Journal of Multivariate Analysis 124 (2014) 237–246

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Invariance properties of the likelihood ratio for covariance matrix estimation in some complex elliptically contoured distributions

Olivier Besson [a,*], Yuri I. Abramovich [b]

[a] University of Toulouse, ISAE, Department of Electronics, Optronics and Signal, 10 Avenue Edouard Belin, 31055 Toulouse, France

[b] W R Systems Ltd., 11351 Random Hills Road, Suite 400, Fairfax, VA 22030, USA

HIGHLIGHTS

- We consider a class of complex elliptically contoured matrix distributions (ECD).
- We investigate properties of the likelihood ratio (LR).
- We derive stochastic representations of the LR for covariance matrix estimation (CME).
- Its p.d.f. evaluated at the true CM $\boldsymbol{R}_0$ does not depend on the latter.
- This extends the expected likelihood approach for regularized CME.

ARTICLE INFO

Article history:
Received 26 December 2012
Available online 11 November 2013

AMS subject classifications:
62H10
62H15

Keywords:
Covariance matrix estimation
Elliptically contoured distribution
Expected likelihood
Likelihood ratio

ABSTRACT

The likelihood ratio (LR) for testing if the covariance matrix of the observation matrix $\boldsymbol{X}$ is $\boldsymbol{R}$ has some invariance properties that can be exploited for covariance matrix estimation purposes. More precisely, it was shown in Abramovich et al. (2004, 2007, 2007) that, in the Gaussian case, $LR(\boldsymbol{R}_0|\boldsymbol{X})$, where $\boldsymbol{R}_0$ stands for the true covariance matrix of the observations $\boldsymbol{X}$, has a distribution which does not depend on $\boldsymbol{R}_0$ but only on known parameters. This paved the way to the expected likelihood (EL) approach, which aims at assessing and possibly enhancing the quality of any covariance matrix estimate (CME) by comparing its LR to that of $\boldsymbol{R}_0$. Such invariance properties of $LR(\boldsymbol{R}_0|\boldsymbol{X})$ were recently proven for a class of elliptically contoured distributions (ECD) in Abramovich and Besson (2013) and Besson and Abramovich (2013) where regularized CME were also presented. The aim of this paper is to derive the distribution of $LR(\boldsymbol{R}_0|\boldsymbol{X})$ for other classes of ECD not covered yet, so as to make the EL approach feasible for a larger class of distributions.

1. Introduction and problem statement

The expected likelihood (EL) approach was introduced and developed in [3–5] as a statistical tool to assess the quality of a covariance matrix estimate $\hat{\boldsymbol{R}}$ from observation of a $M \times T$ matrix variate $\boldsymbol{X}$. The EL approach relies on some invariance properties of the likelihood ratio (LR) for testing $H_0 : \mathcal{E}\{\boldsymbol{X}\boldsymbol{X}^H\} = \boldsymbol{R}$ against the alternative $\mathcal{E}\{\boldsymbol{X}\boldsymbol{X}^H\} \neq \boldsymbol{R}$. More precisely, the LR is given by

$$LR(\boldsymbol{R}|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|\boldsymbol{R})}{\max_{\boldsymbol{R}} p(\boldsymbol{X}|\boldsymbol{R})} = \frac{p(\boldsymbol{X}|\boldsymbol{R})}{p(\boldsymbol{X}|\boldsymbol{R}_{\mathrm{ML}})}, \tag{1}$$

* Corresponding author.
E-mail addresses: olivier.besson@isae.fr (O. Besson), yabramovich@wrsystems.com (Y.I. Abramovich).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2013.10.024

where $p(\boldsymbol{X}|\boldsymbol{R})$ stands for the probability density function (p.d.f.) of the observations (which are assumed to be zero-mean) and $\boldsymbol{R}_{\mathrm{ML}}$ denotes the maximum likelihood estimator (MLE) of $\boldsymbol{R}$. As demonstrated in [3–5] for Gaussian distributed data, the p.d.f of $LR(\boldsymbol{R}_0|\boldsymbol{X})$, where $\boldsymbol{R}_0$ is the *true* covariance matrix of $\boldsymbol{X}$, does not depend on $\boldsymbol{R}_0$ but is fully determined by M and T. Moreover, the effective support of this p.d.f. lies on an interval whose values are much below $1 = LR(\boldsymbol{R}_{\mathrm{ML}}|\boldsymbol{X})$, see [3–5] for illustrative examples. In other words, the LR evaluated at the true covariance matrix is much lower than the LR evaluated at the MLE. This naturally raises the question of whether it would not make more sense that an estimate $\boldsymbol{R}(\hat{\boldsymbol{\beta}})$ of $\boldsymbol{R}_0$, where $\boldsymbol{R}(\boldsymbol{\beta})$ is either a parameterized model for the covariance matrix or a regularized estimate (e.g., shrinkage of the MLE to some target covariance matrix), results in a LR which is commensurate with that of $\boldsymbol{R}_0$. This is the gist of the EL approach which estimates $\boldsymbol{\beta}$ by enforcing that $LR(\boldsymbol{R}(\hat{\boldsymbol{\beta}})|\boldsymbol{X})$ takes values which are compatible with the support of the p.d.f. of $LR(\boldsymbol{R}_0|\boldsymbol{X})$. To be more specific, let us consider a classical regularized covariance matrix estimate (CME) based on shrinkage of the MLE to a target matrix $\boldsymbol{R}_t$, i.e.,

$$\boldsymbol{R}(\beta) = (1-\beta)\boldsymbol{R}_{\mathrm{ML}} + \beta \boldsymbol{R}_t.$$

The EL approach for selection of the shrinkage factor β could possibly take the following form [4,1]:

$$\beta_{\mathrm{EL}} = \arg\min_{\beta} \left| LR^{1/T}(\boldsymbol{R}(\beta)|\boldsymbol{X}) - \mathrm{med}\left[\omega(LR|M,T)\right]\right|,$$

where $\omega(LR|M,T)$ is the true p.d.f. of $LR^{1/T}(\boldsymbol{R}_0|\boldsymbol{X})$ and $\mathrm{med}\left[\omega(LR|M,T)\right]$ stands for the median value. In other words, the shrinkage factor is chosen such that the resulting LR of $\boldsymbol{R}(\beta_{\mathrm{EL}})$ is comparable with that of $\boldsymbol{R}_0$. It is well known that regularization is particularly effective in low sample support and the EL principle was shown in [4,1] to provide a quite efficient mechanism to tune the regularization parameters. Various uses of the EL approach are possible and their effectiveness has been illustrated in different applications. For instance, it has been used successfully to detect severely erroneous MUSIC-based DOA estimates in low signal to noise ratio and it provided a mechanism to rectify the set of these estimates to meet the expected likelihood ratio values [3,5]. Accordingly, the EL approach was proven to be instrumental in designing efficient adaptive detectors in low sample support [4].

In [1,8] we extended the EL approach to a class of complex elliptically contoured distributions (ECD) (namely the $\mathcal{EMS}_{M,T}(\boldsymbol{0},\boldsymbol{R},\phi)$ type of distributions, as referred to in this paper) and we provided regularization schemes for covariance matrix estimation. Regularized covariance matrix estimation has been studied extensively in the literature, see e.g. [18,12,19,20,27] for a few examples within the framework of elliptically contoured distributions. In the latter references, the regularization parameters are selected with a view to minimize either the mean-square error or Stein loss. Our goal in this paper is not to derive and compare new covariance estimation schemes, as in [1]. Rather we focus herein in deriving invariance properties of the LR for other classes of complex elliptically contoured distributions, so as to extend the class of distributions for which the EL approach of covariance matrix estimation is feasible. How the EL approach will be used in this framework is beyond the scope of the present paper. The starting point of the present study is the following. While there is a general agreement and usually no ambiguity for defining vector elliptically contoured distributions, when it comes to extending ECD to matrix-variate, a certain number of options are possible [11]. Indeed, Fang and Zhang distinguish four classes of matrix-variate ECD whose p.d.f. and stochastic representations are different. As we shall see shortly, considering as in [1] the columns of $\boldsymbol{X}$ as independent and identically distributed (i.i.d.) elliptically distributed random vectors (r.v.) results in $\boldsymbol{X} \sim \mathcal{EMS}_{M,T}(\boldsymbol{0},\boldsymbol{R},\phi)$ (obtained from a multivariate spherical distribution in the terminology of [11]). On the other hand, the ECD considered, e.g., in [23,24] are obtained assuming that $\mathrm{vec}(\boldsymbol{X}) \in \mathbb{C}^{MT\times 1}$ follows a vector ECD, which we will denote as $\boldsymbol{X} \sim \mathcal{EVS}_{M,T}(\boldsymbol{0},\boldsymbol{R},\phi)$.

In this paper we shall examine the p.d.f. of the likelihood ratio for two classes of complex ECD not covered in [1], namely $\boldsymbol{X} \sim \mathcal{ESS}_{M,T}(\boldsymbol{0},\boldsymbol{R},\phi)$ and $\boldsymbol{X} \sim \mathcal{EVS}_{M,T}(\boldsymbol{0},\boldsymbol{R},\phi)$. For the former, we will pay special attention to the matrix-variate Student distribution. The latter category was considered in [23,24] where Richmond proved the quite remarkable result that Kelly's generalized likelihood ratio test (GLRT) for Gaussian distributed data [17] was also the GLRT for this class of ECD. A main result of this paper includes stochastic representations of the likelihood ratio (and proof of invariance) in both the over-sampled case ($T \geq M$) and the under-sampled scenario where the number of available samples in less than the size of the observation space ($T \leq M$), in which case regularization is mandatory. Note that invariance properties of some likelihood ratios for elliptically contoured distributions (mostly EVS) have been studied, e.g., in [16,15,10,7,6], but the likelihood ratios are somewhat different from what we consider here and they serve different purposes.

2. A brief review of elliptically contoured distributions

In this section, we provide a brief summary of ECD with the only purpose of providing sufficient background for derivation and analysis of the LR in the next sections. We refer the reader to [11,6] for details that are skipped here and for an exhaustive analysis: our presentation here will follow the terminology of [11]. We also point to the recent paper [22] for an excellent comprehensive overview and applications to array processing. A vector $\boldsymbol{x} \in \mathbb{C}^M$ is said to be spherically distributed if its characteristic function $\mathcal{E}\{e^{i\mathrm{Re}\{\boldsymbol{t}^H\boldsymbol{x}\}}\} = \phi(\boldsymbol{t}^H\boldsymbol{t})$: we will denote it as $\boldsymbol{x} \sim \mathcal{S}_M(\phi)$. Assuming that $\boldsymbol{x}$ has a density (which we will do through this document) the latter only depends on $\boldsymbol{x}^H\boldsymbol{x}$. A vector $\boldsymbol{x} \in \mathbb{C}^M$ is said to follow an elliptically contoured distribution if

$$\mathcal{E}\{e^{i\mathrm{Re}\{\boldsymbol{t}^H\boldsymbol{x}\}}\} = \mathcal{E}\{e^{i\mathrm{Re}\{\boldsymbol{t}^H\boldsymbol{m}\}}\}\phi(\boldsymbol{t}^H\boldsymbol{R}\boldsymbol{t}). \tag{2}$$

Such a vector admits the following stochastic representation:

$$\boldsymbol{x} \stackrel{d}{=} \boldsymbol{m} + \mathcal{R}\boldsymbol{B}\boldsymbol{u} \tag{3}$$

where $\stackrel{d}{=}$ means "has the same distribution as". In (3), $\mathcal{R} = \sqrt{\mathcal{Q}}$ is a non-negative real random variable, called modular variate, and is independent of the complex random vector $\boldsymbol{u}$ which is uniformly distributed on the unit complex sphere in $\mathbb{C}\mathcal{S}^R$ denoted as $\boldsymbol{u} \sim \mathcal{U}_R$. $\boldsymbol{B}$ is a $M \times R$ matrix such that $\boldsymbol{B}\boldsymbol{B}^H = \boldsymbol{R}$. The latter is usually referred to as the scatter matrix: with some abuse of language we will refer to it as the covariance matrix, keeping in mind that the true covariance matrix of $\boldsymbol{x}$ is indeed $M^{-1}\mathcal{E}\{\mathcal{R}^2\}\boldsymbol{R}$ under mild assumptions. In the sequel, we will consider the absolutely continuous case for which $\boldsymbol{R}$ is non-singular and hence $R = M$. In such a case, the p.d.f. of $\boldsymbol{x}$ exists and can be written as

$$p(\boldsymbol{x}|\boldsymbol{R}) = |\boldsymbol{R}|^{-1} g\left(\boldsymbol{x}^H \boldsymbol{R}^{-1} \boldsymbol{x}\right) \tag{4}$$

where $g : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is called the density generator and is related to the p.d.f. of $\mathcal{Q}$ [11,22]. We will denote $\boldsymbol{x} \sim \mathcal{E}\mathcal{C}_M(\boldsymbol{\mu}, \boldsymbol{R}, g)$ or $\boldsymbol{x} \sim \mathcal{E}\mathcal{C}_M(\boldsymbol{\mu}, \boldsymbol{R}, \phi)$. Let us now consider the definitions of ECD for a matrix-variate $\boldsymbol{X} \in \mathbb{C}^{M\times T}$: typically, the columns of $\boldsymbol{X}$ correspond to the T observed snapshots from the output of an array with M elements. In order to define ECD, we need as a pre-requisite to consider spherical matrix distributions. We simply give below their definitions along with some properties and refer the reader to [11] for further details and proofs. A matrix $\boldsymbol{Y} \in \mathbb{C}^{M\times T}$ is said to be left-spherical, which we denote as $\boldsymbol{Y} \sim \mathcal{L}\mathcal{S}_{M,T}(\phi)$, if $\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\Gamma}\boldsymbol{Y}$ for any $\boldsymbol{\Gamma} \in O(M)$ where $O(M) = \left\{\boldsymbol{\Gamma} \in \mathbb{C}^{M\times M} \text{ such that } \boldsymbol{\Gamma}^H\boldsymbol{\Gamma} = \boldsymbol{I}_M\right\}$. We have the following properties:

$$\boldsymbol{Y} \sim \mathcal{L}\mathcal{S}_{M,T}(\phi) \Rightarrow \psi_{\boldsymbol{Y}}(\boldsymbol{T}) = \mathcal{E}\{\exp\{i\mathrm{Re}\{\boldsymbol{T}^H\boldsymbol{Y}\}\}\} = \phi(\boldsymbol{T}^H\boldsymbol{T}) \tag{5a}$$

$$\boldsymbol{Y} \sim \mathcal{L}\mathcal{S}_{M,T}(\phi) \Rightarrow \boldsymbol{Y} \stackrel{d}{=} \boldsymbol{U}\boldsymbol{A} \tag{5b}$$

$$\boldsymbol{Y} \sim \mathcal{L}\mathcal{S}_{M,T}(\phi) \Rightarrow p(\boldsymbol{Y}) = g(\boldsymbol{Y}^H\boldsymbol{Y}) \tag{5c}$$

where, in the last line, we have assumed that the p.d.f. of $\boldsymbol{Y}$ exists. In the above equation, $\boldsymbol{U} \sim \mathcal{U}_{M,T}$ is said to be uniformly distributed, i.e., $\boldsymbol{U}$ is left-spherical and $\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}_T$, and $\boldsymbol{A}$ is some random matrix independent of $\boldsymbol{U}$. A matrix $\boldsymbol{Y}$ is said to be spherical if both $\boldsymbol{Y}$ and $\boldsymbol{Y}^H$ are left-spherical: we denote $\boldsymbol{Y} \sim \mathcal{S}\mathcal{S}_{M,T}(\phi)$. This amounts to saying that $\boldsymbol{Y} \stackrel{d}{=} \boldsymbol{\Gamma}\boldsymbol{Y}\boldsymbol{\Omega}$ for any $\boldsymbol{\Gamma} \in O(M)$ and $\boldsymbol{\Omega} \in O(T)$. We have the following properties:

$$\boldsymbol{Y} \sim \mathcal{S}\mathcal{S}_{M,T}(\phi) \Rightarrow \psi_{\boldsymbol{Y}}(\boldsymbol{T}) = \phi(\lambda(\boldsymbol{T}^H\boldsymbol{T})) \tag{6a}$$

$$\boldsymbol{Y} \sim \mathcal{S}\mathcal{S}_{M,T}(\phi) \Rightarrow \boldsymbol{Y} \stackrel{d}{=} \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^H \tag{6b}$$

$$\boldsymbol{Y} \sim \mathcal{S}\mathcal{S}_{M,T}(\phi) \Rightarrow p(\boldsymbol{Y}) = g(\lambda(\boldsymbol{Y}^H\boldsymbol{Y})). \tag{6c}$$

In the above $\lambda(.)$ stands for the diagonal matrix of the (non-zero) eigenvalues of the matrix between parentheses, $\boldsymbol{U}$, $\boldsymbol{\Lambda}$ and $\boldsymbol{V}$ are independent. The third class of matrix spherical distributions is the so-called multivariate spherical distributions. $\boldsymbol{Y} = \begin{bmatrix}\boldsymbol{y}_1 & \cdots & \boldsymbol{y}_T\end{bmatrix} \sim \mathcal{M}\mathcal{S}_{M,T}(\phi)$ if $\boldsymbol{Y} \stackrel{d}{=} \begin{bmatrix}\boldsymbol{\Gamma}_1\boldsymbol{y}_1 & \cdots & \boldsymbol{\Gamma}_T\boldsymbol{y}_T\end{bmatrix}$ for every $\boldsymbol{\Gamma}_t \in O(M)$, $t = 1, \ldots, T$. One has the following properties:

$$\boldsymbol{Y} \sim \mathcal{M}\mathcal{S}_{M,T}(\phi) \Rightarrow \psi_{\boldsymbol{Y}}(\boldsymbol{T}) = \phi(\boldsymbol{t}_1^H\boldsymbol{t}_1, \ldots, \boldsymbol{t}_T^H\boldsymbol{t}_T) \tag{7a}$$

$$\boldsymbol{Y} \sim \mathcal{M}\mathcal{S}_{M,T}(\phi) \Rightarrow \boldsymbol{Y} \stackrel{d}{=} \begin{bmatrix}\mathcal{R}_1\boldsymbol{u}_1 & \cdots & \mathcal{R}_T\boldsymbol{u}_T\end{bmatrix} = \boldsymbol{U}_2\mathrm{diag}(\mathcal{R}_1, \ldots, \mathcal{R}_T) \tag{7b}$$

$$\boldsymbol{Y} \sim \mathcal{M}\mathcal{S}_{M,T}(\phi) \Rightarrow p(\boldsymbol{Y}) = g(\boldsymbol{y}_1^H\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T^H\boldsymbol{y}_T). \tag{7c}$$

In (7), $\boldsymbol{T} = \begin{bmatrix}\boldsymbol{t}_1 & \cdots & \boldsymbol{t}_T\end{bmatrix}$, $\boldsymbol{u}_t$ are i.i.d. r.v. uniformly distributed on the complex unit sphere $\mathbb{C}\mathcal{S}^M$, i.e., $\boldsymbol{u}_t \sim\sim \mathcal{U}_M$, and $\mathcal{R}_t \geq 0$ are i.i.d. and independent of $\boldsymbol{U}_2$.

Finally, one can define vector-spherical distributions: $\boldsymbol{Y} \sim \mathcal{V}\mathcal{S}_{M,T}(\phi)$ if $\mathrm{vec}(\boldsymbol{Y}) \sim \mathcal{S}_{MT}(\phi)$ which yields

$$\boldsymbol{Y} \sim \mathcal{V}\mathcal{S}_{M,T}(\phi) \Rightarrow \psi_{\boldsymbol{Y}}(\boldsymbol{T}) = \phi(\mathrm{Tr}\{\boldsymbol{T}^H\boldsymbol{T}\}) \tag{8a}$$

$$\boldsymbol{Y} \sim \mathcal{V}\mathcal{S}_{M,T}(\phi) \Rightarrow \boldsymbol{Y} \stackrel{d}{=} \mathcal{R}\boldsymbol{U}_3 \tag{8b}$$

$$\boldsymbol{Y} \sim \mathcal{V}\mathcal{S}_{M,T}(\phi) \Rightarrow p(\boldsymbol{Y}) = g(\mathrm{Tr}\{\boldsymbol{T}^H\boldsymbol{T}\}) \tag{8c}$$

with $\mathrm{vec}(\boldsymbol{U}_3) \sim \mathcal{U}_{MT}$ and with $\mathcal{R} \geq 0$ independent of $\boldsymbol{U}_3$.

Elliptically contoured distributions essentially follow from the transformation

$$\boldsymbol{X} = \boldsymbol{M} + \boldsymbol{B}\boldsymbol{Y} \tag{9}$$

where $\boldsymbol{Y}$ follows a spherical distribution and $\boldsymbol{B}\boldsymbol{B}^H = \boldsymbol{R}$. In the sequel, similarly to [11], we will use the following notation: $\boldsymbol{Y} \sim \mathcal{L}\mathcal{S}_{M,T}(\phi) \Rightarrow \boldsymbol{X} \sim \mathcal{E}\mathcal{L}\mathcal{S}_{M,T}(\boldsymbol{M}, \boldsymbol{R}, \phi)$, $\boldsymbol{Y} \sim \mathcal{S}\mathcal{S}_{M,T}(\phi) \Rightarrow \boldsymbol{X} \sim \mathcal{E}\mathcal{S}\mathcal{S}_{M,T}(\boldsymbol{M}, \boldsymbol{R}, \phi)$, $\boldsymbol{Y} \sim \mathcal{M}\mathcal{S}_{M,T}(\phi) \Rightarrow \boldsymbol{X} \sim \mathcal{E}\mathcal{M}\mathcal{S}_{M,T}(\boldsymbol{M}, \boldsymbol{R}, \phi)$ and $\boldsymbol{Y} \sim \mathcal{V}\mathcal{S}_{M,T}(\phi) \Rightarrow \boldsymbol{X} \sim \mathcal{E}\mathcal{V}\mathcal{S}_{M,T}(\boldsymbol{M}, \boldsymbol{R}, \phi)$. We will use indifferently ϕ or g in the notation since we assume that the p.d.f. exists. The stochastic representations of $\boldsymbol{X}$ follow immediately from those of $\boldsymbol{Y}$ and the corresponding p.d.f can be written as

$$p(\boldsymbol{X}|\boldsymbol{R},g) = |\boldsymbol{R}|^{-T} g\left((\boldsymbol{X}-\boldsymbol{M})^H\boldsymbol{R}^{-1}(\boldsymbol{X}-\boldsymbol{M})\right) \tag{ELS}$$

$$p(\boldsymbol{X}|\boldsymbol{R},g) = |\boldsymbol{R}|^{-T} g\left(\lambda((\boldsymbol{X}-\boldsymbol{M})^H\boldsymbol{R}^{-1}(\boldsymbol{X}-\boldsymbol{M}))\right) \tag{ESS}$$

$$p(\boldsymbol{X}|\boldsymbol{R},g) = |\boldsymbol{R}|^{-T} g((\boldsymbol{y}_1-\boldsymbol{m}_1)^H\boldsymbol{R}^{-1}(\boldsymbol{y}_1-\boldsymbol{m}_1),\ldots,(\boldsymbol{y}_T-\boldsymbol{m}_T)^H\boldsymbol{R}^{-1}(\boldsymbol{y}_T)-\boldsymbol{m}_T) \tag{EMS}$$

$$p(\boldsymbol{X}|\boldsymbol{R},g) = |\boldsymbol{R}|^{-T} g\left(\mathrm{Tr}\{(\boldsymbol{X}-\boldsymbol{M})^H\boldsymbol{R}^{-1}(\boldsymbol{X}-\boldsymbol{M})\}\right). \tag{EVS}$$

In [1] we considered in fact the EL approach for the EMS class of distributions. In the present paper, we extend this approach to ESS distributions (of which the usual multivariate Student distribution $p(\boldsymbol{X}|\boldsymbol{R}) \propto |\boldsymbol{R}|^{-T}|\boldsymbol{I}_T + (\boldsymbol{X}-\boldsymbol{M})^H\boldsymbol{R}^{-1}(\boldsymbol{X}-\boldsymbol{M})|^{-(d+M)}$ is a member) and to the EVS distributions which have been considered e.g., in [23,24].

3. Likelihood ratio for $\mathcal{EVS}_{M,T}(\mathbf{0},\boldsymbol{R},g)$ distributions

In this section we assume that $\boldsymbol{X}$ follows a vector-spherical distribution with zero-mean, i.e., $\boldsymbol{X} \sim \mathcal{EVS}_{M,T}(\mathbf{0},\boldsymbol{R},g)$ and our goal is to estimate $\boldsymbol{R}$ using the expected likelihood approach. Towards this goal, the first step is to derive the likelihood ratio $L(\boldsymbol{R}(\boldsymbol{\beta})|\boldsymbol{X})$ for any (possibly parameterized) covariance matrix, and to show that $L(\boldsymbol{R}_0|\boldsymbol{X})$ does not depend on $\boldsymbol{R}_0$. Similarly to [11,23,24] we will assume that $g(.)$ is a continuous and non-increasing function, which guarantees that $t^{MT}g(Mt)$ achieves a maximum at $t_{\max}$.

3.1. The over-sampled case

Let us consider first that the number of measurements T exceeds the number of elements in the array M, i.e., $T \geq M$. Under such hypothesis, the maximum likelihood estimate (MLE) of $\boldsymbol{R}$ is given by [23,24]

$$\boldsymbol{R}_{\mathrm{ML}} = \frac{\boldsymbol{X}\boldsymbol{X}^H}{t_{\max}}. \tag{11}$$

It thus follows that the LR can be written as

$$\begin{aligned} LR(\boldsymbol{R}(\boldsymbol{\beta})|\boldsymbol{X}) &= \frac{p(\boldsymbol{X}|\boldsymbol{R}(\boldsymbol{\beta}))}{p(\boldsymbol{X}|\boldsymbol{R}_{\mathrm{ML}})} \\ &= \frac{|\boldsymbol{R}(\boldsymbol{\beta})|^{-T} g\left(\mathrm{Tr}\{\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}\}\right)}{|\boldsymbol{R}_{\mathrm{ML}}|^{-T} g\left(\mathrm{Tr}\{\boldsymbol{X}^H\boldsymbol{R}_{\mathrm{ML}}^{-1}\boldsymbol{X}\}\right)} \\ &= |\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{R}_{\mathrm{ML}}|^T \frac{g\left(\mathrm{Tr}\{\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}\}\right)}{g\left(\mathrm{Tr}\{t_{\max}\boldsymbol{I}_M\}\right)} \\ &= C|\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}\boldsymbol{X}^H|^T g\left(\mathrm{Tr}\{\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}\}\right) \end{aligned} \tag{12}$$

where $C = t_{\max}^{-MT}/g(Mt_{\max}) = \left[\max_t t^{MT}g(Mt)\right]^{-1}$. Now using the stochastic representation in (8b), we have $\boldsymbol{X} \overset{d}{=} \mathcal{R}\boldsymbol{R}_0^{-1/2}\boldsymbol{U}_3$ and thus

$$\begin{aligned} LR(\boldsymbol{R}_0|\boldsymbol{X}) &= C|\boldsymbol{R}_0^{-1}\boldsymbol{X}\boldsymbol{X}^H|^T g\left(\mathrm{Tr}\{\boldsymbol{X}^H\boldsymbol{R}_0^{-1}\boldsymbol{X}\}\right) \\ &\overset{d}{=} C|\mathcal{R}^2\boldsymbol{U}_3\boldsymbol{U}_3^H|^T g\left(\mathrm{Tr}\{\mathcal{R}^2\boldsymbol{U}_3\boldsymbol{U}_3^H\}\right) \\ &= C|\mathcal{R}^2\boldsymbol{U}_3\boldsymbol{U}_3^H|^T g\left(\mathcal{R}^2\right), \end{aligned} \tag{13}$$

where we used the fact that $\mathrm{Tr}\{\boldsymbol{U}_3^H\boldsymbol{U}_3\} = \mathrm{vec}(\boldsymbol{U}_3)^H\mathrm{vec}(\boldsymbol{U}_3) = 1$ since $\mathrm{vec}(\boldsymbol{U}_3) \sim \mathcal{U}_{MT}$. The latter property implies that

$$\mathrm{vec}(\boldsymbol{U}_3) \overset{d}{=} \frac{\tilde{\boldsymbol{n}}}{\|\tilde{\boldsymbol{n}}\|} \quad \text{with } \tilde{\boldsymbol{n}} \sim \mathcal{N}(\mathbf{0},\boldsymbol{I}_{MT}). \tag{14}$$

Let $\boldsymbol{N}$ be the $M\times T$ matrix such that $\mathrm{vec}(\boldsymbol{N}) = \tilde{\boldsymbol{n}}$: we thus have $\boldsymbol{U}_3 \overset{d}{=} \frac{\boldsymbol{N}}{\mathrm{Tr}\{\boldsymbol{N}\boldsymbol{N}^H\}^{1/2}}$ and therefore

$$|\boldsymbol{U}_3\boldsymbol{U}_3^H| \overset{d}{=} \frac{|\boldsymbol{N}\boldsymbol{N}^H|}{\mathrm{Tr}\{\boldsymbol{N}\boldsymbol{N}^H\}^M}. \tag{15}$$

Now observe that the matrix $\boldsymbol{W} = \boldsymbol{N}\boldsymbol{N}^H$ has a complex Wishart distribution with T degrees of freedom, i.e., $\boldsymbol{W} \sim \mathcal{W}(T,\boldsymbol{I}_M)$ [13,21]. This matrix can thus be decomposed as $\boldsymbol{W} = \boldsymbol{C}^H\boldsymbol{C}$ where $\boldsymbol{C}$ is an $M\times M$ upper-triangular matrix. It is well known [13,21] that all random variables $\boldsymbol{C}_{ij}$ are independent. Moreover, $|\boldsymbol{C}_{ii}|^2 \sim \chi^2_{T-i+1}$ where χ^2_n stands for the complex central chi-square distribution with n degrees of freedom, whose p.d.f. is given by $f_{\chi^2_n}(x) = \Gamma^{-1}(n)x^{n-1}\exp\{-x\}$. Additionally, one has $\boldsymbol{C}_{ij} \sim \mathcal{N}(0, 1)$ for $i \neq j$. It then ensues that

$$\frac{|\boldsymbol{W}|}{\mathrm{Tr}\{\boldsymbol{W}\}^M} = \frac{\prod_{m=1}^{M} |\boldsymbol{C}_{mm}|^2}{\left[\sum_{m=1}^{M} |\boldsymbol{C}_{mm}|^2 + \sum_{i\neq j} |\boldsymbol{C}_{ij}|^2\right]^M} \stackrel{d}{=} \frac{\prod_{m=1}^{M} \chi^2_{T-m+1}}{\left[\chi^2_{M(M-1)/2} + \sum_{m=1}^{M} \chi^2_{T-m+1}\right]^M}. \tag{16}$$

Finally, we obtain the following stochastic representation:

$$LR(\boldsymbol{R}_0|\boldsymbol{X}) \stackrel{d}{=} C\mathcal{R}^{2MT} g\left(\mathcal{R}^2\right) \frac{\left(\prod_{m=1}^{M} \chi^2_{T-m+1}\right)^T}{\left[\chi^2_{M(M-1)/2} + \sum_{m=1}^{M} \chi^2_{T-m+1}\right]^{MT}} \tag{17}$$

whose p.d.f. obviously does not depend on $\boldsymbol{R}_0$. However the p.d.f. of $LR(\boldsymbol{R}_0|\boldsymbol{X})$ depends on $g(.)$. For a given density generator, the p.d.f. of $LR(\boldsymbol{R}_0|\boldsymbol{X})$ can be easily simulated which provides a target value (for instance the median value of $LR(\boldsymbol{R}_0|\boldsymbol{X})$ [1]) for the LR of any regularized covariance matrix estimate, and hence a way to select the regularization parameters.

3.2. The under-sampled case

Let us address now the case $T \leq M$ which requires a specific analysis since the MLE of $\boldsymbol{R}$ does not exist any longer. In fact, with $T < M$ training samples, information about the covariance matrix can be retrieved only in the T-dimensional subspace spanned by the columns of the data matrix $\boldsymbol{X}$ [2]. Let $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^H$ be the sample covariance matrix and let $\boldsymbol{S} = \hat{\boldsymbol{U}}_T\hat{\boldsymbol{\Lambda}}_T\hat{\boldsymbol{U}}_T^H$ denote its eigenvalue decomposition where $\hat{\boldsymbol{U}}_T$ is a $M \times T$ matrix of orthonormal eigenvectors i.e., $\hat{\boldsymbol{U}}_T^H\hat{\boldsymbol{U}}_T = \boldsymbol{I}_T$. Inference about any covariance matrix $\boldsymbol{R}(\boldsymbol{\beta})$ is thus possible only in the range space of $\hat{\boldsymbol{U}}_T$. Therefore, for any given $\boldsymbol{R}(\boldsymbol{\beta})$, we need to find the rank-T Hermitian matrix $\boldsymbol{D}_T$, such that the construct $\boldsymbol{R}_T = \hat{\boldsymbol{U}}_T\boldsymbol{D}_T\hat{\boldsymbol{U}}_T^H$ is "closest" to the model $\boldsymbol{R}(\boldsymbol{\beta})$. In [2] it was proven that $\boldsymbol{D}_T$ is given by

$$\boldsymbol{D}_T = \left[\hat{\boldsymbol{U}}_T^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\hat{\boldsymbol{U}}_T\right]^{-1} \tag{18}$$

and hence

$$\boldsymbol{R}_T = \hat{\boldsymbol{U}}_T\left[\hat{\boldsymbol{U}}_T^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\hat{\boldsymbol{U}}_T\right]^{-1}\hat{\boldsymbol{U}}_T^H. \tag{19}$$

$\boldsymbol{R}_T$, which is rank-deficient, can be considered as the singular covariance matrix of an EVS distribution.

At this stage, we need to consider EVS distributions with singular $\boldsymbol{R}$ and to define singular densities in a particular subspace [23,9], in the same way that singular Gaussian distributions can be defined, see e.g., [26,25,14]. Let us thus temporarily consider an $\mathcal{EVS}_{M,T}(\boldsymbol{0}, \boldsymbol{R}, g)$ distribution with a rank-deficient $\boldsymbol{R}$ that can be decomposed as $\boldsymbol{R} = \boldsymbol{U}_R\boldsymbol{D}_R\boldsymbol{U}_R^H$ where $\boldsymbol{U}_R$ is an $M \times R$ matrix whose orthonormal columns form a basis for the range space of $\boldsymbol{R}$, and $\boldsymbol{D}_R$ is a positive definite $R \times R$ Hermitian matrix. Then [11,22], we have $\boldsymbol{X} \stackrel{d}{=} \mathcal{R}\boldsymbol{U}_R\boldsymbol{D}_R^{1/2}\boldsymbol{Y}_R$ with $\mathrm{vec}(\boldsymbol{Y}_R) \sim \mathcal{U}_{RT}$. Although a density cannot be defined on the set of $M \times T$ matrices, one can define a density on the set of $M \times T$ matrices $\boldsymbol{X}$ such that $\left(\boldsymbol{U}_R^{\perp}\right)^H\boldsymbol{X} = \boldsymbol{0}$ [23,9], where $\boldsymbol{U}_R^{\perp}$ is an orthonormal basis for the complement of $\boldsymbol{U}_R$, i.e., $\left(\boldsymbol{U}_R^{\perp}\right)^H\boldsymbol{U}_R^{\perp} = \boldsymbol{I}_{M-R}$ and $\left(\boldsymbol{U}_R^{\perp}\right)^H\boldsymbol{U}_R = \boldsymbol{0}$. Let $\boldsymbol{U} = \begin{bmatrix}\boldsymbol{U}_R & \boldsymbol{U}_R^{\perp}\end{bmatrix}$ and let us make the change of variables

$$\tilde{\boldsymbol{X}} = \boldsymbol{U}^H\boldsymbol{X} = \begin{bmatrix}\boldsymbol{U}_R^H\boldsymbol{X} \\ \left(\boldsymbol{U}_R^{\perp}\right)^H\boldsymbol{X}\end{bmatrix} = \begin{bmatrix}\tilde{\boldsymbol{X}}_R \\ \boldsymbol{0}\end{bmatrix} \tag{20}$$

where $\tilde{\boldsymbol{X}}_R \stackrel{d}{=} \mathcal{R}\boldsymbol{U}_R\boldsymbol{D}_R^{1/2}\boldsymbol{Y}_R \sim \mathcal{EVS}_{R,T}(\boldsymbol{0}, \boldsymbol{D}_R, g_R)$. The p.d.f. of $\tilde{\boldsymbol{X}}_R$ exists and is given by

$$\begin{aligned} p(\tilde{\boldsymbol{X}}_R|\boldsymbol{U}_R, \boldsymbol{D}_R) &= |\boldsymbol{D}_R|^{-T}g_R\left(\mathrm{Tr}\{\tilde{\boldsymbol{X}}_R^H\boldsymbol{D}_R^{-1}\tilde{\boldsymbol{X}}_R\}\right) \\ &= |\boldsymbol{D}_R|^{-T}g_R\left(\mathrm{Tr}\{\boldsymbol{X}^H\boldsymbol{U}_R\boldsymbol{D}_R^{-1}\boldsymbol{U}_R^H\boldsymbol{X}\}\right) \\ &= |\boldsymbol{D}_R|^{-T}g_R\left(\mathrm{Tr}\{\boldsymbol{X}^H\boldsymbol{R}^{-}\boldsymbol{X}\}\right) \end{aligned} \tag{21}$$

where $\boldsymbol{R}^{-}$ is the pseudo-inverse of $\boldsymbol{R}$ and the notation $g_R(.)$ emphasizes that the density is well-defined as a function over $\mathbb{C}^R$ [23]. Since the Jacobian from $\boldsymbol{X}$ to $\tilde{\boldsymbol{X}}$ is 1, one can define a density on the set $\left\{\boldsymbol{X} \in \mathbb{C}^{M\times T} / \left(\boldsymbol{U}_R^{\perp}\right)^H \boldsymbol{X} = \boldsymbol{0}\right\}$ as [23]

$$p(\boldsymbol{X}|\boldsymbol{U}_R, \boldsymbol{D}_R) = |\boldsymbol{D}_R|^{-T} g_R\left(\mathrm{Tr}\{\boldsymbol{X}^H \boldsymbol{U}_R \boldsymbol{D}_R^{-1} \boldsymbol{U}_R^H \boldsymbol{X}\}\right). \tag{22}$$

Now, back to our specific application with $\boldsymbol{R}_T$ in (19) being an admissible singular covariance matrix, we get

$$\begin{aligned} p(\boldsymbol{X}|\boldsymbol{R}(\boldsymbol{\beta})) &= |\boldsymbol{D}_T|^{-T} g_T\left(\mathrm{Tr}\{\boldsymbol{X}^H \hat{\boldsymbol{U}}_T \boldsymbol{D}_T^{-1} \hat{\boldsymbol{U}}_T^H \boldsymbol{X}\}\right) \\ &= |\hat{\boldsymbol{U}}_T^H \boldsymbol{R}^{-1}(\boldsymbol{\beta}) \hat{\boldsymbol{U}}_T|^T g_T\left(\mathrm{Tr}\{\boldsymbol{X}^H \hat{\boldsymbol{U}}_T \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{R}^{-1}(\boldsymbol{\beta}) \hat{\boldsymbol{U}}_T\right] \hat{\boldsymbol{U}}_T^H \boldsymbol{X}\}\right). \end{aligned} \tag{23}$$

The MLE of $\boldsymbol{D}_T$ is given by

$$\boldsymbol{D}_T^{ML} = \frac{\hat{\boldsymbol{U}}_T^H \boldsymbol{X}\boldsymbol{X}^H \hat{\boldsymbol{U}}_T}{\tilde{t}_{\max}} = \frac{\hat{\boldsymbol{\Lambda}}_T}{\tilde{t}_{\max}} \tag{24}$$

where $\tilde{t}_{\max} = \arg\max_t t^{T^2} g_T(tT)$. It then follows that the LR has the following expression:

$$\begin{aligned} LR^u(\boldsymbol{R}(\boldsymbol{\beta})|\boldsymbol{X}) &= |\hat{\boldsymbol{U}}_T^H \boldsymbol{R}^{-1}(\boldsymbol{\beta}) \hat{\boldsymbol{U}}_T \boldsymbol{D}_T^{ML}|^T \frac{g_T\left(\mathrm{Tr}\{\boldsymbol{X}^H \boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}\}\right)}{g_T\left(\mathrm{Tr}\{\tilde{t}_{\max}\boldsymbol{I}_T\}\right)} \\ &= \tilde{C}|\boldsymbol{X}^H \boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}|^T g_T\left(\mathrm{Tr}\{\boldsymbol{X}^H \boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}\}\right) \end{aligned} \tag{25}$$

with $\tilde{C} = \tilde{t}_{\max}^{-T^2}/g(T\tilde{t}_{\max}) = \left[\max_t t^{T^2} g_T(tT)\right]^{-1}$. Using the fact that $\boldsymbol{X} \overset{d}{=} \mathcal{R}\boldsymbol{R}_0^{-1/2}\boldsymbol{U}_3$, the under-sampled likelihood ratio, evaluated at the true covariance matrix, can thus be written as

$$\begin{aligned} LR^u(\boldsymbol{R}_0|\boldsymbol{X}) &= \tilde{C}|\boldsymbol{X}^H \boldsymbol{R}_0^{-1}\boldsymbol{X}|^T g_T\left(\mathrm{Tr}\{\boldsymbol{X}^H \boldsymbol{R}_0^{-1}\boldsymbol{X}\}\right) \\ &\overset{d}{=} \tilde{C}|\mathcal{R}^2 \boldsymbol{U}_3^H \boldsymbol{U}_3|^T g_T\left(\mathrm{Tr}\{\mathcal{R}^2 \boldsymbol{U}_3^H \boldsymbol{U}_3\}\right) \\ &= \tilde{C}\frac{|\mathcal{R}^2 \boldsymbol{N}^H \boldsymbol{N}|^T}{\mathrm{Tr}\{\boldsymbol{N}^H \boldsymbol{N}\}^{T^2}} g_T\left(\mathcal{R}^2\right), \end{aligned} \tag{26}$$

where $\boldsymbol{N}$ is an $M \times T$ matrix such that $\mathrm{vec}(\boldsymbol{N}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{MT})$. Consequently $\boldsymbol{W} = \boldsymbol{N}^H\boldsymbol{N} \sim \mathcal{W}(M, \boldsymbol{I}_T)$ and $\boldsymbol{W}$ can be decomposed as $\boldsymbol{W} = \boldsymbol{C}^H\boldsymbol{C}$ where $\boldsymbol{C}$ is an $T \times T$ upper-triangular matrix. As before, all random variables $\boldsymbol{C}_{ij}$ are independent and now $|\boldsymbol{C}_{ii}|^2 \sim \chi^2_{M-i+1}$. Therefore, one has

$$\begin{aligned} \frac{|\boldsymbol{W}|}{\mathrm{Tr}\{\boldsymbol{W}\}^T} &= \frac{\prod_{t=1}^{T} |\boldsymbol{C}_{tt}|^2}{\left[\sum_{t=1}^{T} |\boldsymbol{C}_{tt}|^2 + \sum_{i\neq j} |\boldsymbol{C}_{ij}|^2\right]^T} \\ &\overset{d}{=} \frac{\prod_{t=1}^{T} \chi^2_{M-t+1}}{\left[\chi^2_{T(T-1)/2} + \sum_{t=1}^{T} \chi^2_{M-t+1}\right]^T}. \end{aligned} \tag{27}$$

Finally, in the under-sampled case, one obtains

$$LR^u(\boldsymbol{R}_0|\boldsymbol{X}) \overset{d}{=} \tilde{C}\mathcal{R}^{2T^2} g_T\left(\mathcal{R}^2\right) \frac{\left(\prod_{t=1}^{T} \chi^2_{M-t+1}\right)^T}{\left[\chi^2_{T(T-1)/2} + \sum_{t=1}^{T} \chi^2_{M-t+1}\right]^{T^2}}. \tag{28}$$

Similarly to the over-sampled case, the p.d.f. of $LR^u(\boldsymbol{R}_0|\boldsymbol{X})$ is independent of $\boldsymbol{R}_0$ but depends on $g(.)$. It is noteworthy by comparing (17) and (28) that, when $T = M$, the over-sampled LR and the under-sampled LR coincide.

It is possible to gather (17) and (28) under the same umbrella, which yields a *unified representation* of $LR(\boldsymbol{R}_0|\boldsymbol{X})$ as

$$LR(\boldsymbol{R}_0|\boldsymbol{X}) \overset{d}{=} \left[\max_t t^{PT} g(tP)\right]^{-1} \mathcal{R}^{2TP} g\left(\mathcal{R}^2\right) \frac{\left(\prod_{p=1}^{P} \chi^2_{Q-p+1}\right)^T}{\left[\chi^2_{P(P-1)/2} + \sum_{p=1}^{P} \chi^2_{Q-p+1}\right]^{PT}} \tag{29}$$

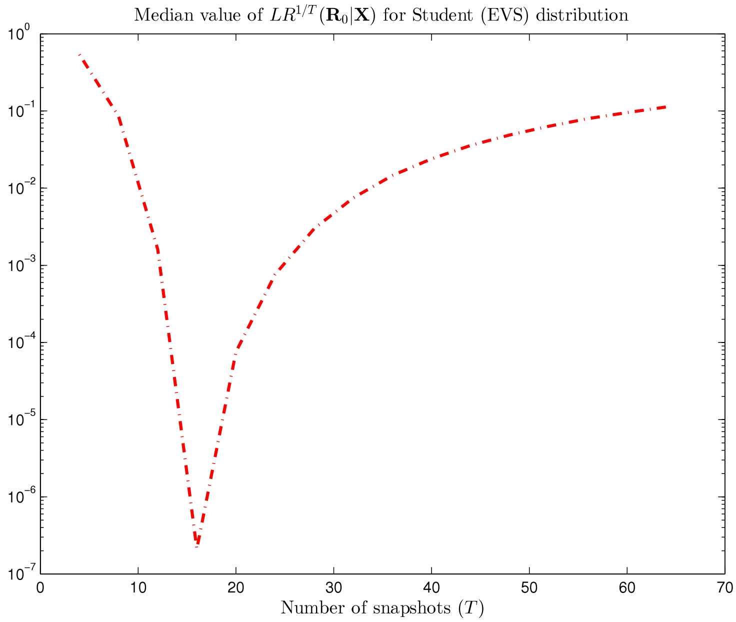

Fig. 1. Median value of $LR^{1/T}(\boldsymbol{R}_0|\boldsymbol{X})$ for a Student (EVS-type) distribution versus T. $M = 16$ and $d = 1$.

where $P = \min(M, T)$ and $Q = \max(M, T)$. As an illustration, Fig. 1 displays the median value of $LR^{1/T}(\boldsymbol{R}_0|\boldsymbol{X})$ for a Student (EVS-type) distribution with d degrees of freedom, which corresponds to $g(t) \propto \left[1 + d^{-1}\mathrm{Tr}\{\boldsymbol{X}^H\boldsymbol{R}_0^{-1}\boldsymbol{X}\}\right]^{-(MT+d)}$. As can be observed, this median value is well below one and hence only regularization of the MLE can drive the LR down to values compatible with those of $\boldsymbol{R}_0$.

To summarize, the required-for-EL-implementation invariance properties of $LR(\boldsymbol{R}_0|\boldsymbol{X})$ have been proven for EVS distributions, and the stochastic representations derived above allow for easy simulation of the p.d.f. of $LR(\boldsymbol{R}_0|\boldsymbol{X})$ for any M, T and $g(.)$. This makes the EL approach for regularized CME possible in the EVS class of distributions.

4. Likelihood ratio for $\mathcal{ESS}_{M,T}(\boldsymbol{0}, \boldsymbol{R}, g)$ distributions

In this section, we briefly consider the LR for $\mathcal{ESS}_{M,T}(\boldsymbol{0}, \boldsymbol{R}, g)$ distributions and we give an illustration on the matrix-variate Student distribution. We begin with the over-sampled case $T \geq M$. Let $\boldsymbol{X} \sim \mathcal{ESS}_{M,T}(\boldsymbol{0}, \boldsymbol{R}, g)$ so that

$$p(\boldsymbol{X}|\boldsymbol{R}, g) = |\boldsymbol{R}|^{-T} g\left(\lambda(\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1})\right)$$

where $\lambda(\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1})$ stands for the diagonal matrix of the (M non-zero) eigenvalues of $\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1}$. As shown in [11] the MLE of $\boldsymbol{R}$ is given by

$$\boldsymbol{R}_{\mathrm{ML}} = \frac{\boldsymbol{X}\boldsymbol{X}^H}{\alpha} \tag{30}$$

where $\alpha = \arg\max_t t^{MT} g(t\boldsymbol{I}_M)$. The LR is thus given by

$$\begin{aligned}
LR(\boldsymbol{R}(\boldsymbol{\beta})|\boldsymbol{X}) &= \frac{p(\boldsymbol{X}|\boldsymbol{R}(\boldsymbol{\beta}))}{p(\boldsymbol{X}|\boldsymbol{R}_{\mathrm{ML}})} \\
&= |\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{R}_{\mathrm{ML}}|^T \frac{g\left(\lambda(\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta}))\right)}{g\left(\lambda(\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}_{\mathrm{ML}}^{-1})\right)} \\
&= |\alpha^{-1}\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})|^T \frac{g\left(\lambda(\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta}))\right)}{g\left(\lambda(\alpha\boldsymbol{I}_M)\right)} \\
&= C|\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})|^T g\left(\lambda(\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta}))\right)
\end{aligned} \tag{31}$$

with $C = \alpha^{-MT}/g(\alpha\boldsymbol{I}_M) = \left[\max_t t^{MT} g(t\boldsymbol{I}_M)\right]^{-1}$. From the stochastic representation $\boldsymbol{X} \overset{d}{=} \boldsymbol{R}_0^{1/2}\boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^H$ with $\boldsymbol{U} \in O(M)$, $\boldsymbol{V} \sim \mathcal{U}_{T,M}$, it ensues that

$$\begin{aligned} LR(\boldsymbol{R}_0|\boldsymbol{X}) &= C|\boldsymbol{R}_0^{-1/2}\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}_0^{-1/2}|^T g\left(\lambda(\boldsymbol{R}_0^{-1/2}\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}_0^{-1/2})\right) \\ &\overset{d}{=} C|\boldsymbol{\Lambda}|^T g(\boldsymbol{\Lambda}) \end{aligned} \tag{32}$$

which is clearly independent of $\boldsymbol{R}_0$ but of course depends on $g(.)$.

For illustration purposes, let us consider the more insightful and practically important case of the matrix-variate Student distribution with $d\,(\geq T)$ degrees of freedom, defined as

$$\begin{aligned} p(\boldsymbol{X}|\boldsymbol{R}) &\propto |\boldsymbol{R}|^{-T}|\boldsymbol{I}_T + \boldsymbol{X}^H\boldsymbol{R}^{-1}\boldsymbol{X}|^{-(d+M)} \\ &\propto |\boldsymbol{R}|^{-T}|\boldsymbol{I}_M + \boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1}|^{-(d+M)}. \end{aligned}$$

It is well known that the MLE of $\boldsymbol{R}$, under the assumption $T \geq M$, is given by

$$\boldsymbol{R}_{\mathrm{ML}} = \frac{d+M-T}{T}\boldsymbol{X}\boldsymbol{X}^H. \tag{33}$$

Therefore, the LR for any covariance matrix $\boldsymbol{R}(\boldsymbol{\beta})$ takes the form

$$LR(\boldsymbol{R}(\boldsymbol{\beta})|\boldsymbol{X}) = C|\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})|^T|\boldsymbol{I}_M + \boldsymbol{X}\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})|^{-(d+M)}. \tag{34}$$

The stochastic representation of $\boldsymbol{X}$ writes

$$\boldsymbol{X} \overset{d}{=} \boldsymbol{R}_0^{1/2}\boldsymbol{N}_1\left(\boldsymbol{N}_2\boldsymbol{N}_2^H\right)^{-1/2} \tag{35}$$

where $\boldsymbol{N}_1 \in \mathbb{C}^{M\times T} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_M, \boldsymbol{I}_T)$ and $\boldsymbol{N}_2 \in \mathbb{C}^{T\times d} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_T, \boldsymbol{I}_d)$ are independent. Then, at the true covariance matrix, the LR admits the representation

$$LR(\boldsymbol{R}_0|\boldsymbol{X}) \overset{d}{=} C|\boldsymbol{F}|^T|\boldsymbol{I}_M + \boldsymbol{F}|^{-(d+M)} \tag{36}$$

where $\boldsymbol{F} = \boldsymbol{N}_1\left(\boldsymbol{N}_2\boldsymbol{N}_2^H\right)^{-1/2}\boldsymbol{N}_1^H$ follows a F-distribution [13].

Let us now investigate the under-sampled scenario where $T \leq M$. As a preliminary it is requested to consider ESS distributions with a rank $-R$ matrix $\boldsymbol{R} = \boldsymbol{U}_R\boldsymbol{D}_R\boldsymbol{U}_R^H$. Making the same change of variables as in (20), one has $\tilde{\boldsymbol{X}}_R \sim \mathcal{ESS}_{R,T}(\boldsymbol{0}, \boldsymbol{D}_R, g_R)$, i.e.,

$$\begin{aligned} p(\tilde{\boldsymbol{X}}_R|\boldsymbol{U}_R, \boldsymbol{D}_R) &= |\boldsymbol{D}_R|^{-T} g_R\left(\lambda(\tilde{\boldsymbol{X}}_R\tilde{\boldsymbol{X}}_R^H\boldsymbol{D}_R^{-1})\right) \\ &= |\boldsymbol{D}_R|^{-T} g_R\left(\lambda(\boldsymbol{U}_R^H\boldsymbol{X}\boldsymbol{X}^H\boldsymbol{U}_R\boldsymbol{D}_R^{-1})\right) \\ &= |\boldsymbol{D}_R|^{-T} g_R\left(\lambda(\boldsymbol{X}^H\boldsymbol{R}^{-}\boldsymbol{X})\right). \end{aligned} \tag{37}$$

Then one can define a density on the set $\left\{\boldsymbol{X} \in \mathbb{C}^{M\times T} / \left(\boldsymbol{U}_R^{\perp}\right)^H\boldsymbol{X} = \boldsymbol{0}\right\}$ as

$$p(\boldsymbol{X}|\boldsymbol{U}_R, \boldsymbol{D}_R) = |\boldsymbol{D}_R|^{-T} g_R\left(\lambda(\boldsymbol{X}^H\boldsymbol{R}^{-}\boldsymbol{X})\right). \tag{38}$$

Therefore, when $T \leq M$ and $\boldsymbol{R}_T$ in (19) being an admissible singular covariance matrix, we get

$$\begin{aligned} p(\boldsymbol{X}|\boldsymbol{R}(\boldsymbol{\beta})) &= |\boldsymbol{D}_T|^{-T} g_T\left(\lambda(\boldsymbol{X}^H\hat{\boldsymbol{U}}_T\boldsymbol{D}_T^{-1}\hat{\boldsymbol{U}}_T^H\boldsymbol{X})\right) \\ &= |\hat{\boldsymbol{U}}_T^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\hat{\boldsymbol{U}}_T|^T g_T\left(\lambda(\boldsymbol{X}^H\hat{\boldsymbol{U}}_T\left[\hat{\boldsymbol{U}}_T^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\hat{\boldsymbol{U}}_T\right]\hat{\boldsymbol{U}}_T^H\boldsymbol{X})\right). \end{aligned} \tag{39}$$

The MLE of $\boldsymbol{D}_T$ is now given by

$$\boldsymbol{D}_T^{ML} = \frac{\hat{\boldsymbol{U}}_T^H\boldsymbol{X}\boldsymbol{X}^H\hat{\boldsymbol{U}}_T}{\tilde{t}_{\max}} = \frac{\hat{\boldsymbol{\Lambda}}_T}{\tilde{t}_{\max}} \tag{40}$$

where $\tilde{t}_{\max} = \arg\max_t t^{T^2} g_T(t\boldsymbol{I}_T)$. The under-sampled LR is thus given by

$$\begin{aligned} LR^u(\boldsymbol{R}(\boldsymbol{\beta})|\boldsymbol{X}) &= |\hat{\boldsymbol{U}}_T^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\hat{\boldsymbol{U}}_T\boldsymbol{D}_T^{ML}|^T\frac{g_T\left(\lambda(\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X})\right)}{g_T\left(\tilde{t}_{\max}\boldsymbol{I}_T\right)} \\ &= \tilde{C}|\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}|^T g_T\left(\lambda(\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X})\right) \end{aligned} \tag{41}$$

with $\tilde{C} = \tilde{t}_{\max}^{-T^2}/g(T\tilde{t}_{\max}) = \left[\max_t t^{T^2} g_T(t\boldsymbol{I}_T)\right]^{-1}$. The stochastic representation of $\boldsymbol{X}$ writes $\boldsymbol{X} \overset{d}{=} \boldsymbol{R}_0^{1/2}\boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^H$ with $\boldsymbol{U} \sim \mathcal{U}_{M,T}$, $\boldsymbol{V} \in O(T)$, and hence

$$\begin{aligned} LR^u(\boldsymbol{R}_0|\boldsymbol{X}) &= \tilde{C}|\boldsymbol{X}^H\boldsymbol{R}_0^{-1}\boldsymbol{X}|^T g_T\left(\lambda(\boldsymbol{X}^H\boldsymbol{R}_0^{-1}\boldsymbol{X})\right) \\ &\overset{d}{=} \tilde{C}|\boldsymbol{\Lambda}|^T g_T(\boldsymbol{\Lambda}). \end{aligned} \tag{42}$$

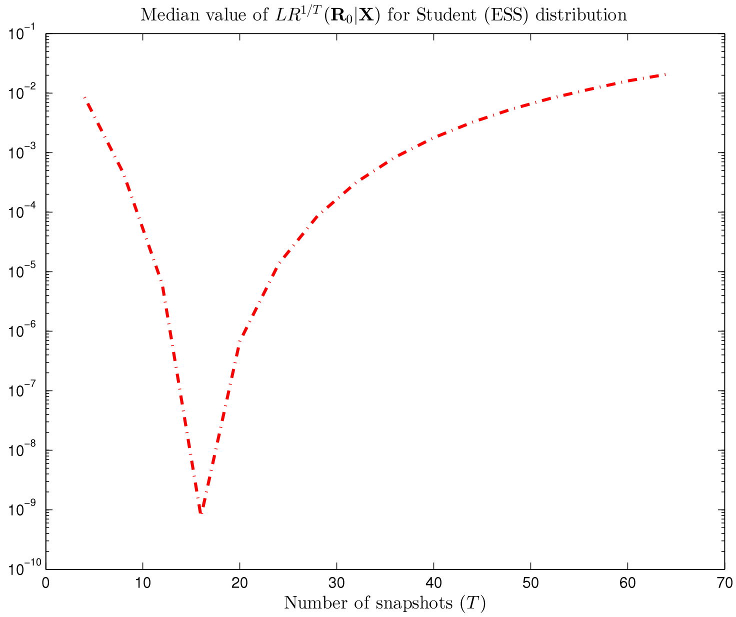

Fig. 2. Median value of $LR^{1/T}(\boldsymbol{R}_0|\boldsymbol{X})$ for a Student (ESS-type) distribution versus T. $M = 16$ and $d = T$.

For illustration purposes, let us study again the multivariate Student distribution. We first consider $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{G}\boldsymbol{S}_2^{-1/2}$ where $\boldsymbol{G} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R} = \boldsymbol{U}_R\boldsymbol{D}_R\boldsymbol{U}_R^H)$ and $\boldsymbol{S}_2 \sim \mathcal{W}(d, \boldsymbol{I}_T)$. The p.d.f. of $\tilde{\boldsymbol{X}}_R = \boldsymbol{U}_R^H\boldsymbol{X}$ is easily obtained as

$$\begin{aligned} p(\tilde{\boldsymbol{X}}_R|\boldsymbol{U}_R, \boldsymbol{D}_R) &\propto |\boldsymbol{D}_R|^{-T}|\boldsymbol{I}_T + \tilde{\boldsymbol{X}}_R^H\boldsymbol{D}_R^{-1}\tilde{\boldsymbol{X}}_R|^{-(d+R)} \\ &\propto |\boldsymbol{D}_R|^{-T}|\boldsymbol{I}_T + \boldsymbol{X}^H\boldsymbol{U}_R\boldsymbol{D}_R^{-1}\boldsymbol{U}_R^H\boldsymbol{X}|^{-(d+R)} \end{aligned} \tag{43}$$

from which one can define the density

$$p(\boldsymbol{X}|\boldsymbol{R}) \propto |\boldsymbol{D}_R|^{-T}|\boldsymbol{I}_T + \boldsymbol{X}^H\boldsymbol{R}^{-}\boldsymbol{X}|^{-(d+R)}\delta\left((\boldsymbol{U}_R^{\perp})^H\boldsymbol{X}\right). \tag{44}$$

In the under-sampled case, the p.d.f. of the observations can thus be written as

$$\begin{aligned} p(\boldsymbol{X}|\boldsymbol{R}(\beta)) &\propto |\boldsymbol{D}_T|^{-T}|\boldsymbol{I}_T + \boldsymbol{X}^H\hat{\boldsymbol{U}}_T\boldsymbol{D}_T^{-1}\hat{\boldsymbol{U}}_T^H\boldsymbol{X}|^{-(d+T)} \\ &\propto |\boldsymbol{D}_T|^{-T}|\boldsymbol{I}_T + \boldsymbol{X}^H\hat{\boldsymbol{U}}_T\left[\hat{\boldsymbol{U}}_T^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\hat{\boldsymbol{U}}_T\right]\hat{\boldsymbol{U}}_T^H\boldsymbol{X}|^{-(d+T)} \\ &\propto |\hat{\boldsymbol{U}}_T^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\hat{\boldsymbol{U}}_T|^T|\boldsymbol{I}_T + \boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}|^{-(d+T)}. \end{aligned} \tag{45}$$

It ensues that the LR for $\boldsymbol{R}(\boldsymbol{\beta})$ is given by

$$\begin{aligned} LR^u(\boldsymbol{R}(\boldsymbol{\beta})|\boldsymbol{X}) &= \left|\hat{\boldsymbol{U}}_T^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\hat{\boldsymbol{U}}_T\frac{d}{T}\boldsymbol{\Lambda}_T\right|^T \frac{|\boldsymbol{I}_T + \boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}|^{-(d+T)}}{|\boldsymbol{I}_T + d^{-1}T\boldsymbol{I}_T|^{-(d+T)}} \\ &= \tilde{C}|\boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}|^T|\boldsymbol{I}_T + \boldsymbol{X}^H\boldsymbol{R}^{-1}(\boldsymbol{\beta})\boldsymbol{X}|^{-(d+T)} \end{aligned} \tag{46}$$

with $\tilde{C} = (d^{-1}T)^{T^2}(1 + d^{-1}T)^{(d+T)T}$. Finally, using $\boldsymbol{X} \stackrel{d}{=} \boldsymbol{R}_0^{1/2}\boldsymbol{N}_1\boldsymbol{S}_2^{-1/2}$, the under-sampled LR evaluated at $\boldsymbol{R}_0$ is finally given by

$$LR^u(\boldsymbol{R}_0|\boldsymbol{X}) \stackrel{d}{=} \tilde{C}|\boldsymbol{S}_2^{-1}\boldsymbol{S}_1|^T|\boldsymbol{I}_T + \boldsymbol{S}_2^{-1}\boldsymbol{S}_1|^{-(d+T)} \tag{47}$$

with $\boldsymbol{S}_1 \sim \mathcal{W}(M, \boldsymbol{I}_T)$. In Fig. 2, we display the median value of $LR^{1/T}(\boldsymbol{R}_0|\boldsymbol{X})$ in the case of a multivariate Student (ESS-type) distribution as a function of the number of snapshots. Again, it is observed that this median value is well below 1.

5. Conclusions

Invariance properties of the likelihood ratio for testing a covariance matrix were extended to a class of complex elliptically contoured distributions. The stochastic representations of $LR(\boldsymbol{R}_0|\boldsymbol{X})$ derived herein allow for a straightforward evaluation of its p.d.f. for a given set of values of M, T and $g(.)$. This paves the way to application of the EL approach to regularized covariance matrix where regularization parameters $\boldsymbol{\beta}$ can be chosen such that $LR(\boldsymbol{R}(\boldsymbol{\beta})|\boldsymbol{X})$ falls in the support of the p.d.f. of $LR(\boldsymbol{R}_0|\boldsymbol{X})$.

References

[1] Y.I. Abramovich, O. Besson, Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach – Part 1: the oversampled case, IEEE Trans. Signal Process. 61 (23) (2013) 5807–5818.
[2] Y.I. Abramovich, B.A. Johnson, GLRT-based detection-estimation for undersampled training conditions, IEEE Trans. Signal Process. 56 (8) (2008) 3600–3612.
[3] Y. Abramovich, N. Spencer, A. Gorokhov, Bounds on maximum likelihood ratio—part I: application to antenna array detection–estimation with perfect wavefront coherence, IEEE Trans. Signal Process. 52 (6) (2004) 1524–1536.
[4] Y.I. Abramovich, N.K. Spencer, A.Y. Gorokhov, Modified GLRT and AMF framework for adaptive detectors, IEEE Trans. Aerosp. Electron. Syst. 43 (3) (2007) 1017–1051.
[5] Y.I. Abramovich, N.K. Spencer, A.Y. Gorokhov, GLRT-based threshold detection–estimation performance improvement and application to uniform circular antenna arrays, IEEE Trans. Signal Process. 55 (1) (2007) 20–31.
[6] T.W. Anderson, K.-T. Fang, Theory and applications of elliptically contoured and related distributions, in: Tech. Rep. 24, U. S. Army Research Office, 1990.
[7] T.W. Anderson, K.-T. Fang, H. Hsu, Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions, Canad. J. Statist. 14 (1) (1986) 55–59.
[8] O. Besson, Y.I. Abramovich, Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach – Part 2: the undersampled case, IEEE Trans. Signal Process. 61 (23) (2013) 5819–5829.
[9] J.A. Díaz-García, G. González-Farías, Singular random matrix decompositions: distributions, J. Multivariate Anal. 94 (1) (2005) 109–122.
[10] K.-T. Fang, H.-F. Chen, Relationships among classes of spherical matrix distributions, in: Tech. Rep. no. 10, U. S. Army Research Office, 1984.
[11] K.T. Fang, Y.T. Zhang, Generalized Multivariate Analysis, Springer Verlag, Berlin, 1990.
[12] D. Fourdrinier, W.E. Strawderman, M.T. Wells, Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix, J. Multivariate Anal. 85 (1) (2003) 24–39.
[13] N.R. Goodman, Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction), Ann. of Math. Stud. 34 (1) (1963) 152–177.
[14] A.K. Gupta, D.K. Nagar, Matrix Variate Distributions, Chapman & Hall/CRC, Boca Raton, FL, 2000.
[15] D.R. Jensen, I.J. Good, Invariant distributions associated with matrix laws under structural symmetry, J. R. Stat. Soc. Ser. B Stat. 43 (3) (1981) 427–432.
[16] T. Kariya, Robustness of multivariate tests, Ann. Statist. 9 (6) (1981) 1267–1281.
[17] E.J. Kelly, An adaptive detection algorithm, IEEE Trans. Aerosp. Electron. Syst. 22 (1) (1986) 115–127.
[18] T. Kubokawa, M.S. Srivastava, Robust improvement in estimation of a covariance matrix in an elliptically contoured distribution, Ann. Statist. 27 (2) (1999) 600–609.
[19] O. Ledoit, M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivariate Analysis 88 (2) (2004) 365–411.
[20] P.L. Leung, F.Y. Ng, Improved estimation of a covariance matrix in an elliptically contoured matrix distribution, J. Multivariate Anal. 88 (1) (2004) 131–137.
[21] R.J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley, 1982.
[22] E. Ollila, D. Tyler, V. Koivunen, H. Poor, Complex elliptically symmetric distributions: survey, new results and applications, IEEE Trans. Signal Process. 60 (11) (2012) 5597–5625.
[23] C.D. Richmond, Adaptive array signal processing and performance analysis in non-Gaussian environments, Ph.D. Thesis, Massachusetts Institute of Technology (1996).
[24] C.D. Richmond, A note on non-Gaussian adaptive array detection and signal parameter estimation, IEEE Signal Process. Lett. 3 (8) (1996) 251–252.
[25] M. Siotani, T. Hayakawa, Y. Fujikoto, Modern Multivariate Statistical Analysis, American Science Press, Cleveland, OH, 1985.
[26] M.S. Srivastava, C.G. Khatri, An Introduction to Multivariate Statistics, Elsevier, North Holland Inc., New York, 1979.
[27] A. Wiesel, Unified framework to regularized covariance estimation in scaled Gaussian models, IEEE Trans. Signal Process. 60 (1) (2012) 29–38.