1} \\
&= - \left(\sum_{k=1}^K L_k \right) \mathbf{M}_p^{-1} + \mathbf{M}_p^{-1} \\
&\quad \times \left[\sum_{k=1}^K (\nu_k + L_k) (\nu_k-m)^{-1} \mathbf{Z}_k \right. \\
&\quad \left. \times \left[\mathbf{I} + (\nu_k-m)^{-1} \mathbf{Z}_k^H \mathbf{M}_p^{-1} \mathbf{Z}_k \right]^{-1} \mathbf{Z}_k^H \right] \mathbf{M}_p^{-1}. \quad (31)
\end{aligned}$$

Therefore, the MLE of \mathbf{M}_p satisfies the following equation

$$\begin{aligned}
\left(\sum_{k=1}^K L_k \right) \mathbf{M}_p &= \sum_{k=1}^K (\nu_k + L_k) (\nu_k-m)^{-1} \mathbf{Z}_k \\
&\quad \times \left[\mathbf{I} + (\nu_k-m)^{-1} \mathbf{Z}_k^H \mathbf{M}_p^{-1} \mathbf{Z}_k \right]^{-1} \mathbf{Z}_k^H. \quad (32)
\end{aligned}$$

This is an implicit equation of the type $\mathbf{M}_p = g(\mathbf{M}_p, \mathbf{Z})$. In order to solve it, we propose an iterative scheme. Starting with an initial value $\mathbf{M}_p^{(0)}$, the new value of \mathbf{M}_p is updated as $\mathbf{M}_p^{(i+1)} = g(\mathbf{M}_p^{(i)}, \mathbf{Z})$. Note that this kind of iterative solutions has already been encountered in different problems, see, e.g., [24]–[27]. For instance, [25] considers the Expectation-Maximization (EM) algorithm for covariance matrix estimation with compound-Gaussian clutter, and ends up with a similar implicit equation. As mentioned in [25], there is no guarantee that the EM algorithm converges, and the same observation applies to our MLE. Accordingly, care is to be taken when choosing an initial estimate as it may impact the final estimate. In fact, proving that such a scheme converges, and moreover to a unique point, is rather intricate. A proof of convergence to a unique solution for a similar iterative scheme was given in [26] and [27]. However, this proof cannot be easily adapted to the matrix equation (32). Hence, we can only conjecture that the same property may apply here. However, we did not encounter any convergence problem in our numerical experiments.

V. NUMERICAL SIMULATIONS

We now illustrate the performance of the various estimation schemes introduced in the previous sections. In all simulations, we consider an array with $m = 8$ elements and the covariance matrix $\bar{\mathbf{M}}_p$ is of the form $\bar{\mathbf{M}}_p(k, \ell) = 0.9^{|k-\ell|}$. The total number of training samples is set to $\sum_{k=1}^K L_k = 3m = 24$ but we vary the number of groups K . More specifically, we consider three cases with $K = 2$, $K = 3$ and $K = 6$, respectively. This means that there exist 2, 3 or 6 different covariance matrices among the training samples with different degrees of homogeneity specified by the values of the parameters ν_k , for $k = 1, \dots, K$. The three cases are defined by:

- Case 1: $\mathbf{L} = [16 \ 8]^T$ and $\boldsymbol{\nu} = [\nu_1 \ 10]^T$;
- Case 2: $\mathbf{L} = [8 \ 8 \ 8]^T$ and $\boldsymbol{\nu} = [\nu_1 \ 10 \ 24]^T$;
- Case 3: $\mathbf{L} = [4 \ 4 \ 4 \ 4 \ 4 \ 4]^T$ and $\boldsymbol{\nu} = [\nu_1 \ 10 \ 12 \ 16 \ 20 \ 24]^T$.

The value of ν_1 is varied between 10 and 24. In the first case, the second group with eight samples is rather heterogeneous compared to the CUT (i.e., \mathbf{M}_2 may be far from \mathbf{M}_p), and the degree of heterogeneity of the first group is varied, resulting in training samples being more homogeneous with the CUT as ν_1 increases. In the second scenario, the second group of samples is rather heterogeneous compared to the CUT while, for the third group, \mathbf{M}_3 may be quite close to \mathbf{M}_p . The degree of heterogeneity of the first group varies as previously. In the third scenario, the K groups have K different degrees of heterogeneity. We consider two different values for μ , i.e., $\mu = 10$ or $\mu = 20$ adjusting the importance of the *a priori* information used in the model. In the former case, the prior knowledge about $\bar{\mathbf{M}}_p$ is not very informative while, in the latter case, \mathbf{M}_p will be closer to $\bar{\mathbf{M}}_p$.

A. Convergence of the Gibbs Sampler

We first study the convergence of the Gibbs sampler and show how to determine the values of N_{bi} and N_r that are sufficient to ensure convergence of the Gibbs sampler. Usually, a two-step procedure is used [28]. First, a rough estimate of N_{bi} and N_r is obtained by observing the MSE along the iterations of the Gibbs sampler. More precisely, one begins with large values of

N_{bi} and N_r , so that a reference estimate is obtained. In order to determine N_r , the MSE between this reference and the estimate obtained from N_r iterations is computed. The number of iterations N_r is selected as the value above which this MSE is sufficiently small. In our case, we observed that $N_r = 200$ iterations were sufficient to obtain a small MSE. Keeping this value of N_r , we next computed the MSE for different values of the number of burn-in iterations N_{bi} . We observed that a very short burn-in period was sufficient, and that the MSE was approximately constant as soon as $N_{bi} \geq 20$. Therefore this value was retained. Once N_{bi} and N_r are chosen, there exists a rigorous procedure to assess convergence [28]. It consists of running M chains in parallel, each one of length (N_{bi}, N_r) but with different initial values, and to observe the variability of the estimates within and between the chains. More precisely, let $\mathbf{M}_p^{(i,j)}$ be the matrix obtained at the i th iteration of the j th chain and let us note

$$\hat{\mathbf{M}}_p^{(\cdot,j)} = N_r^{-1} \sum_{i=N_{bi}+1}^{N_{bi}+N_r} \mathbf{M}_p^{(i,j)}$$

$$\check{\mathbf{M}}_p = M^{-1} \sum_{j=1}^M \hat{\mathbf{M}}_p^{(\cdot,j)}$$

where $\hat{\mathbf{M}}_p^{(\cdot,j)}$ corresponds to the MMSE for the j th chain, and $\check{\mathbf{M}}_p$ is the average value over the M chains. Let \mathbf{B} and \mathbf{W} be the between-sequence and within-sequence covariance matrices for the M Markov chains, whose (p, q) element are defined as

$$[\mathbf{B}]_{pq} = \frac{N_r}{M-1} \sum_{j=1}^M \left(\left[\hat{\mathbf{M}}_p^{(\cdot,j)} \right]_{pq} - [\check{\mathbf{M}}_p]_{pq} \right)^2$$

$$[\mathbf{W}]_{pq} = \frac{1}{M} \sum_{j=1}^M \frac{1}{N_r-1} \times \sum_{i=N_{bi}+1}^{N_{bi}+N_r} \left(\left[\mathbf{M}_p^{(i,j)} \right]_{pq} - \left[\hat{\mathbf{M}}_p^{(\cdot,j)} \right]_{pq} \right)^2.$$

The convergence of the Gibbs sampler can be monitored by the multivariate potential scale reduction factor, which is defined as [29]

$$\rho = \frac{N_r - 1}{N_r} + \frac{M + 1}{MN_r} \lambda_{\max}(\mathbf{W}^{-1}\mathbf{B}) \quad (33)$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the matrix between parentheses. Fig. 1 below shows the value of ρ in case 1 for 100 independent realizations. As can be observed, the average value of ρ is around 1.12, which is deemed sufficient to conclude that the Gibbs sampler effectively converges with $N_{bi} = 20$ and $N_r = 200$ (a value of ρ below 1.2 is recommended for convergence in [28]). Therefore, in the sequel we will use these values.

B. Performance Analysis

We now study the performance of the various estimators derived in Sections III and IV, namely:

- the new heterogeneous Gibbs sampler operating on the matrices $\{\mathbf{M}_k\}$, see (10), (11), and the Gibbs sampler of [15]

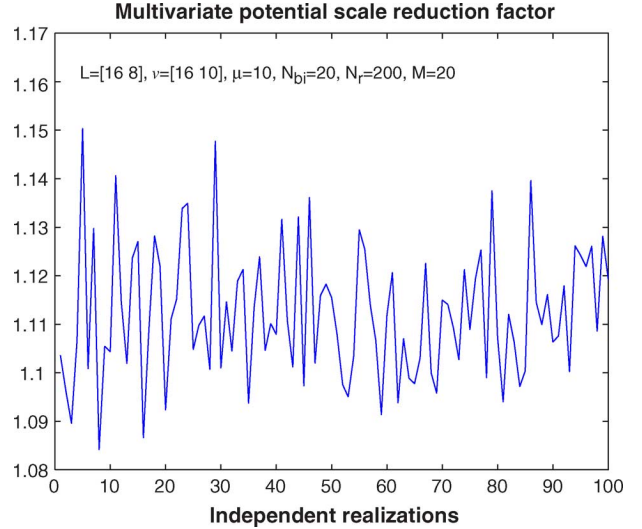


Fig. 1. Multivariate potential scale reduction factor in case 1.

which assumes that all training samples share the same covariance matrix \mathbf{M}_s ;

- the SCM-type estimators, more precisely the conventional SCM of (14), and the weighted SCM estimators of (17) and (18);
- the estimates based on diagonal loading, cf. (24) and (28), or colored loading, see (19) and (23);
- the MLE obtained as the solution of (32).

All simulations were conducted with 5000 Monte-Carlo runs and the MSE of each estimator was estimated from these 5000 trials. The MLE is initialized with the SCM, and the number of iterations in the MLE procedure is set to 50. The MSE of all estimates is displayed in the following way. The MMSE estimator appears in solid lines while the method of [15] appears with a circle. The SCM, weighted SCM with \mathbf{w}_u and weighted SCM with \mathbf{w}_o appear with a square, a star, and a cross marker, respectively. Diagonal [respectively, colored] loading is displayed by a triangle (up) [respectively, triangle (down)]. Figs. 2 – 7 compare the performance of all estimators. From inspection of these figures, the following observations can be made.

- The proposed heterogeneous Gibbs sampler always provides the smallest MSE. The improvement compared to the Gibbs sampler of [15] which assumes that all training samples have the same covariance matrix is about 1 dB. This shows that it is beneficial to take into account the fact that the covariance matrices of the \mathbf{Z}_k may be different. However, in the present case, the improvement is not very important, mainly because all covariance matrices \mathbf{M}_k have the same average value, namely \mathbf{M}_p . Hence, assuming that all snapshots \mathbf{Z}_k share the same covariance matrix \mathbf{M}_s , with average value \mathbf{M}_p , does not result in a significant degradation. In contrast, the performance is expected to be more degraded as soon as the average value of each \mathbf{M}_k is different.
- The MSE of the heterogeneous Gibbs sampler is between 4 and 8 dB lower than that of the weighted SCM estimators, which is significant. However, the latter perform quite well

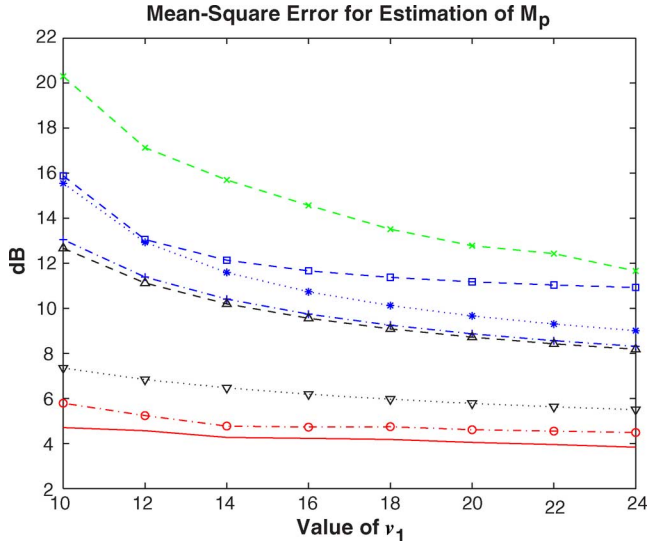


Fig. 2. Performance of the various estimation schemes versus ν_1 in case 1. $\mu = 10$.

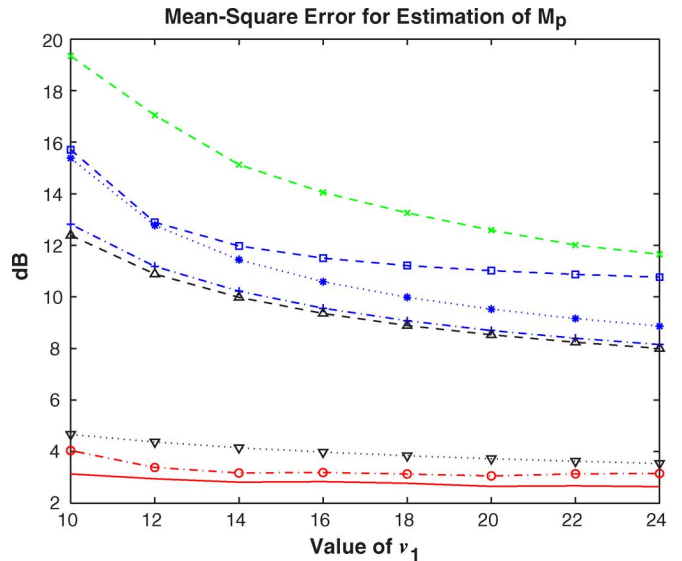


Fig. 3. Performance of the various estimation schemes versus ν_1 in case 1. $\mu = 20$.

given that they are very simple schemes. Using \mathbf{w}_o instead of \mathbf{w}_u results in a 1-dB gain; hence, enforcing unbiasedness of $\hat{\mathbf{S}}$ for any \mathbf{M}_p may not be the best strategy, at least when the MSE is concerned.

- More importantly, loading of the weighted SCM appears to be a very effective solution, notably colored loading. Diagonal loading is slightly superior to the best weighted SCM, and colored loading offers a considerable improvement, between 6 and 8 dB compared to diagonal loading. In some cases, especially when the number of groups increases, colored loading happens to be better than the method of [15] which assumes that all training samples share the same covariance. The MSE of the colored loading estimator is also only 1 to 3 dB above that of the MMSE. This makes colored loading an interesting scheme as it offers an excellent tradeoff between performance and computational complexity.
- The performance of the MMSE, the weighted SCM and colored loading estimators is approximately constant when ν_1 varies. An explanation is that only one over K values of ν_k is varied, and, hence, the overall set of data does not undergo important modifications. Also, all algorithms know the value of $\boldsymbol{\nu}$, and, hence, take it into account to achieve a nearly constant MSE.
- Using *a priori* information results in significant gain. This is for instance illustrated by the fact that the weighted SCM estimators perform better than the MLE (although being simpler): this is due to the fact that they use $\hat{\mathbf{M}}_p$ in contrast to the MLE. Therefore, *a priori* knowledge turns out to be useful as it enables one to have more accurate estimates while keeping the computational complexity lower.
- As μ goes from $\mu = 10$ to $\mu = 20$ (i.e., as \mathbf{M}_p is closer to $\hat{\mathbf{M}}_p$) the MSE of the MMSE and colored loading estimators decreases by about 1.3 dB. This corresponds to the improvement provided by a more and more reliable *a priori* information, i.e., \mathbf{M}_p being closer to $\hat{\mathbf{M}}_p$.

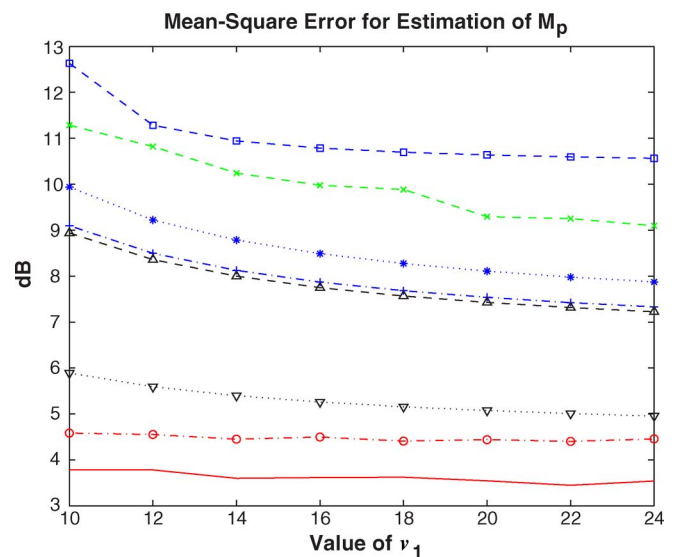


Fig. 4. Performance of the various estimation schemes versus ν_1 in case 2. $\mu = 10$.

To summarize this section, the MMSE was observed to be the most performant method, at the price of high computational complexity due to the implementation of a Gibbs sampler. Two other methods can be seen as competitors: the weighted SCM with optimal weight \mathbf{w}_o and colored loading. Both of them are simple from a computational point of view. The latter is more performant but it assumes knowledge of $\hat{\mathbf{M}}_p$ while the former depends on $\hat{\mathbf{M}}_p$ only through $\text{Tr}\{\hat{\mathbf{M}}_p^2\}$ and $\text{Tr}\{\hat{\mathbf{M}}_p\}^2$. Therefore, it may be more robust to imprecise knowledge of $\hat{\mathbf{M}}_p$, an issue that is addressed in Section V-C.

C. Robustness Analysis

In this section, we study the robustness of the estimators to imprecise knowledge of the parameters describing the *a priori*

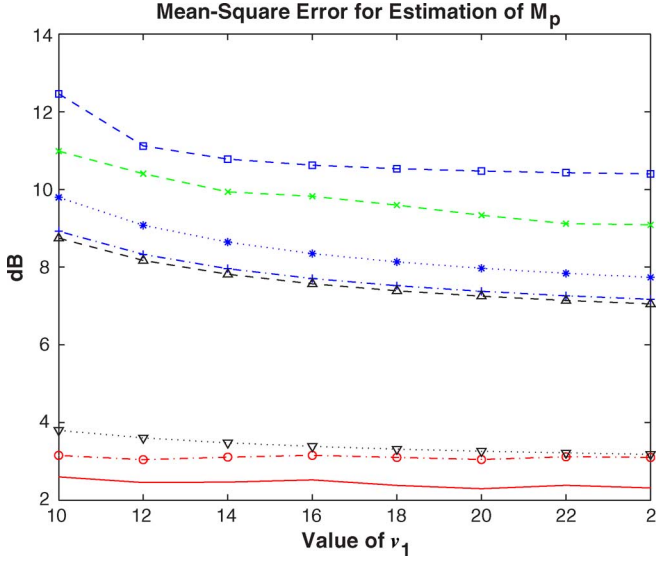


Fig. 5. Performance of the various estimation schemes versus ν_1 in case 2. $\mu = 20$.

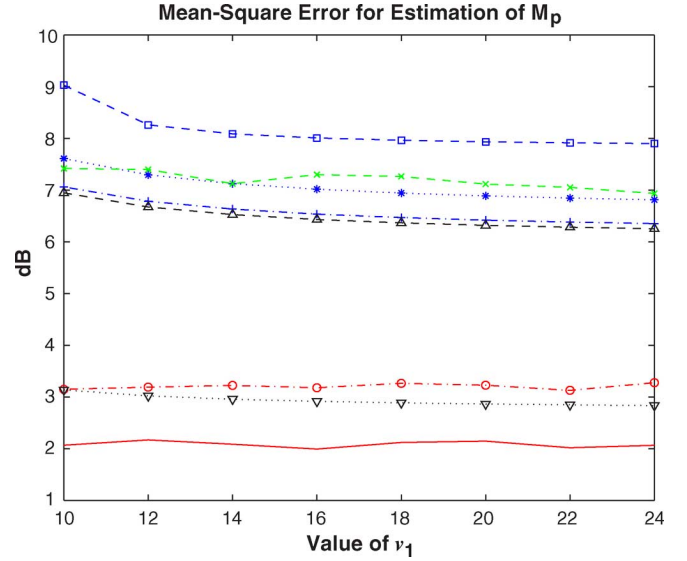


Fig. 7. Performance of the various estimation schemes versus ν_1 in case 3. $\mu = 20$.

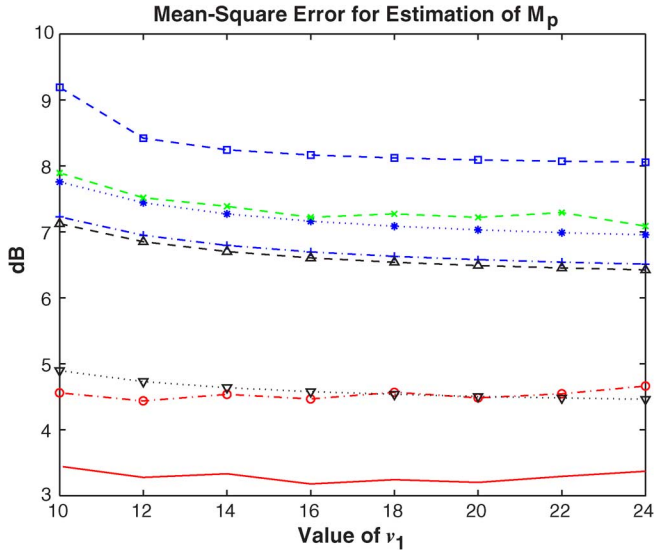


Fig. 6. Performance of the various estimation schemes versus ν_1 in case 3. $\mu = 10$.

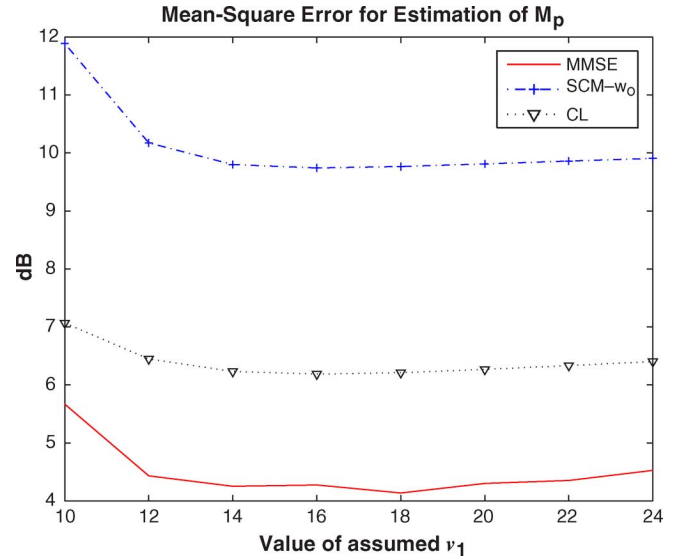


Fig. 8. Robustness of the estimation schemes to a mismatch in ν_1 in case 1. $\mu = 10$.

information. Since these parameters are numerous, we focus on case 1, and we consider a possible mismatch in ν_1 or $\bar{\mathbf{M}}_p$. In order to test the robustness towards a mismatch of ν_1 , the data was generated with $\nu_1 = 16$ and the algorithms were run with an assumed value of ν_1 , which varies between 10 and 24. In other words, the algorithms do not use the correct value of ν_1 which rules the degree of heterogeneity of the first group of snapshots. For the sake of readability and because they offer the best performance, we only consider here the MMSE estimator, the optimal weighted SCM and colored loading. Fig. 8 displays the MSE of these three methods versus the assumed value of ν_1 . As can be observed, the estimation schemes are rather robust to an imprecise knowledge of ν_1 : their MSE does not significantly vary, which is an appealing feature.

In Fig. 9 we test the robustness of the algorithms to an imprecise knowledge of $\bar{\mathbf{M}}_p$. Towards this end, the data is gener-

ated with the same $\bar{\mathbf{M}}_p$ as in previous simulations. In order to simulate an error in $\bar{\mathbf{M}}_p$, we generate a random matrix drawn from a Wishart distribution with mean the true $\bar{\mathbf{M}}_p$ and q degrees of freedom. In other words, the algorithms are used with a wrong value of $\bar{\mathbf{M}}_p$, randomly distributed around the true $\bar{\mathbf{M}}_p$. For the sake of notational convenience let us denote by $\bar{\mathbf{M}}'_p$ this matrix. As q increases, the distance between the two matrices decreases, and, hence, the mismatch level decreases. In order to give an order of magnitude, the MSE between $\bar{\mathbf{M}}_p$ and $\bar{\mathbf{M}}'_p$ is 8, 4.2, and 1 dB for $q = 10$, $q = 24$, and $q = 50$, respectively. Given that, for known $\bar{\mathbf{M}}_p$, the MSE of the MMSE estimator is about 4 dB, see Fig. 2, $q = 10$ is a very severe case, i.e., a strong mismatch. $q = 24$ corresponds to a situation where the mismatch in $\bar{\mathbf{M}}_p$ is comparable to the MSE for known $\bar{\mathbf{M}}_p$, and the mismatch is reasonably small for $q = 50$.

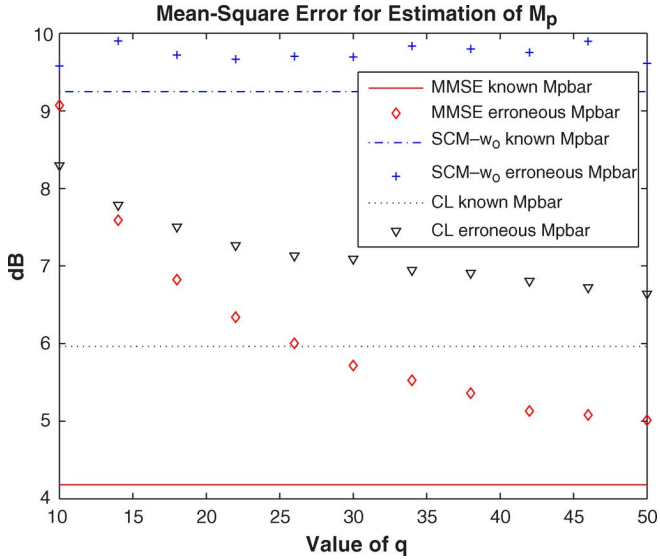


Fig. 9. Robustness of the estimation schemes to a mismatch in $\bar{\mathbf{M}}_p$ in case 1. $\mu = 10$.

The results are shown in Fig. 9. The weighted SCM is seen to be quite insensitive to variations in q : this is due to the fact that it depends on $\bar{\mathbf{M}}'_p$ only through $\text{Tr}\{\bar{\mathbf{M}}'^2_p\}$ and $\text{Tr}\{\bar{\mathbf{M}}'^2_p\}$. However, in the scenario considered, $\mathcal{E}\{\text{Tr}\{\bar{\mathbf{M}}'_p\}\} = \text{Tr}\{\bar{\mathbf{M}}_p\}$ and $\mathcal{E}\{\text{Tr}\{\bar{\mathbf{M}}'^2_p\}\} = \text{Tr}\{\bar{\mathbf{M}}^2_p\} + q^{-1}\text{Tr}\{\bar{\mathbf{M}}_p\}^2$. Hence, even if there is a mismatch between the actual $\bar{\mathbf{M}}_p$ and its presumed value, the impact is not necessarily serious, as one does not make a large error on $\text{Tr}\{\bar{\mathbf{M}}^2_p\}$ and $\text{Tr}\{\bar{\mathbf{M}}_p\}^2$. In contrast, colored loading appears to be more sensitive to $\bar{\mathbf{M}}_p$. Compared to its MSE for known $\bar{\mathbf{M}}_p$, it exhibits a loss of about 2.3 dB for $q = 10$ and 0.7 dB for $q = 50$. The MMSE estimator is the most sensitive estimator, hence, the least robust, as one can observe a loss of 4.9 dB [respectively, 0.83 dB] for $q = 10$ [respectively, $q = 50$] compared to its MSE for known $\bar{\mathbf{M}}_p$. For some values of q —but these values correspond to a very imprecise *a priori* knowledge—colored loading is even better than the MMSE estimator, indicating that the latter needs a more accurate knowledge of $\bar{\mathbf{M}}_p$ than the former.

VI. CONCLUSION

This paper extended the model presented in [15] to the case where the training samples do not share a common covariance matrix. Within this new framework, we presented a Gibbs sampling strategy to implement the MMSE estimator, and we derived optimal sample covariance matrix type estimators. Furthermore, in order to have the less *a priori* knowledge about \mathbf{M}_p , we also considered it as a deterministic quantity, and we derived its MLE. Numerical results indicate that the MMSE estimator implemented with the heterogeneous Gibbs sampler outperforms the estimators presented in [15], when there exist heterogeneities between the groups of snapshots. It was also shown that the simple weighted SCM estimators provide a

rather good performance. Moreover, they can be improved in a straightforward manner by considering either diagonal loading, or better colored loading. The latter scheme was shown to provide an excellent tradeoff between performance and computational complexity. Another important conclusion is the impact of *a priori* knowledge on the estimation performance. Bayesian methods using prior information provide significantly better performance than the MLE which does not use this prior information. Accordingly, the weighted SCM-type estimators, with possibly loading, which also use *a priori* information, although simpler than the MLE perform better. Finally, it was observed that the proposed estimation schemes were rather robust to a mismatch with respect to the knowledge of the degree of heterogeneity but less robust to a mismatch of $\bar{\mathbf{M}}_p$. Possible extensions of this work deals with modeling of more heterogeneous environments for which the covariance matrices \mathbf{M}_k may have a different average value.

APPENDIX PROOF OF (15) AND (16)

In this appendix, we derive the theoretical MSE of the weighted SCM $\tilde{\mathbf{S}}$ defined in (13). Toward this end, we proceed in two steps. First, we derive the MSE of $\tilde{\mathbf{S}}$ conditionally to \mathbf{M}_p ; this enables us to obtain the optimal value of \mathbf{w} for any \mathbf{M}_p . Of course, this optimal weight vector is hypothetical as \mathbf{M}_p is unknown. Next, we average the MSE for a given \mathbf{M}_p , over the density of \mathbf{M}_p , in order to obtain the final MSE. First note that

$$\mathcal{E}\left\{\|\tilde{\mathbf{S}} - \mathbf{M}_p\|^2 | \mathbf{M}_p\right\} = \text{Tr}\left\{\mathcal{E}\left\{\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H | \mathbf{M}_p\right\}\right. \\ \left. - 2\text{Tr}\left\{\mathcal{E}\left\{\tilde{\mathbf{S}} | \mathbf{M}_p\right\}\mathbf{M}_p\right\} + \text{Tr}\left\{\mathbf{M}_p^2\right\}\right\}. \quad (34)$$

However, since $\mathbf{Z}_k | \mathbf{M}_k \sim \mathcal{CN}_{m, L_k}(\mathbf{0}, \mathbf{M}_k, \mathbf{I}_{L_k})$ and $\mathcal{E}\{\mathbf{M}_k | \mathbf{M}_p\} = \mathbf{M}_p$, we have

$$\mathcal{E}\{\tilde{\mathbf{S}} | \mathbf{M}_p\} = \mathcal{E}_{\{\mathbf{M}_k\} | \mathbf{M}_p}\left\{\mathcal{E}\left\{\tilde{\mathbf{S}} | \{\mathbf{M}_k\}, \mathbf{M}_p\right\}\right\} \\ = \mathcal{E}_{\{\mathbf{M}_k\} | \mathbf{M}_p}\left\{\sum_{k=1}^K w_k \mathcal{E}\left\{\mathbf{S}_k | \mathbf{M}_k, \mathbf{M}_p\right\}\right\} \\ = \mathcal{E}_{\{\mathbf{M}_k\} | \mathbf{M}_p}\left\{\sum_{k=1}^K w_k L_k \mathbf{M}_k\right\} \\ = \left(\sum_{k=1}^K w_k L_k\right) \mathbf{M}_p. \quad (35)$$

Let us now turn to the average value of $\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H$, conditionally to \mathbf{M}_p . From the definition of $\tilde{\mathbf{S}}$, one can write that

$$\mathcal{E}\{\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H | \mathbf{M}_p\} \\ = \mathcal{E}_{\{\mathbf{M}_k\} | \mathbf{M}_p}\left\{\mathcal{E}\left\{\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H | \{\mathbf{M}_k\}, \mathbf{M}_p\right\}\right\} \\ = \mathcal{E}_{\{\mathbf{M}_k\} | \mathbf{M}_p}\left\{\mathcal{E}\left\{\sum_{k, \ell=1}^K w_k w_\ell \mathbf{S}_k \mathbf{S}_\ell^H | \{\mathbf{M}_k\}, \mathbf{M}_p\right\}\right\}. \quad (36)$$

Using the fact that $\mathbf{Z}_k|\mathbf{M}_k \sim \mathcal{CN}_{m,L_k}(\mathbf{0}, \mathbf{M}_k, \mathbf{I}_{L_k})$, we have

$$\begin{aligned}
& \mathcal{E} \left\{ \sum_{k,\ell=1}^K w_k w_\ell \mathbf{S}_k \mathbf{S}_\ell^H | \{\mathbf{M}_k\}, \mathbf{M}_p \right\} \\
&= \sum_{k=1}^K w_k^2 \mathcal{E} \left\{ \mathbf{S}_k^2 | \mathbf{M}_k, \mathbf{M}_p \right\} \\
&+ \sum_{k \neq \ell} w_k w_\ell \mathcal{E} \left\{ \mathbf{S}_k | \mathbf{M}_k, \mathbf{M}_p \right\} \mathcal{E} \left\{ \mathbf{S}_\ell | \mathbf{M}_\ell, \mathbf{M}_p \right\} \\
&= \sum_{k=1}^K w_k^2 \left[L_k^2 \mathbf{M}_k^2 + L_k \text{Tr}\{\mathbf{M}_k\} \mathbf{M}_k \right] \\
&+ \sum_{k \neq \ell} w_k w_\ell L_k L_\ell \mathbf{M}_k \mathbf{M}_\ell
\end{aligned} \tag{37}$$

and, hence

$$\begin{aligned}
\mathcal{E} \{ \tilde{\mathbf{S}} \tilde{\mathbf{S}}^H | \mathbf{M}_p \} &= \left(\sum_{k=1}^K w_k L_k \mathcal{E} \{ \mathbf{M}_k | \mathbf{M}_p \} \right)^2 \\
&+ \sum_{k=1}^K w_k^2 L_k^2 \left[\mathcal{E} \{ \mathbf{M}_k^2 | \mathbf{M}_p \} - \mathcal{E} \{ \mathbf{M}_k | \mathbf{M}_p \}^2 \right] \\
&+ \sum_{k=1}^K w_k^2 L_k \mathcal{E} \{ \text{Tr}\{\mathbf{M}_k\} \mathbf{M}_k | \mathbf{M}_p \}.
\end{aligned} \tag{38}$$

Since $\mathbf{M}_k | \mathbf{M}_p \sim \mathcal{CW}_m^{-1}((\nu_k - m) \mathbf{M}_p, \nu_k)$, it can be readily shown [18] that, for $\nu_k > m + 1$

$$\begin{aligned}
& \mathcal{E} \left\{ \mathbf{M}_k^2 | \mathbf{M}_p \right\} \\
&= \frac{(\nu_k - m)^2 \mathbf{M}_p^2 + (\nu_k - m) \text{Tr}\{\mathbf{M}_p\} \mathbf{M}_p}{(\nu_k - m + 1)(\nu_k - m - 1)}
\end{aligned} \tag{39}$$

$$\begin{aligned}
& \mathcal{E} \left\{ \text{Tr}\{\mathbf{M}_k\}^2 | \mathbf{M}_p \right\} \\
&= \frac{(\nu_k - m) \text{Tr}\{\mathbf{M}_p^2\} + (\nu_k - m)^2 \text{Tr}\{\mathbf{M}_p\}^2}{(\nu_k - m + 1)(\nu_k - m - 1)}.
\end{aligned} \tag{40}$$

Therefore, it follows that:

$$\begin{aligned}
& \text{Tr} \left\{ \mathcal{E} \{ \tilde{\mathbf{S}} \tilde{\mathbf{S}}^H | \mathbf{M}_p \} \right\} \\
&= \left(\sum_{k=1}^K w_k L_k \right)^2 \text{Tr} \left\{ \mathbf{M}_p^2 \right\} \\
&+ \sum_{k=1}^K w_k^2 L_k^2 \frac{\text{Tr} \left\{ \mathbf{M}_p^2 \right\} + (\nu_k - m) \text{Tr}\{\mathbf{M}_p\}^2}{(\nu_k - m + 1)(\nu_k - m - 1)} \\
&+ \sum_{k=1}^K w_k^2 L_k \frac{(\nu_k - m) \text{Tr} \left\{ \mathbf{M}_p^2 \right\} + (\nu_k - m)^2 \text{Tr}\{\mathbf{M}_p\}^2}{(\nu_k - m + 1)(\nu_k - m - 1)}.
\end{aligned} \tag{41}$$

Finally, the MSE of $\tilde{\mathbf{S}}$, conditionally to \mathbf{M}_p , is given by

$$\begin{aligned}
& \mathcal{E} \left\{ \|\tilde{\mathbf{S}} - \mathbf{M}_p\|^2 | \mathbf{M}_p \right\} \\
&= \left[\left(\sum_{k=1}^K w_k L_k \right) - 1 \right]^2 \text{Tr} \left\{ \mathbf{M}_p^2 \right\} \\
&+ \sum_{k=1}^K w_k^2 L_k^2 \frac{\text{Tr} \left\{ \mathbf{M}_p^2 \right\} + (\nu_k - m) \text{Tr}\{\mathbf{M}_p\}^2}{(\nu_k - m + 1)(\nu_k - m - 1)} \\
&+ \sum_{k=1}^K w_k^2 L_k \frac{(\nu_k - m) \text{Tr} \left\{ \mathbf{M}_p^2 \right\} + (\nu_k - m)^2 \text{Tr}\{\mathbf{M}_p\}^2}{(\nu_k - m + 1)(\nu_k - m - 1)}.
\end{aligned} \tag{42}$$

We now average the previous equation with respect to the prior distribution of \mathbf{M}_p . Since $\mathbf{M}_p | \bar{\mathbf{M}}_p \sim \mathcal{CW}_m(\mu^{-1} \bar{\mathbf{M}}_p, \mu)$, we have from [18]

$$\mathcal{E}_{\mathbf{M}_p} \left\{ \text{Tr} \left\{ \mathbf{M}_p^2 \right\} \right\} = \text{Tr} \left\{ \bar{\mathbf{M}}_p^2 \right\} + \mu^{-1} \text{Tr}\{\bar{\mathbf{M}}_p\}^2 \tag{43}$$

$$\mathcal{E}_{\mathbf{M}_p} \left\{ \text{Tr}\{\mathbf{M}_p\}^2 \right\} = \mu^{-1} \text{Tr} \left\{ \bar{\mathbf{M}}_p^2 \right\} + \text{Tr}\{\bar{\mathbf{M}}_p\}^2. \tag{44}$$

Reporting these values in the conditional MSE, and after some straightforward manipulations, we end up with

$$\begin{aligned}
& \mathcal{E} \left\{ \|\tilde{\mathbf{S}} - \mathbf{M}_p\|^2 \right\} \\
&= \left[\left(\sum_{k=1}^K w_k L_k \right) - 1 \right]^2 \\
&\times \left[\text{Tr} \left\{ \bar{\mathbf{M}}_p^2 \right\} + \mu^{-1} \text{Tr}\{\bar{\mathbf{M}}_p\}^2 \right] \\
&+ \sum_{k=1}^K w_k^2 \frac{(\nu_k + L_k - m) L_k}{(\nu_k - m + 1)(\nu_k - m - 1)} \\
&\times \left[(\nu_k - m + \mu^{-1}) \text{Tr}\{\bar{\mathbf{M}}_p\}^2 \right. \\
&\quad \left. + [\mu^{-1}(\nu_k - m) + 1] \text{Tr} \left\{ \bar{\mathbf{M}}_p^2 \right\} \right]
\end{aligned} \tag{45}$$

which concludes the proof.

REFERENCES

- [1] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.
- [2] I. S. Reed, J. D. Mallett, and L. E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 10, no. 6, pp. 853–863, Nov. 1974.
- [3] 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.
- [4] 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.
- [5] W. L. Melvin, "A STAP overview," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pt. 2, pp. 19–35, Jan. 2004.
- [6] R. Klemm, *Principles of Space-Time Adaptive Processing*, ser. IEE Radar, Sonar, Navigation and Avionics Series 12. London, U.K.: The Inst. Elect. Eng., 2002.
- [7] J. R. Guerci and E. J. Baranoski, "Knowledge-aided adaptive radar at DARPA—An overview," *IEEE Signal Process. Mag.*, vol. 23, no. 1, pp. 41–50, Jan. 2006.
- [8] W. L. Melvin and J. R. Guerci, "Knowledge-aided signal processing: A new paradigm for radar and other advanced sensors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 3, pp. 983–996, Jul. 2006.

- [9] K. Gerlach and M. L. Picciolo, "Airborne/spacebased radar STAP using a structured covariance matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 1, pp. 269–281, Jan. 2003.
- [10] S. Blunt, K. Gerlach, and M. Rangaswamy, "STAP using knowledge-aided covariance estimation and the FRACTA algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 3, pp. 1043–1057, Jul. 2006.
- [11] W. L. Melvin and G. A. Showman, "An approach to knowledge-aided covariance estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 3, pp. 1021–1042, Jul. 2006.
- [12] J. S. Bergin, C. M. Teixeira, P. M. Techau, and J. R. Guerci, "Improved clutter mitigation performance using knowledge-aided space-time adaptive processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 42, no. 3, pp. 997–1009, Jul. 2006.
- [13] L. Svensson, "Bayesian inference with unknown noise covariance," Ph.D. dissertation, Chalmers Univ. Technol., Göteborg, Sweden, Nov. 2004.
- [14] L. Svensson and M. Lundberg, "On posterior distributions for signals in Gaussian noise with unknown covariance matrix," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3554–3571, Sep. 2005.
- [15] S. Bidon, O. Besson, and J.-Y. Tournet, "A Bayesian approach to adaptive detection in non-homogeneous environments," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 205–217, Jan. 2007.
- [16] O. Besson, J.-Y. Tournet, and S. Bidon, "Bayesian estimation of covariance matrices in non-homogeneous environments," in *Proc. ICASSP*, Honolulu, HI, Apr. 15–20, 2007, vol. III, pp. 1037–1040.
- [17] C. T. Carpraro, G. T. Carpraro, I. Bradaric, D. D. Weiner, M. C. Wicks, and W. J. Baldygo, "Implementing digital terrain data in knowledge-aided space-time adaptive processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 3, pp. 1080–1099, Jul. 2006.
- [18] J. A. Tague and C. I. Caldwell, "Expectations of useful complex Wishart forms," *Multidimen. Syst. Signal Process.*, vol. 5, pp. 263–279, 1994.
- [19] J. Ward, *Space-Time Adaptive Processing for Airborne Radar*, Lincoln Lab., Mass. Inst. Technol., Lexington, MA, Tech. Rep. 1015, Dec. 1994.
- [20] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer-Verlag, 2004.
- [21] Y. I. Abramovich and A. I. Nevrev, "An analysis of effectiveness of adaptive maximization of the signal to noise ratio which utilizes the inversion of the estimated covariance matrix," *Radio Eng. Electron. Phys.*, vol. 26, pp. 67–74, Dec. 1981.
- [22] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, Jul. 1988.
- [23] S. T. Smith, "Covariance, subspace and intrinsic Cramér-Rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1610–1630, May 2005.
- [24] R. J. A. Little and D. B. Rubin, *Statistical Analysis With Missing Data*. New York: Wiley, 1987.
- [25] M. Rangaswamy, "Statistical analysis of the nonhomogeneity detector for non-Gaussian interference backgrounds," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2101–2111, Jun. 2005.
- [26] F. Pascal, P. Forster, J.-P. Ovarlez, and P. Larzabal, "Theoretical analysis of an improved covariance matrix estimator in non-Gaussian noise," in *Proc. ICASSP 2005*, Philadelphia, PA, Mar. 18–25, 2005, vol. IV, pp. 69–72.
- [27] F. Pascal, "Détection et estimation en environnement non Gaussien," (in French) Ph.D. dissertation, Univ. Paris X, Nanterre, France, Dec. 2006.
- [28] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed. New York: Chapman and Hall, 2004.

- [29] S. P. Brooks and A. Gelman, "General methods for monitoring convergence of iterative simulations," *J. Computat. Graph. Statist.*, vol. 7, no. 4, pp. 434–455, Dec. 1998.



Olivier Besson (S'90–M'92–SM'04) received the Ph.D. degree in signal processing in 1992, and the "Habilitation à Diriger des Recherches" in 1998, both from Institut National Polytechnique, Toulouse, France.

Since October 1993, he has been with the Department of Avionics and Systems of ENSICA, where he is now Professor and head of the Signal Processing Group. His research interests are in the general area of statistical signal and array processing with particular interest to robustness issues in detection/estimation problems for radar and communications.

Dr. Besson is a past Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS. He is a member of the IEEE Sensor Array and Multichannel (SAM) Technical Committee, and served as the co-technical chairman of the IEEE SAM2004 workshop. He has held visiting positions at Brigham Young University, Provo, UT, and the Università del Salento, Lecce, Italy.



Stéphanie Bidon (S'06) received the Engineer degree in aeronautics and the M.S. degree in signal processing from ENSICA, Toulouse, France, in 2004 and 2005, respectively.

She is currently pursuing the Ph.D. degree with the Department of Avionics and Systems, ENSICA. Her thesis deals with space-time adaptive processing for airborne radars.



Jean-Yves Tournet (M'94) received the ingénieur degree in electrical engineering from Ecole Nationale Supérieure d'Electronique, d'Electrotechnique, d'Informatique et d'Hydraulique (ENSEEIH), Toulouse, France. He received the Ph.D. degree from the National Polytechnic Institute, Toulouse, in 1992.

He is currently a Professor with the University of Toulouse (ENSEEIH), where he is a member of the IRIT Laboratory (UMR 5505 of the CNRS). His research activity is centered around estimation, detection, and classification of non-Gaussian and nonstationary processes.

Dr. Tournet was the Program Chair of the European Conference on Signal Processing (EUSIPCO), held in Toulouse in September 2002. He was also a member of the organizing committee for the International Conference on Acoustics, Speech, and Signal Processing, held in Toulouse in May 2006 (ICASSP'06). He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) committee of the IEEE Signal Processing Society.