



HAL
open science

On the Expected Likelihood Approach for Assessment of Regularization Covariance Matrix

Yuri Abramovich, Olivier Besson

► **To cite this version:**

Yuri Abramovich, Olivier Besson. On the Expected Likelihood Approach for Assessment of Regularization Covariance Matrix. *IEEE Signal Processing Letters*, 2015, vol. 22 (n° 6), pp. 777-781. 10.1109/LSP.2014.2369232 . hal-01131285

HAL Id: hal-01131285

<https://hal.science/hal-01131285>

Submitted on 13 Mar 2015

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.



Open Archive Toulouse Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <http://oatao.univ-toulouse.fr/>
Eprints ID: 13622

To link to this article: DOI: 10.1109/LSP.2014.2369232

URL: <http://dx.doi.org/10.1109/LSP.2014.2369232>

To cite this version: Abramovich, Yuri and Besson, Olivier *On the Expected Likelihood Approach for Assessment of Regularization Covariance Matrix*. (2015) IEEE Signal Processing Letters, vol. 22 (n° 6). pp. 777-781. ISSN 1070-9908

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@inp-toulouse.fr

On the Expected Likelihood Approach for Assessment of Regularization Covariance Matrix

Yuri I. Abramovich, *Fellow, IEEE*, and Olivier Besson, *Senior Member, IEEE*

Abstract—Regularization, which consists in shrinkage of the sample covariance matrix to a target matrix, is a commonly used and effective technique in low sample support covariance matrix estimation. Usually, a target matrix is chosen and optimization of the shrinkage factor is carried out, based on some relevant metric. In this letter, we rather address the choice of the target matrix. More precisely, we aim at evaluating, from observation of the data matrix, whether a given target matrix is a good regularizer. Towards this end, the expected likelihood (EL) approach is investigated. At a first step, we re-interpret the regularized covariance matrix estimate as the minimum mean-square error estimate in a Bayesian model where the target matrix serves as a prior. The likelihood function of the data is then derived, and the EL principle is subsequently applied. Over-sampled and under-sampled scenarios are considered.

Index Terms—Covariance matrix estimation, expected likelihood, regularization.

I. PROBLEM STATEMENT AND MOTIVATION

ESTIMATION of the covariance matrix $\mathbf{R} = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}$ of a random vector $\mathbf{x} \in \mathbb{C}^M$ from a finite number T of independent observations $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_T]$ is a fundamental problem in many engineering applications. For instance, in adaptive radar detection where it is desired to detect a target buried in Gaussian noise, the optimal filter depends on the noise covariance matrix, and the latter is usually estimated from training samples which contains noise only noise only [1], [2]. However, substituting the sample covariance matrix (SCM) $T^{-1}\mathbf{X}\mathbf{X}^H$ for \mathbf{R} in the optimal filter results in a significant loss in terms of output signal to noise ratio (SNR) [3], [4]. Indeed, the corresponding SNR loss is beta distributed and approximately $T = 2M$ samples are required to achieve an average SNR loss less than 3 dB. In cases where M is large, this number can be prohibitive and it is more customary to have to operate in low sample support. To cope with such situations, a widely used technique consists in regularization, or shrinkage of the SCM

towards a given matrix $\bar{\mathbf{R}}$, i.e., estimate \mathbf{R} as, see e.g., [5], [6], [7], [8]

$$\hat{\mathbf{R}}(\beta) = (1 - \beta)\frac{\mathbf{X}\mathbf{X}^H}{T} + \beta\bar{\mathbf{R}} \quad (1)$$

where $\beta \in [0, 1]$ and $\bar{\mathbf{R}}$ is some matrix, which is deemed to be close to \mathbf{R} [6] or is meant at regularizing the problem. The particular case of $\bar{\mathbf{R}} = \mathbf{I}$ (often referred to as diagonal loading) has proven to be particularly effective, especially when the noise has a high-power low-rank component plus a white noise component [9], [10], [11]. However, other possible choices are possible, including colored loading [12], [13]. Most often, $\bar{\mathbf{R}}$ is fixed and its choice is not questioned: rather, the focus is on optimization of β , so as, for instance, to achieve minimum mean-square error of the estimate in (1). Even if β is adequately selected, the choice of $\bar{\mathbf{R}}$ may also be very influential. For instance, diagonal loading is known to perform better when the eigenspectrum of the covariance matrix consists of a few dominant eigenvalues plus a fixed floor: it may not be as effective when the eigenspectrum has a smoothly decreasing profile.

In this paper, we address the selection of $\bar{\mathbf{R}}$. More specifically, we wish to examine whether $\bar{\mathbf{R}}$ is a good choice as a regularizer, from observation of \mathbf{X} . Towards this end, we propose to use the expected likelihood (EL) approach [14], [15], [16] to assess the plausibility of $\bar{\mathbf{R}}$. The EL was introduced as a tool to assess the quality of a covariance matrix estimate, say for example $\hat{\mathbf{R}}(\beta)$. It relies on some invariance properties of the likelihood ratio (LR) for testing if $\mathcal{E}\{\mathbf{X}\mathbf{X}^H\} = \hat{\mathbf{R}}(\beta)$. In [14], [15], [16], it was proved that the LR, evaluated at the true covariance matrix of \mathbf{X} , has a distribution that only depends on M and T . This property led the authors of [14], [15], [16] to select β such that the value of $LR(\hat{\mathbf{R}}(\beta)|\mathbf{X})$ is commensurate with that taken at the true covariance matrix. In this paper, we investigate using the EL approach not for selection of β , rather for that of $\bar{\mathbf{R}}$.

II. ASSESSMENT OF $\bar{\mathbf{R}}$ THROUGH EXPECTED LIKELIHOOD

In this section, we use the EL approach to evaluate the plausibility of the regularization covariance matrix $\bar{\mathbf{R}}$. First, the estimate in (1) is re-interpreted as the minimum mean-square error (MMSE) estimate of \mathbf{R} in a Bayesian framework where $\bar{\mathbf{R}}$ serves as a prior covariance. Then, the likelihood $p(\mathbf{X}|\bar{\mathbf{R}})$ corresponding to this model is derived and shown to be of a multivariate Student distribution type. Finally, the EL approach is applied to this Student distribution. Let us assume that the columns of \mathbf{X} are independent and identically distributed random vectors drawn from a zero-mean complex multivariate

Y. Abramovich is with WR Systems, Ltd., Fairfax, VA 22030 USA (e-mail: yabramovich@wrsystems.com).

O. Besson is with the University of Toulouse, ISAE-Supaero, Department Electronics Optronics Signal, Toulouse, France (e-mail: olivier.besson@isae-supaero.fr).

Gaussian distribution, which we denote as $\mathbf{X} \sim \mathbb{C}\mathcal{N}(\mathbf{o}, \mathbf{R}, \mathbf{I}_T)$. Then, the probability density function (p.d.f.) of \mathbf{X} is given by

$$p(\mathbf{X}|\mathbf{R}) = \pi^{-MT} |\mathbf{R}|^{-T} \text{etr} \{-\mathbf{X}^H \mathbf{R}^{-1} \mathbf{X}\} \quad (2)$$

where $\text{etr}\{\cdot\}$ stands for the exponential of the trace. Suppose now that \mathbf{R} is drawn from an inverse Wishart distribution with ν degrees of freedom and mean $\bar{\mathbf{R}}$, i.e., its p.d.f. is given by

$$\pi(\mathbf{R}|\bar{\mathbf{R}}) \propto |(\nu - M)\bar{\mathbf{R}}|^\nu |\mathbf{R}|^{-(\nu+M)} \text{etr} \{-(\nu - M)\bar{\mathbf{R}}\mathbf{R}^{-1}\} \quad (3)$$

where \propto means proportional to. We denote this distribution as $\mathbf{R}|\bar{\mathbf{R}} \sim \mathbb{C}\mathcal{W}_M^{-1}(\nu, (\nu - M)\bar{\mathbf{R}})$. Then, the posterior distribution of $\mathbf{R}|\mathbf{X}$ is

$$p(\mathbf{R}|\mathbf{X}) \propto |\mathbf{R}|^{-(\nu+T+M)} \text{etr} \{-\mathbf{R}^{-1} [\mathbf{S} + (\nu - M)\bar{\mathbf{R}}]\} \quad (4)$$

with $\mathbf{S} = \mathbf{X}\mathbf{X}^H$. Therefore, $\mathbf{R}|\mathbf{X} \sim \mathbb{C}\mathcal{W}_M^{-1}(\nu + T, \mathbf{S} + (\nu - M)\bar{\mathbf{R}})$ and the MMSE of \mathbf{R} is the mean of (4), i.e., [17]

$$\begin{aligned} \mathbf{R}_{MMSE} &= \frac{1}{\nu + T - M} [\mathbf{S} + (\nu - M)\bar{\mathbf{R}}] \\ &= (1 - \beta) \frac{\mathbf{X}\mathbf{X}^H}{T} + \beta \bar{\mathbf{R}}; \quad \beta = \frac{\nu - M}{\nu + T - M} \end{aligned} \quad (5)$$

which is exactly of the form (1). It follows that the regularizing matrix $\bar{\mathbf{R}}$ in (1) is equivalent to a prior covariance matrix in the Bayesian model (2)-(3). Choosing $\bar{\mathbf{R}}$ thus amounts to choosing a prior covariance matrix.

Next, this interpretation paves the way to using the EL approach. At first glance, it is not obvious why and how the EL approach could be used in the purpose of testing the plausibility of $\bar{\mathbf{R}}$ in the Bayesian hierarchical model described by (2)-(3), as the latter is quite different from the framework the EL approach was originally based upon. However, one should observe that $\mathcal{E}\{\mathbf{X}\mathbf{X}^H\} = \mathcal{E}\{\mathcal{E}\{\mathbf{X}\mathbf{X}^H|\mathbf{R}\}\} = \mathcal{E}\{\mathbf{R}\} = \bar{\mathbf{R}}$, and hence $\bar{\mathbf{R}}$ is the ‘‘average’’ covariance matrix of \mathbf{X} . Additionally, the p.d.f. of \mathbf{X} can be written as [18]

$$\begin{aligned} p(\mathbf{X}|\bar{\mathbf{R}}) &\propto |(\nu - M)\bar{\mathbf{R}}|^\nu \\ &\times \int |\mathbf{R}|^{-(\nu+T+M)} \text{etr} \{-\mathbf{R}^{-1} [\mathbf{S} + (\nu - M)\bar{\mathbf{R}}]\} d\mathbf{R} \\ &\propto |(\nu - M)\bar{\mathbf{R}}|^{-T} |\mathbf{I}_M + (\nu - M)^{-1} \bar{\mathbf{R}}^{-1} \mathbf{X}\mathbf{X}^H|^{-(\nu+T)} \end{aligned} \quad (6)$$

which is recognized as a multivariate Student distribution with $\nu + T - M$ degrees of freedom and parameter matrix $(\nu - M)\bar{\mathbf{R}}$ [19], [20]. Before pursuing our derivations, we would like to offer the following comments. Let $\bar{\mathbf{G}}$ denote a square-root of $\bar{\mathbf{R}}$, i.e., $\bar{\mathbf{G}}\bar{\mathbf{G}}^H = \bar{\mathbf{R}}$. In the Bayesian model (2)-(3), one has [18], [19] $\mathbf{R} \stackrel{d}{=} (\nu - M)\bar{\mathbf{G}}\mathbf{W}_{M,\nu}^{-1}\bar{\mathbf{G}}^H$ where $\stackrel{d}{=}$ means ‘‘is distributed as’’. In the previous equation, $\mathbf{W}_{M,\nu}$ follows a Wishart distribution with ν degrees of freedom and parameter matrix \mathbf{I}_M , i.e., $p(\mathbf{W}_{M,\nu}) \propto |\mathbf{W}_{M,\nu}|^{\nu-M} \text{etr}\{-\mathbf{W}_{M,\nu}\}$. We denote the Wishart distribution as $\mathbf{W}_{M,\nu} \sim \mathbb{C}\mathcal{W}_M(\nu, \mathbf{I}_M)$. It ensues that

$$\mathbf{X} \stackrel{d}{=} (\nu - M)^{1/2} \bar{\mathbf{G}} [\mathbf{W}_{M,\nu}^{-1}]^{1/2} \mathbf{N}_{M,T} \quad (7)$$

with $\mathbf{N}_{M,T} \sim \mathbb{C}\mathcal{N}(\mathbf{o}, \mathbf{I}_M, \mathbf{I}_T)$ independent of $\mathbf{W}_{M,\nu}$. In contrast, (6) yields the representation [19], [20]

$$\mathbf{X} \stackrel{d}{=} (\nu - M)^{1/2} \bar{\mathbf{G}} \mathbf{N}_{M,T} [\mathbf{W}_{T,\nu+T-M}]^{-1/2} \quad (8)$$

where $\mathbf{W}_{T,\nu+T-M} \sim \mathbb{C}\mathcal{W}_T(\nu + T - M, \mathbf{I}_T)$. Albeit the two mechanisms for generating \mathbf{X} are different, from a likelihood point of view the two representations are equivalent, as far as only assessment of $\bar{\mathbf{R}}$ from $p(\mathbf{X}|\bar{\mathbf{R}})$ is involved. Of course, in the Bayesian model (2)-(3), $\bar{\mathbf{R}}$ serves as a prior, and interest is on estimating \mathbf{R} , while in (6), $\bar{\mathbf{R}}$ is viewed as the covariance matrix in a Student distribution. Nevertheless, from our perspective of evaluating the plausibility of $\bar{\mathbf{R}}$, we will be using (6) and therefore the EL approach can be advocated. Observe that the difference compared to the original Gaussian frequentist framework of [14], [15], [16], is that one needs to deal with a Bayesian hierarchical framework which results in a non Gaussian likelihood $p(\mathbf{X}|\bar{\mathbf{R}})$. This being so, the EL approach was recently extended to the class of elliptically contoured distributions (ECD) [21], [22], [23] in [24], [25], [26]. Herein, we build upon the results of [26] with a few differences due to the fact that $\bar{\mathbf{R}}$ is the average covariance in a Bayesian framework. We first investigate the over-sampled case ($T \geq M$), then the under-sampled case ($T \leq M$) which deserves a specific treatment.

A. Over-Sampled Case

When $T \geq M$, the (generalized) likelihood ratio for a candidate $\bar{\mathbf{R}}_c$ is given by

$$LR(\bar{\mathbf{R}}_c) = \frac{p(\mathbf{X}|\bar{\mathbf{R}}_c)}{\max_{\bar{\mathbf{R}}} p(\mathbf{X}|\bar{\mathbf{R}})} = \frac{p(\mathbf{X}|\bar{\mathbf{R}}_c)}{p(\mathbf{X}|\bar{\mathbf{R}}_{ML})} \quad (9)$$

where $\bar{\mathbf{R}}_{ML}$ is the MLE of $\bar{\mathbf{R}}$, given by [21]

$$(\nu - M)\bar{\mathbf{R}}_{ML} = \frac{\nu}{T} \mathbf{X}\mathbf{X}^H. \quad (10)$$

Therefore, the LR for the candidate $\bar{\mathbf{R}}_c$ can be rewritten as

$$\begin{aligned} LR(\bar{\mathbf{R}}_c) &= \frac{|(\nu - M)\bar{\mathbf{R}}_c|^{-T} |\mathbf{I}_M + \frac{\bar{\mathbf{R}}_c^{-1} \mathbf{X}\mathbf{X}^H}{\nu - M}|^{-(\nu+T)}}{|(\nu - M)\bar{\mathbf{R}}_{ML}|^{-T} |\mathbf{I}_M + \frac{\bar{\mathbf{R}}_{ML}^{-1} \mathbf{X}\mathbf{X}^H}{\nu - M}|^{-(\nu+T)}} \\ &= C(\nu, M, T) |(\nu - M)^{-1} \bar{\mathbf{R}}_c^{-1} \mathbf{X}\mathbf{X}^H|^T \\ &\times |\mathbf{I}_M + (\nu - M)^{-1} \bar{\mathbf{R}}_c^{-1} \mathbf{X}\mathbf{X}^H|^{-(\nu+T)} \end{aligned} \quad (11)$$

where $C(\nu, M, T) = (1 + \nu^{-1}T)^{M(\nu+T)} (\frac{\nu}{T})^{MT}$. Let us now evaluate this LR at the true matrix $\bar{\mathbf{R}}$. From (7)-(8), it ensues that

$$\begin{aligned} LR(\bar{\mathbf{R}}) &\stackrel{d}{=} C(\nu, M, T) |\mathbf{W}_{M,\nu}^{-1} \mathbf{W}_{M,T}|^T \\ &\times |\mathbf{I}_M + \mathbf{W}_{M,\nu}^{-1} \mathbf{W}_{M,T}|^{-(\nu+T)} \end{aligned} \quad (12a)$$

$$\begin{aligned} &\stackrel{d}{=} C(\nu, M, T) |\mathbf{N}_{M,T} \mathbf{W}_{T,\nu+T-M}^{-1} \mathbf{N}_{M,T}^H|^T \\ &\times |\mathbf{I}_M + \mathbf{N}_{M,T} \mathbf{W}_{T,\nu+T-M}^{-1} \mathbf{N}_{M,T}^H|^{-(\nu+T)} \end{aligned} \quad (12b)$$

where $\mathbf{W}_{M,T} \sim \mathcal{CW}_M(T, \mathbf{I}_M)$. Hence, the likelihood ratio, when evaluated at the true matrix $\bar{\mathbf{R}}$ (i.e., when $\mathbf{X} \sim \mathcal{CN}(\mathbf{o}, \mathbf{R}, \mathbf{I}_T)$ and $\mathbf{R} \sim \mathcal{CW}_M^{-1}(\nu, (\nu - M)\bar{\mathbf{R}})$) has a distribution that only depends on ν , M and T . This p.d.f. can thus be computed in advance and the LR for a candidate regularization matrix $\bar{\mathbf{R}}_c$, as given by (11), can be compared to, say, the median value of $LR(\bar{\mathbf{R}})$, to decide if $\bar{\mathbf{R}}_c$ is a “good” regularization matrix.

B. Under-Sampled Case

When the number of observations T is less than the size of the observation space, the above theory does no longer hold since, with probability one, the data matrix \mathbf{X} belongs to a subspace of dimension T . More precisely, if we let $\mathbf{X} = \mathbf{U}_T \Sigma_T \mathbf{V}_T^H$ be the thin singular value decomposition of \mathbf{X} , inference about the covariance matrix of \mathbf{X} can be made only in the subspace $\mathcal{R}(\mathbf{U}_T)$ spanned by the columns of \mathbf{U}_T [27]: in other words, only $\mathbf{U}_T^H \bar{\mathbf{R}} \mathbf{U}_T$ is identifiable. As argued in [27] for Gaussian settings, if one wants to assess $\bar{\mathbf{R}}_c$ as the covariance matrix of \mathbf{X} , at best one can test the “closest” matrix to $\bar{\mathbf{R}}_c$ in $\mathcal{R}(\mathbf{U}_T)$. The latter is given by $\mathbf{U}_T \bar{\mathbf{D}}_{c,T} \mathbf{U}_T^H$ with $\bar{\mathbf{D}}_{c,T} = [\mathbf{U}_T^H \bar{\mathbf{R}}_c^{-1} \mathbf{U}_T]^H$, and can be interpreted as a singular covariance matrix [27]. Therefore, the under-sampled scenario is closely related to distributions with singular covariance matrices: this is indeed the starting point of the EL approach when $T \leq M$, see e.g., [27], [25] for details.

Thus, let us start with the Student distribution (8) in the case where $\bar{\mathbf{R}}$ has rank R , i.e., $\bar{\mathbf{R}} = \mathbf{U}_R \bar{\mathbf{D}} \mathbf{U}_R^H$ where $\mathbf{U}_R^H \mathbf{U}_R = \mathbf{I}_R$ and $\bar{\mathbf{D}}$ is an arbitrary full-rank positive definite Hermitian matrix. We assume temporarily that \mathbf{U}_R is known (it will be replaced by \mathbf{U}_T when coming back to our original problem). Also, let \mathbf{U}_R^\perp be an orthonormal basis for the complement of \mathbf{U}_R , i.e., $(\mathbf{U}_R^\perp)^H \mathbf{U}_R^\perp = \mathbf{I}_{M-R}$ and $(\mathbf{U}_R^\perp)^H \mathbf{U}_R = \mathbf{o}$. Then, one can define a singular density on the set $\mathcal{S} = \{\mathbf{X} \in \mathbb{C}^{M \times T}; (\mathbf{U}_R^\perp)^H \mathbf{X} = \mathbf{o}\}$ as [28], [29]

$$p(\mathbf{X}|\bar{\mathbf{D}}) \propto |(\nu - M)\bar{\mathbf{D}}|^{-T} \times |\mathbf{I}_R + (\nu - M)^{-1} \bar{\mathbf{D}}^{-1} \mathbf{U}_R^H \mathbf{X} \mathbf{X}^H \mathbf{U}_R|^{-(\nu+T-M+R)}. \quad (13)$$

The MLE of $(\nu - M)\bar{\mathbf{D}}$ is given by, see (10), $(\nu - M + R)T^{-1} \mathbf{U}_R^H \mathbf{X} \mathbf{X}^H \mathbf{U}_R$. It follows that the likelihood ratio, for the candidate $\bar{\mathbf{D}}_c$ is given by (14) with $C^u(\nu, M, R, T) = \left(\frac{\nu - M + R}{T}\right)^{RT} (1 + (\nu - M + R)^{-1} T)^{R(\nu+T-M+R)}$. (See (14)–(15), shown at the bottom of the page.)

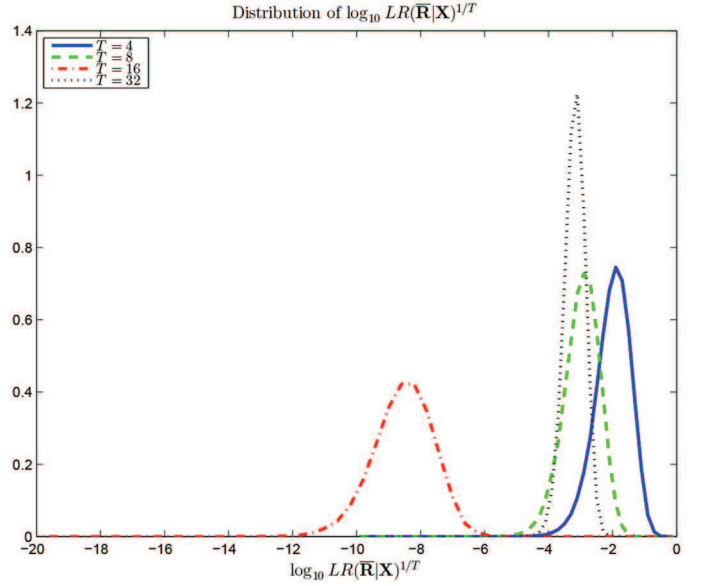


Fig. 1. Probability density function of $\log_{10} LR(\bar{\mathbf{R}}|\mathbf{X})^{1/T}$ for various T . $M = 16$ and $\nu = M + 1$.

Let us now come back to assessing a candidate matrix $\bar{\mathbf{R}}_c$: as argued before, one can only assess the closest matrix to $\bar{\mathbf{R}}_c$ in $\mathcal{R}(\mathbf{U}_T)$, namely $\mathbf{U}_T \bar{\mathbf{D}}_{c,T} \mathbf{U}_T^H$ with $\bar{\mathbf{D}}_{c,T} = [\mathbf{U}_T^H \bar{\mathbf{R}}_c^{-1} \mathbf{U}_T]^H$. The MLE of $(\nu - M)\bar{\mathbf{D}}_{c,T}$ is now given by $(\nu + T - M)T^{-1} \mathbf{U}_T^H \mathbf{X} \mathbf{X}^H \mathbf{U}_T = (\nu + T - M)T^{-1} \Lambda_T$. From (14), the likelihood ratio is thus given by (15) where we used the fact that $\mathbf{U}_T \mathbf{U}_T^H \mathbf{X} = \mathbf{X}$. When evaluated at the true $\bar{\mathbf{R}}$, the stochastic representations in (7)–(8) yield

$$LR^u(\bar{\mathbf{R}}) \stackrel{d}{=} C^u(\nu, M, T, T) |\mathbf{W}_{T, \nu+T-M}^{-1} \mathbf{W}_{T, M}|^T \times |\mathbf{I}_T + \mathbf{W}_{T, \nu+T-M}^{-1} \mathbf{W}_{T, M}|^{-(\nu+2T-M)} \quad (16a)$$

$$\stackrel{d}{=} C(\nu, M, T) |\mathbf{N}_{M, T}^H \mathbf{W}_{M, \nu}^{-1} \mathbf{N}_{M, T}|^T \times |\mathbf{I}_T + \mathbf{N}_{M, T}^H \mathbf{W}_{M, \nu}^{-1} \mathbf{N}_{M, T}|^{-(\nu+2T-M)}. \quad (16b)$$

III. NUMERICAL ILLUSTRATIONS

We now illustrate how the above procedure can be helpful in assessing the validity of a given prior (or regularization) matrix $\bar{\mathbf{R}}_c$. We consider a uniform linear array with $M = 16$ elements spaced a half wavelength apart. The data are generated according to the Bayesian model (2)–(3). In Fig. 1 we display the distribution of the log likelihood ratio for different values of

$$LR^u(\bar{\mathbf{D}}_c) = \frac{|(\nu - M)\bar{\mathbf{D}}_c|^{-T} |\mathbf{I}_R + (\nu - M)^{-1} \bar{\mathbf{D}}_c^{-1} \mathbf{U}_R^H \mathbf{X} \mathbf{X}^H \mathbf{U}_R|^{-(\nu+T-M+R)}}{\left|\frac{\nu - M + R}{T} \mathbf{U}_R^H \mathbf{X} \mathbf{X}^H \mathbf{U}_R\right|^{-T} (1 + (\nu - M + R)^{-1} T)^{R(\nu+T-M+R)}} \times |\mathbf{I}_R + (\nu - M)^{-1} \bar{\mathbf{D}}_c^{-1} \mathbf{U}_R^H \mathbf{X} \mathbf{X}^H \mathbf{U}_R|^{-(\nu+T-M+R)} \quad (14)$$

$$LR^u(\bar{\mathbf{D}}_{c,T}) = C^u(\nu, M, T, T) |(\nu - M)^{-1} \bar{\mathbf{D}}_{c,T}^{-1} \mathbf{U}_T^H \mathbf{X} \mathbf{X}^H \mathbf{U}_T|^T |\mathbf{I}_T + (\nu - M)^{-1} \bar{\mathbf{D}}_{c,T}^{-1} \mathbf{U}_T^H \mathbf{X} \mathbf{X}^H \mathbf{U}_T|^{-(\nu+2T-M)} \times |\mathbf{I}_T + (\nu - M)^{-1} \mathbf{X}^H \bar{\mathbf{R}}_c^{-1} \mathbf{X}|^{-(\nu+2T-M)} \quad (15)$$

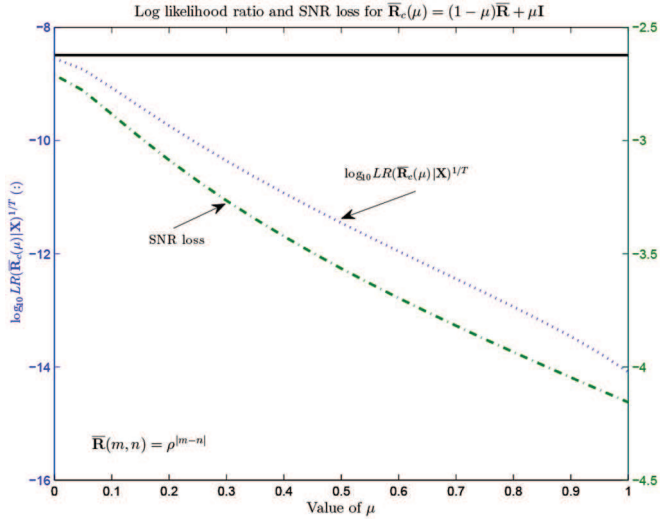


Fig. 2. Likelihood ratio of $\bar{\mathbf{R}}_c(\mu) = (1 - \mu)\bar{\mathbf{R}} + \mu\mathbf{I}$ and SNR loss of associated filter $\mathbf{w} \propto [\mathbf{S} + (\nu - M)\bar{\mathbf{R}}_c(\mu)]^{-1}\mathbf{s}$ versus μ . $M = 16$, $T = 16$ and $\nu = M + 1$. $\bar{\mathbf{R}}(m, n) = \rho^{|m-n|}$.

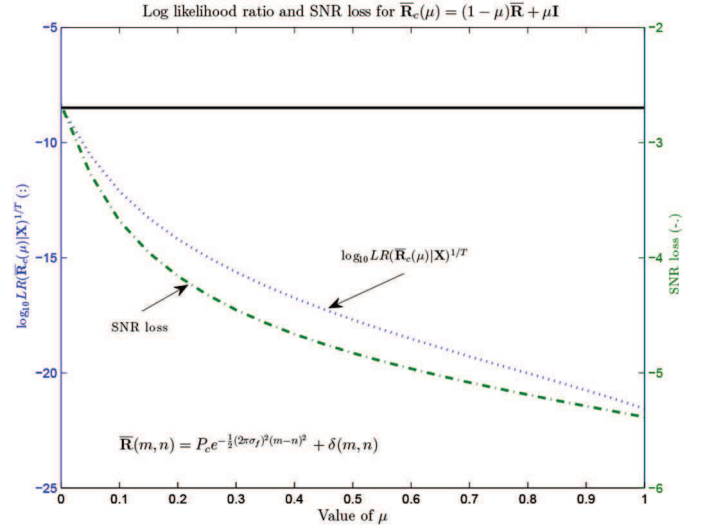


Fig. 3. Likelihood ratio of $\bar{\mathbf{R}}_c(\mu) = (1 - \mu)\bar{\mathbf{R}} + \mu\mathbf{I}$ and SNR loss of associated filter $\mathbf{w} \propto [\mathbf{S} + (\nu - M)\bar{\mathbf{R}}_c(\mu)]^{-1}\mathbf{s}$ versus μ . $M = 16$, $T = 16$ and $\nu = M + 1$. $\bar{\mathbf{R}}(m, n) = P_c \exp\{-\frac{1}{2}(2\pi\sigma_f)^2(m-n)^2\} + \delta(m, n)$.

T . Similarly to what was observed in [16], [25], the log likelihood ratio, when evaluated at the true $\bar{\mathbf{R}}$ takes very small values, and hence a candidate $\bar{\mathbf{R}}_c$ should be retained if its corresponding LR matches that of the true $\bar{\mathbf{R}}$. Let us now investigate if this procedure results in a “good” choice for $\bar{\mathbf{R}}_c$. We consider three types of covariance matrix $\bar{\mathbf{R}}$:

- 1) $\bar{\mathbf{R}}(m, n) = P_c e^{-\frac{1}{2}(2\pi\sigma_f)^2(m-n)^2} + \delta(m, n)$ with $\sigma_f = 0.01$ and $P_c = 100$.
- 2) $\bar{\mathbf{R}}(m, n) = \rho^{|m-n|}$ with $\rho = 0.9$.
- 3) $\bar{\mathbf{R}} = \sum_{j=1}^3 P_j \mathbf{a}(\theta_j) \mathbf{a}^H(\theta_j) + \mathbf{I}$ with $\mathbf{a}(\theta) = [1 e^{i\pi \sin \theta} \dots e^{i\pi(M-1) \sin \theta}]^T$, $\theta_j = -20^\circ, 20^\circ, 40^\circ$ and $10 \log_{10} P_j = 20, 20, 30$ dB

For each type, we consider as a candidate $\bar{\mathbf{R}}_c(\mu) = (1 - \mu)\bar{\mathbf{R}} + \mu\mathbf{I}$ with $\mu \in [0, 1]$ and we evaluate the mean value of $\log_{10} LR(\bar{\mathbf{R}}_c(\mu)|\mathbf{X})^{1/T}$. Note that, a priori, the best choice is $\bar{\mathbf{R}}_c(0) = \bar{\mathbf{R}}$. In order to assess $\bar{\mathbf{R}}_c(\mu)$, we consider the adaptive filter $\mathbf{w} \propto [\mathbf{S} + (\nu - M)\bar{\mathbf{R}}_c(\mu)]^{-1}\mathbf{s}$ where $\mathbf{s} = \mathbf{a}(5^\circ)$ is the signature of the signal of interest. The SNR loss, evaluated at the output of this adaptive filter, will serve as a figure of merit for assessment of $\bar{\mathbf{R}}_c(\mu)$. In Fig. 2–4, we display the mean value of $\log_{10} LR(\bar{\mathbf{R}}_c(\mu)|\mathbf{X})^{1/T}$ (the solid line represents the target value, namely the median value of $\log_{10} LR(\bar{\mathbf{R}}|\mathbf{X})^{1/T}$), as well as the SNR loss. These figures confirm two facts. Firstly, there is a good consistency between $LR(\bar{\mathbf{R}}_c(\mu)|\mathbf{X}) \simeq \text{med}[LR(\bar{\mathbf{R}}|\mathbf{X})^{1/T}]$ and the fact that the filter based on $\bar{\mathbf{R}}_c(\mu)$ is effective. In other words, selecting $\bar{\mathbf{R}}_c(\mu)$ from the EL principle helps finding a good regularization matrix and, subsequently, a performant adaptive filter. Secondly, diagonal loading is seen to be more effective in the case of a low-rank plus white noise type of covariance matrix: indeed, the LR remains close to $\text{med}[LR(\bar{\mathbf{R}}|\mathbf{X})^{1/T}]$ for a large range of values of μ , and so is for the SNR loss. In fact, choosing the identity matrix as a regularizer is as good as selecting the true $\bar{\mathbf{R}}$. In contrast, diagonal loading is less effective for the two other types of covariance matrix: when μ

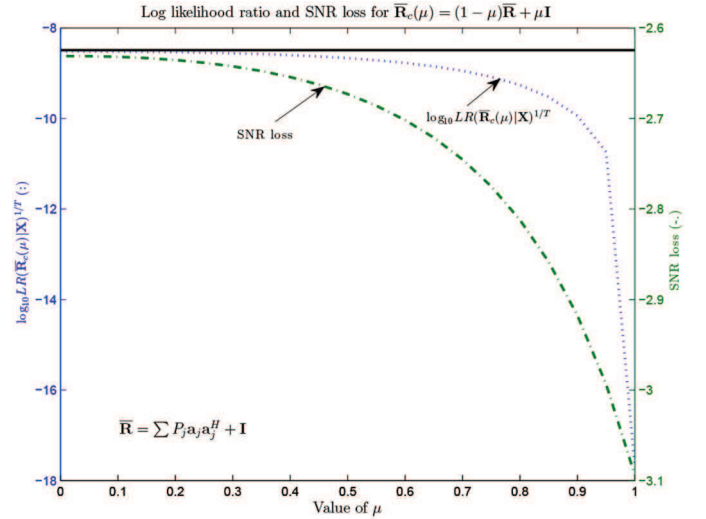


Fig. 4. Likelihood ratio of $\bar{\mathbf{R}}_c(\mu) = (1 - \mu)\bar{\mathbf{R}} + \mu\mathbf{I}$ and SNR loss of associated filter $\mathbf{w} \propto [\mathbf{S} + (\nu - M)\bar{\mathbf{R}}_c(\mu)]^{-1}\mathbf{s}$ versus μ . $M = 16$, $T = 16$ and $\nu = M + 1$. $\bar{\mathbf{R}} = \sum_{j=1}^3 P_j \mathbf{a}(\theta_j) \mathbf{a}^H(\theta_j) + \mathbf{I}$.

increases, the LR departs from its target value and SNR loss is worst.

IV. CONCLUSIONS

In this letter, we addressed the problem of selecting the regularization matrix $\bar{\mathbf{R}}$ in estimation schemes which consist of shrinkage of the sample covariance matrix to a given regularization matrix. We interpreted the latter as a prior covariance matrix in a Bayesian model. The likelihood function $p(\mathbf{X}|\bar{\mathbf{R}})$ of the latter was derived as a function of $\bar{\mathbf{R}}$, and the expected likelihood approach was advocated to assess the validity of $\bar{\mathbf{R}}$. It was shown that this approach is instrumental in providing a reliable measure of the quality of $\bar{\mathbf{R}}$. As a by-product, we showed that diagonal loading is effective only in special cases of covariance matrices, and the EL approach proposed was helpful in identifying these cases.

REFERENCES

- [1] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 115–127, Mar. 1986.
- [2] 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.
- [3] 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.
- [4] C. G. Khatri and C. R. Rao, "Effects of estimated noise covariance matrix in optimal signal detection," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 5, pp. 671–679, May 1987.
- [5] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivar. Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.
- [6] P. Stoica, J. Li, X. Zhu, and J. R. Guerci, "On using a priori knowledge in space-time adaptive processing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2598–2602, Jun. 2008.
- [7] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4097–4107, Sep. 2011.
- [8] E. Ollila and D. Tyler, "Regularized M-estimators of scatter matrix," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 6059–6070, Nov. 2014.
- [9] Y. I. Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum SNR," *Radio Eng. Electron. Phys.*, vol. 26, pp. 87–95, Mar. 1981.
- [10] 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.
- [11] O. P. Cheremisin, "Efficiency of adaptive algorithms with regularised sample covariance matrix," *Radio Eng. Electron. Phys.*, vol. 27, no. 10, pp. 69–77, 1982.
- [12] J. D. Hiemstra, "Colored diagonal loading," in *Conf. Proc. IEEE Radar*, 2002, pp. 386–390.
- [13] P. Wang, Z. Sahinoglu, M.-P. Pun, H. Li, and B. Himed, "Knowledge-aided adaptive coherence estimator in stochastic partially homogeneous environments," *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 193–196, Mar. 2011.
- [14] Y. Abramovich, N. Spencer, and A. Gorokhov, "Bounds on maximum likelihood ratio-Part I: Application to antenna array detection-estimation with perfect wavefront coherence," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1524–1536, Jun. 2004.
- [15] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "GLRT-based threshold detection-estimation performance improvement and application to uniform circular antenna arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 20–31, Jan. 2007.
- [16] "Modified GLRT and AMF framework for adaptive detectors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 1017–1051, Jul. 2007.
- [17] J. A. Tague and C. I. Caldwell, "Expectations of useful complex Wishart forms," *Multidimen. Syst. Signal Process.*, vol. 5, pp. 263–279, 1994.
- [18] R. J. Muirhead, *Aspects of Multivariate Statistical Theory.* New York, NY, USA: Wiley, 1982.
- [19] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions.* Boca Raton, FL, USA: Chapman & Hall/CRC, 2000.
- [20] S. Kotz and S. Nadarajah, *Multivariate t Distributions and their applications.* Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [21] K. T. Fang and Y. T. Zhang, *Generalized Multivariate Analysis.* Berlin, Germany: Springer Verlag, 1990.
- [22] T. W. Anderson and K.-T. Fang, Theory and applications of elliptically contoured and related distributions U.S. Army Research Office, Tech. Rep. 24, Sep. 1990.
- [23] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, Nov. 2012.
- [24] Y. I. Abramovich and O. Besson, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach—Part I: The oversampled case," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5807–5818, Dec. 2013.
- [25] O. Besson and Y. I. Abramovich, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach - Part 2: The under-sampled case," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5819–5829, Dec. 2013.
- [26] O. Besson and Y. Abramovich, "Invariance properties of the likelihood ratio for covariance matrix estimation in some complex elliptically contoured distributions," *J. Multivar. Anal.*, vol. 124, pp. 237–246, Feb. 2014.
- [27] Y. I. Abramovich and B. A. Johnson, "GLRT-based detection-estimation for undersampled training conditions," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3600–3612, Aug. 2008.
- [28] M. S. Srivastava and C. G. Khatri, *An Introduction to Multivariate Statistics.* New York, NY, USA: Elsevier/North Holland, 1979.
- [29] M. Siotani, T. Hayakawa, and Y. Fujikoto, *Modern Multivariate Statistical Analysis.* Cleveland, OH, USA: American Science, 1985.