BAYESIAN ESTIMATION OF COVARIANCE MATRICES IN NON-HOMOGENEOUS ENVIRONMENTS

Olivier Besson[*]*, Jean-Yves Tourneret*[†] *and Stéphanie Bidon*[*]

[*]Department of Avionics and Systems, ENSICA, Toulouse, France.
[†]Signal and Communications Group, IRIT/ENSEEIHT, Toulouse, France.

ABSTRACT

In many applications, it is required to detect, from a primary vector, the presence of a signal of interest embedded in noise with unknown statistics. We consider a situation where the training samples used to infer the noise statistics do not share the same covariance matrix as the vector under test. A Bayesian model is proposed where the covariance matrices of the primary and the secondary data are assumed to be random, with some appropriate joint distribution. The prior distributions of these matrices reflect a rough knowledge about the environment. Within this framework, the minimum mean-square error (MMSE) estimator and the the maximum a posteriori (MAP) estimator of the primary data covariance matrix are derived. A Gibbs sampling strategy is presented for the implementation of the MMSE estimator. Numerical simulations illustrate the performances of these estimators and compare them with those of the sample covariance matrix estimator.

Index Terms— Bayesian estimation, covariance matrix, detection, Gibbs sampler, inhomogeneities

1. INTRODUCTION

Detection of a signal of interest in a background of noise is a fundamental task in many applications, including radar, communications or sonar [1]. This is especially the case for radar systems whose core task is to detect a target amongst clutter, thermal noise and possibly jamming. Usually, the presence of a target, with given space and/or time signature $\boldsymbol{s}$, is sought in a (range) cell under test (CUT), given an observation vector $\boldsymbol{z}$ -the primary data- that corresponds to the output of an array of sensors. In the Gaussian case, when the covariance matrix $\boldsymbol{M}_p$ of the noise in the CUT is known, the optimal processor consists of a whitening step followed by matched filtering [1]. However, the statistics of the noise in the CUT are generally unknown and hence $\boldsymbol{M}_p$ must somehow be estimated. Thus, *central to most detection schemes is the problem of estimating the covariance matrix of the noise in the primary data*. This goal is generally achieved through the use of independent training samples $\boldsymbol{z}_k$, $k = 1, \ldots, K$, (the so-called secondary data), which consist of noise only, and would ideally share the same covariance matrix $\boldsymbol{M}_s$, equal to $\boldsymbol{M}_p$. The training samples are usually obtained from range cells close to the cell under test. The principle that underlies this approach is that information about noise in the primary data can be inferred from noise in the secondary data. This implies that the two sets of data "share" some common features. The most widely used assumption is that the noise is Gaussian distributed, and that $\boldsymbol{M}_p = \boldsymbol{M}_s$. In such a situation, $\boldsymbol{M}_p$ can be estimated from the sample covariance matrix of the secondary data, and then used in the optimal detector; this is the essence of the adaptive matched filter (AMF) [2]. A slight deviation to this ideal model is to assume that $\boldsymbol{M}_p$ is only proportional to $\boldsymbol{M}_s$; this is usually referred to as a partially homogeneous environment. In this framework, the generalized likelihood ratio test (GLRT) corresponds to the adaptive coherence estimator, whose properties have been thoroughly studied [3].

However, it has been evidenced that the homogeneous assumption is an idealized situation [4], and that the most commonly encountered situation is that of heterogeneous environments for which $\boldsymbol{M}_p \neq \boldsymbol{M}_s$. This can be due either to the terrain (highly complex and non-stationary clutter environments) or the geometry of the array (e.g. non-linear arrays or non side-looking configurations). Obviously, a pre-requisite to estimate $\boldsymbol{M}_p$ from the $\boldsymbol{z}_k$'s is to have a model relating $\boldsymbol{M}_p$ to $\boldsymbol{M}_s$. The most currently used approach consists of writing the clutter covariance matrix, at each range, as the integral -over clutter patches uniformly distributed in azimuth on an iso-range curve- of the covariance matrix of each clutter patch, weighted by the radar illumination pattern and the ground reflectivity [4, 5]. This model, if it allows one to compute the covariance matrix for each range, does not yield a direct relation between $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$. In addition, it does not enable one to simply model random mismatches between $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$.

In this paper, we propose a new model for non homogeneous environments. Towards this end, a Bayesian approach is advocated, as it is a relevant framework to handle uncertainties, and, at the same time, provides a theoretically sound way to define the relation between $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$. The present paper focuses on the estimation of $\boldsymbol{M}_p$, using this new model. The estimates derived herein can in turn be used for detection purposes, see [6].

2. PROBLEM STATEMENT

This section formulates the hypotheses regarding the data model proposed for non homogeneous environments. As explained previously, we have an observation vector $\boldsymbol{z} \in \mathbb{C}^{m \times 1}$, which consists of Gaussian noise $\boldsymbol{n}$ and possibly a useful signal $\alpha \boldsymbol{s}$, i.e. $\boldsymbol{z} = \alpha \boldsymbol{s} + \boldsymbol{n}$, where $\alpha = 0$ or $\alpha \neq 0$. The covariance matrix of $\boldsymbol{n}$ is $\boldsymbol{M}_p$. Our problem consists of estimating $\boldsymbol{M}_p$ from the observation of K training samples $\boldsymbol{z}_k$, whose covariance matrix $\boldsymbol{M}_s$ might differ from $\boldsymbol{M}_p$. We assume that the vectors $\boldsymbol{z}_k$ are independent and identically Gaussian distributed, $\boldsymbol{z}_k | \boldsymbol{M}_s \sim \mathcal{CN}_m(\boldsymbol{0}, \boldsymbol{M}_s)$, $k = 1, \cdots, K$, where $\boldsymbol{M}_s$ is unknown. Since the $\boldsymbol{z}_k$'s are independent, the joint density of $\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}_1 & \cdots & \boldsymbol{z}_K \end{bmatrix}$, conditionally to $\boldsymbol{M}_s$, is

$$f(\boldsymbol{Z} | \boldsymbol{M}_s) = \pi^{-mK} |\boldsymbol{M}_s|^{-K} \operatorname{etr}\{-\boldsymbol{M}_s^{-1} \boldsymbol{S}\}, \qquad (1)$$

This work was supported by the Délégation Générale pour l'Armement (DGA) and by Thales Systèmes Aéroportés.

where $\mathrm{etr}\{.\}$ stands for the exponential of the trace of the matrix between braces, and $\boldsymbol{S} = \sum_{k=1}^{K} \boldsymbol{z}_k \boldsymbol{z}_k^H$ denotes the sample covariance matrix of the secondary data. In order to model heterogeneity, we assume that $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$ are random, and that the distribution of $\boldsymbol{M}_s$ given $\boldsymbol{M}_p$ is known. More precisely, we assume that the conditional distribution of $\boldsymbol{M}_s|\boldsymbol{M}_p$ is an inverse complex Wishart distribution with $\nu > m$ degrees of freedom, whose mean is $\boldsymbol{M}_p$ [7]:

$$f(\boldsymbol{M}_s|\boldsymbol{M}_p) \propto |\boldsymbol{M}_s|^{-(\nu+m)} \, \mathrm{etr}\{-(\nu-m)\boldsymbol{M}_s^{-1}\boldsymbol{M}_p\}|\boldsymbol{M}_p|^{\nu}. \tag{2}$$

This distribution is denoted as $\boldsymbol{M}_s|\boldsymbol{M}_p \sim \mathcal{CW}_m^{-1}\left((\nu-m)\boldsymbol{M}_p, \nu\right)$. Note that the inverse complex Wishart distribution is the conjugate prior for parameter $\boldsymbol{M}_s$, which will significantly simplify the analysis. Note also that (2) implies that, "on the average", the environment is homogeneous as $\mathcal{E}\left\{\boldsymbol{M}_s|\boldsymbol{M}_p\right\} = \boldsymbol{M}_p$; however, these two matrices will be different with probability one. The parameter ν allows one to adjust the degree of heterogeneity between $\boldsymbol{M}_s$ and $\boldsymbol{M}_p$: when ν increases, $\boldsymbol{M}_s$ is closer to $\boldsymbol{M}_p$ [7].

Let us now turn to the a priori distribution of $\boldsymbol{M}_p$. The choice of this prior is of course a delicate issue. It is usually dictated by two seemingly conflicting arguments. On one hand, the prior $f(\boldsymbol{M}_p)$ should reflect our knowledge about the primary data covariance matrix, or our absence of knowledge, which can be recast through a non informative prior. On the other hand, computational complexity is an important issue. Consequently, the prior distribution of $\boldsymbol{M}_p$ is usually chosen in order to provide tractable posterior densities. We refer the reader to [8] for a very comprehensive discussion about the choice of a prior for covariance matrices. In our context, we assume that we have some rough knowledge about the average value of $\boldsymbol{M}_p$, denoted as $\bar{\boldsymbol{M}}_p$. The latter can be obtained, e.g. by using the model for the clutter covariance matrix described in [4, 5]. Therefore the prior distribution of $\boldsymbol{M}_p$ is supposed to be a complex Wishart distribution with $\mu \geq m$ degrees of freedom and mean $\bar{\boldsymbol{M}}_p$ [7], i.e.

$$f(\boldsymbol{M}_p) \propto |\boldsymbol{M}_p|^{\mu-m} \, \mathrm{etr}\{-\mu\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1}\}. \tag{3}$$

This will be denoted as $\boldsymbol{M}_p \sim \mathcal{CW}_m\left(\mu^{-1}\bar{\boldsymbol{M}}_p, \mu\right)$. As μ increases, $\boldsymbol{M}_p$ is closer to $\bar{\boldsymbol{M}}_p$, and thus the prior density $f(\boldsymbol{M}_p)$ is very informative. On the other hand, for small μ, $\boldsymbol{M}_p$ may significantly depart from $\bar{\boldsymbol{M}}_p$, which results in a vague prior density $f(\boldsymbol{M}_p)$. Hence, the scalar μ enables us to tune the amount of a priori knowledge we have about $\boldsymbol{M}_p$. Furthermore, it should be stressed that $\boldsymbol{M}_p$ will anyway differ from $\bar{\boldsymbol{M}}_p$. The framework we propose thus enter the class of knowledge-aided processing, which is recognized as one of the potentially more effective way to handle heterogeneities [9]. On the other hand, it includes rather heterogeneous environments.

3. ESTIMATION OF $\boldsymbol{M}_p$

We now focus on the problem of estimating $\boldsymbol{M}_p$ from $\boldsymbol{Z}$. As a preliminary step, we obtain the posterior distribution $f(\boldsymbol{M}_p|\boldsymbol{Z})$ which is needed to derive both the MMSE and MAP estimators.

3.1. Posterior distribution of $\boldsymbol{M}_p$

Under the stated hypotheses, the joint distribution of $(\boldsymbol{M}_p, \boldsymbol{M}_s)$, conditionally to $\boldsymbol{Z}$, is given by

$$\begin{aligned} f(\boldsymbol{M}_p, \boldsymbol{M}_s|\boldsymbol{Z}) &\propto f(\boldsymbol{Z}|\boldsymbol{M}_p, \boldsymbol{M}_s) f(\boldsymbol{M}_s|\boldsymbol{M}_p) f(\boldsymbol{M}_p) \\ &\propto |\boldsymbol{M}_s|^{-(\nu+m+K)} |\boldsymbol{M}_p|^{\nu+\mu-m} \\ &\times \mathrm{etr}\{-\boldsymbol{M}_s^{-1}\left[\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p\right]\} \mathrm{etr}\{-\mu\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1}\}. \end{aligned} \tag{4}$$

Integrating (4) with respect to $\boldsymbol{M}_s$, and using (2)-(3), one obtains

$$\begin{aligned} f(\boldsymbol{M}_p|\boldsymbol{Z}) &= \int f(\boldsymbol{M}_p, \boldsymbol{M}_s|\boldsymbol{Z}) \, d\boldsymbol{M}_s, \\ &\propto \frac{|\boldsymbol{M}_p|^{\nu+\mu-m}}{|\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p|^{\nu+K}} \mathrm{etr}\{-\mu\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1}\}, \\ &\propto \frac{|\boldsymbol{M}_p|^{\nu}}{|\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p|^{\nu+K}} f(\boldsymbol{M}_p). \end{aligned} \tag{5}$$

The posterior distribution $f(\boldsymbol{M}_p|\boldsymbol{Z})$ is now used to derive the MMSE and MAP estimators.

3.2. MMSE estimation of $\boldsymbol{M}_p$

The MMSE estimate of $\boldsymbol{M}_p$ is the mean of the posterior distribution, viz

$$\begin{aligned} &\int \boldsymbol{M}_p f(\boldsymbol{M}_p|\boldsymbol{Z}) \, d\boldsymbol{M}_p = \\ &\frac{\int |\boldsymbol{M}_p|^{\nu} |\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p|^{-(\nu+K)} \boldsymbol{M}_p f(\boldsymbol{M}_p) \, d\boldsymbol{M}_p}{\int |\boldsymbol{M}_p|^{\nu} |\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p|^{-(\nu+K)} f(\boldsymbol{M}_p) \, d\boldsymbol{M}_p}. \end{aligned} \tag{6}$$

Unfortunately, no analytical expressions for the integrals in (6) exist, and one must approximate them numerically. Deterministic methods are not appropriate here since these integrals involve functions of high dimensions ($\boldsymbol{M}_p$ is of size $m \times m$). In such situation, it is thus usual to resort to stochastic integration methods such as Markov chain Monte Carlo (MCMC) methods. These methods consist of generating samples distributed according to the posteriors of interest -in this case $f(\boldsymbol{M}_p|\boldsymbol{Z})$- and to average these samples to approximate the integrals to be computed. The interested reader is invited to consult [10] for more details. However, the generation of matrices distributed according to $f(\boldsymbol{M}_p|\boldsymbol{Z})$ is not straightforward, as (5) does not belong to any familiar class of distributions. Instead, this paper proposes to generate matrices $\boldsymbol{M}_p^{(i)}, \boldsymbol{M}_s^{(i)}$ (for $i = 1, \ldots, N_r$) distributed according to the joint distribution $f(\boldsymbol{M}_p, \boldsymbol{M}_s|\boldsymbol{Z})$, using a Gibbs sampling strategy. This recursive strategy has been described in several textbooks such as [10]. Having the matrix $\boldsymbol{M}_s^{(i)}$ at the i-th iteration, the generation of $\boldsymbol{M}_p^{(i+1)}$ and $\boldsymbol{M}_s^{(i+1)}$ is achieved as follows:

- generate $\boldsymbol{M}_p^{(i+1)}$ according to $f(\boldsymbol{M}_p|\boldsymbol{M}_s^{(i)}, \boldsymbol{Z})$,
- generate $\boldsymbol{M}_s^{(i+1)}$ according to $f(\boldsymbol{M}_s|\boldsymbol{M}_p^{(i+1)}, \boldsymbol{Z})$.

Using (4), the conditional distributions of $\boldsymbol{M}_p|\boldsymbol{M}_s, \boldsymbol{Z}$ and $\boldsymbol{M}_s|\boldsymbol{M}_p, \boldsymbol{Z}$ can be expressed as

$$\boldsymbol{M}_p|\boldsymbol{M}_s, \boldsymbol{Z} \sim \mathcal{CW}_m\left(\left[\mu\bar{\boldsymbol{M}}_p^{-1} + (\nu-m)\boldsymbol{M}_s^{-1}\right]^{-1}, \nu+\mu\right), \tag{7}$$

$$\boldsymbol{M}_s|\boldsymbol{M}_p, \boldsymbol{Z} \sim \mathcal{CW}_m^{-1}\left(\boldsymbol{S} + (\nu-m)\boldsymbol{M}_p, \nu+K\right). \tag{8}$$

Consequently, the Gibbs sampling strategy generates iteratively random matrices $\boldsymbol{M}_p$ and $\boldsymbol{M}_s$ drawn from (7) and (8). The scheme is repeated until convergence. Observe that generating matrices according to Wishart or inverse Wishart distributions is straightforward. As is well known [10], the matrices $(\boldsymbol{M}_p, \boldsymbol{M}_s)$ generated with the previous algorithm are asymptotically distributed according to $f(\boldsymbol{M}_p, \boldsymbol{M}_s|\boldsymbol{Z})$. Therefore, the MMSE can be approximated by averaging the "last" matrices generated by the Gibbs sampler. More precisely, the N_{bi} first matrices belonging to the so-called burn-in

period are not used for the estimation, yielding the following estimate

$$\int \boldsymbol{M}_p f(\boldsymbol{M}_p|\boldsymbol{Z})\, d\boldsymbol{M}_p \simeq \frac{1}{N_r} \sum_{i=N_{bi}+1}^{N_{bi}+N_r} \boldsymbol{M}_p^{(i)} \triangleq \widehat{\boldsymbol{M}}_p^{\text{MMSE}}, \quad (9)$$

where N_r is the number of iterations used for the estimation of $\boldsymbol{M}_p$.

3.3. MAP estimation

We now turn to the MAP estimator, which can be obtained by maximizing $f(\boldsymbol{M}_p|\boldsymbol{Z})$. Using (5), it follows that

$$\ln f(\boldsymbol{M}_p|\boldsymbol{Z}) = \text{const.} + (\nu+\mu-m)\ln|\boldsymbol{M}_p| - (\nu+K)\ln|\boldsymbol{S}+(\nu-m)\boldsymbol{M}_p| - \text{Tr}\left\{\mu\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1}\right\}. \quad (10)$$

Differentiating the previous equation and equating the result to zero yields

$$\mu(\nu-m)\boldsymbol{M}_p\bar{\boldsymbol{M}}_p^{-1}\boldsymbol{M}_p - (\nu+\mu-m)\boldsymbol{S} - \boldsymbol{M}_p\left[(\nu-m)(\mu-m-K)\boldsymbol{I} - \mu\bar{\boldsymbol{M}}_p^{-1}\boldsymbol{S}\right] = \boldsymbol{0}. \quad (11)$$

The previous equation is a quadratic matrix equation, which can be solved in closed-form, see [6]. More precisely, it can be shown that the MAP estimate is given by

$$\widehat{\boldsymbol{M}}_p^{\text{MAP}} = \bar{\boldsymbol{M}}_p^{1/2}\boldsymbol{U}\,\text{diag}\,(\lambda_k)\,\boldsymbol{U}^H\bar{\boldsymbol{M}}_p^{1/2}, \quad (12)$$

where $\bar{\boldsymbol{M}}_p^{1/2}$ stands for the Hermitian square-root of $\bar{\boldsymbol{M}}_p$, $\text{diag}\,(\lambda_k)$ is a diagonal matrix with diagonal entries λ_k, $\boldsymbol{U}$ is the matrix of the eigenvectors of $\tilde{\boldsymbol{S}} = \bar{\boldsymbol{M}}_p^{-1/2}\boldsymbol{S}\bar{\boldsymbol{M}}_p^{-1/2} = \boldsymbol{U}\text{diag}\,(\ell_k)\,\boldsymbol{U}^H$, and

$$\lambda_k = \left(\frac{\mu-m-K}{2\mu} - \frac{\ell_k}{2(\nu-m)}\right) + \sqrt{\left(\frac{\mu-m-K}{2\mu} - \frac{\ell_k}{2(\nu-m)}\right)^2 + \frac{\nu+\mu-m}{\mu(\nu-m)}\ell_k}. \quad (13)$$

In contrast to the MMSE estimator, the MAP estimator can be obtained directly, and is thus less computationally expensive.

4. NUMERICAL EXAMPLES

In this section we first study the convergence of the Gibbs sampler. Then we compare the performances of the Bayesian estimators (MAP and MMSE) to those obtained with the sample covariance matrix (SCM) estimator. In all simulations, we consider an array with $m = 8$ elements, and the average value of the nominal primary data covariance matrix is $\bar{\boldsymbol{M}}_p(k,\ell) = 0.9^{|k-\ell|}$. The number of training samples is $K = 2m = 16$.

4.1. Convergence analysis

It is known that the Gibbs-sampler provides random matrices that are asymptotically distributed according to the target distribution. However, a critical issue is to determine the numbers of iterations N_{bi} and N_r (for burn-in and computation, respectively) that are sufficient to have an accurate estimate of $\boldsymbol{M}_p$ with (9). Usually, a two-step procedure is used. First, a rough idea of the values of N_{bi} and N_r is obtained by evaluating the mean-square error (MSE) through the iterations. In a second step, more theoretically sound indicators, such as the potential scale reduction factor [10], are used to confirm or infirm the values selected in the first step. Such a procedure was applied in our case -see [6] for details- showing that a short burn-in period $N_{bi} = 20$ and $N_r = 100$ iterations were sufficient to ensure a good estimation of $\boldsymbol{M}_p$. Once these values are chosen, a rigorous way to assess convergence is to use the between-within variance criterion. The principle is to run M parallel chains of length (N_{bi}, N_r), with different initial values. Let $\boldsymbol{M}_p^{(i,j)}$ be the matrix obtained at the i-th iteration of the j-th chain and let us note

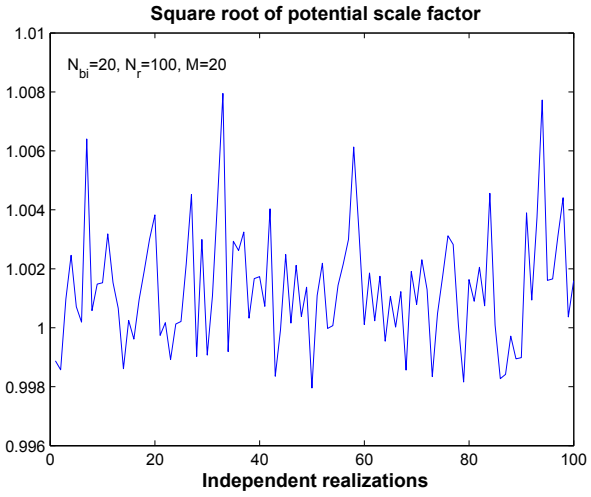

Fig. 1. Potential scale factor. $N_{bi} = 20$, $N_r = 100$ and $M = 20$. $\nu = m+1$, $\mu = m$ and $K = 2m$.

$$\widehat{\boldsymbol{M}}_p^{(.,j)} = N_r^{-1}\sum_{i=N_{bi}+1}^{N_{bi}+N_r} \boldsymbol{M}_p^{(i,j)}, \quad (14)$$

$$\widetilde{\boldsymbol{M}}_p = M^{-1}\sum_{j=1}^{M} \widehat{\boldsymbol{M}}_p^{(.,j)}, \quad (15)$$

where $\widehat{\boldsymbol{M}}_p^{(.,j)}$ corresponds to the MMSE estimate for the j-th chain, and $\widetilde{\boldsymbol{M}}_p$ is the average value over the M chains. The between-sequence and within-sequence variances for the M Markov chains corresponding to the (p,q) element of $\boldsymbol{M}_p$ (denoted as B_{pq} and W_{pq} respectively), are defined by

$$B_{pq} = \frac{N_r}{M-1}\sum_{j=1}^{M}([\widehat{\boldsymbol{M}}_p^{(.,j)}]_{pq} - [\tilde{\boldsymbol{M}}_p]_{pq})^2, \quad (16)$$

$$W_{pq} = \frac{1}{M}\sum_{j=1}^{M}\frac{1}{N_r-1}\sum_{i=N_{bi}+1}^{N_{bi}+N_r}([\boldsymbol{M}_p^{(i,j)}]_{pq} - [\widehat{\boldsymbol{M}}_p^{(.,j)}]_{pq})^2. \quad (17)$$

The convergence of the Gibbs sampler can be monitored by the so-called potential scale factor ρ_{pq} defined as [10]

$$\rho_{pq} = \frac{N_r-1}{N_r} + \frac{1}{N_r}\frac{B_{pq}}{W_{pq}}. \quad (18)$$

A value of $\sqrt{\rho_{pq}}$ less than 1.2 is recommended for convergence assessment in [10]. Figure 1 displays the value of $\sqrt{\rho_{11}}$ when $N_{bi} =$

20, $N_r = 100$ and $M = 20$. The experiment was run a hundred times and the values of $\sqrt{\rho_{11}}$ are plotted for these 100 independent realizations. It is clearly seen that these values of N_{bi} and N_r ensure convergence of the Gibbs-sampler, which validates our selection.

4.2. Estimation performance

This section compares the performance of the MMSE, MAP and SCM estimators. Note that the SCM estimator $K^{-1}\boldsymbol{S}$ is used for estimating $\boldsymbol{M}_p$ in homogeneous environments. Figures 2 and 3 display the MSEs obtained with the three estimation strategies versus μ, for two different values of ν, namely $\nu = m + 1$ and $\nu = 2m$. Note that, as ν increases, the environment is more homogeneous. Accordingly, when μ increases, the a priori knowledge about $\bar{\boldsymbol{M}}_p$ is more important. From inspection of these figures, it can be seen that the MMSE estimator always provides the best performance. The improvement compared to the MAP estimator is about $4 - 6$dB: it does not really depend on ν and it slightly decreases as μ increases. The improvement compared to the SCM estimator is about $6 - 8$dB when $\nu = 2m$ and $14 - 18$dB when $\nu = m + 1$, which is quite significant. The difference between the MMSE and the SCM is more pronounced when the environment is more heterogeneous, as could be expected. In fact, the performance of the SCM-based estimator degrades significantly as the environment becomes more heterogeneous, i.e. as ν decreases. In contrast, the performance of the MMSE estimator remains approximately constant when ν varies, which means that it can accommodate quite heterogeneous environments. Also, the MSE of the MMSE estimator decreases when μ increases, which is a direct consequence of the priori knowledge about $\bar{\boldsymbol{M}}_p$ being more pronounced. The performance of the SCM-based estimator is seen to depend weakly on μ.

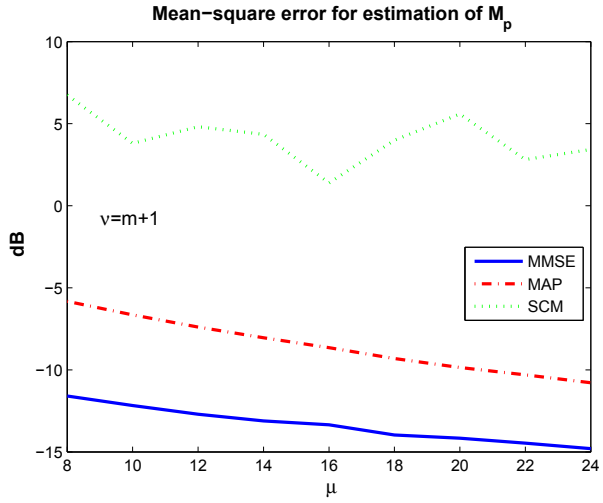

Fig. 2. MSE for estimation of $\boldsymbol{M}_p$ versus μ. $\nu = m + 1$

5. CONCLUSIONS

In this paper, a Bayesian framework was proposed to handle the case of heterogeneous environments, for which there exists a mismatch between the covariance matrix of the primary data and that of the training samples used for adaptation. A flexible model was presented where both the importance of the a priori knowledge and the degree of heterogeneity can be set through scalar variables. Within this framework, the MMSE and MAP estimators of the primary data covariance matrix using training samples were derived. A Gibbs sampling strategy was presented to implement the MMSE estimator. The latter estimator enables one to significantly improve performance compared to the usual sample covariance matrix estimator, and was also shown to outperform the MAP estimator.

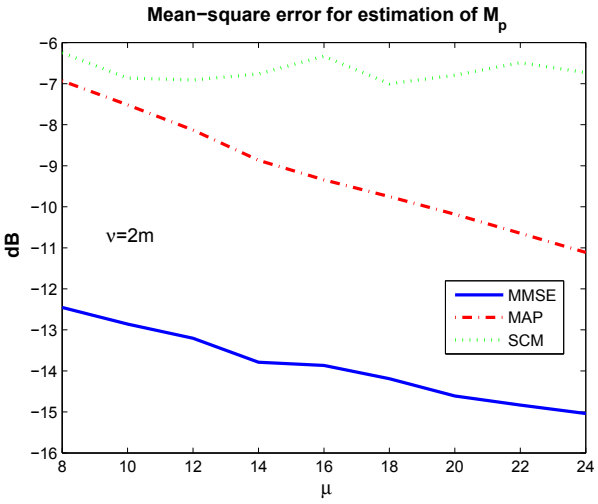

Fig. 3. MSE for estimation of $\boldsymbol{M}_p$ versus μ. $\nu = 2m$

6. REFERENCES

[1] L. L. Scharf. *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*. Addison Wesley, Reading, MA, 1991.

[2] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg. A CFAR adaptive matched filter detector. *IEEE Transactions Aerospace Electronic Systems*, 28(1):208–216, January 1992.

[3] S. Kraut, L. L. Scharf, and R. W. Butler. The adaptive coherence estimator: A uniformly most powerful invariant adaptive detection statistic. *IEEE Transactions Signal Processing*, 53(2):427–438, February 2005.

[4] W. L. Melvin. Space-time adaptive radar performance in heterogeneous clutter. *IEEE Transactions Aerospace Electronic Systems*, 36(2):621–633, April 2000.

[5] J. Ward. Space-time adaptive processing for airborne radar. Technical Report 1015, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, December 1994.

[6] S. Bidon, O. Besson, and J.-Y. Tourneret. A Bayesian approach to adaptive detection in non-homogeneous environments. *IEEE Transactions Signal Processing*, 2006. submitted for publication.

[7] J. A. Tague and C. I. Caldwell. Expectations of useful complex Wishart forms. *Multidimensional Systems and Signal Processing*, 5:263–279, 1994.

[8] L. Svensson and M. Lundberg. On posterior distributions for signals in Gaussian noise with unknown covariance matrix. *IEEE Transactions Signal Processing*, 53(9):3554–3571, September 2005.

[9] J. R. Guerci and E. J. Baranoski. Knowledge-aided adaptive radar at DARPA - An overview. *IEEE Signal Processing Magazine*, 23(1):41–50, January 2006.

[10] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, London, 1995.