COVARIANCE MATRIX ESTIMATION IN COMPLEX ELLIPTIC DISTRIBUTIONS USING THE EXPECTED LIKELIHOOD APPROACH

Yuri I. Abramovich

W R Systems, Ltd.
11351 Random Hills Road, Suite 400
Fairfax, VA 22030
yabramovich@wrsystems.com

Olivier Besson[*]

University of Toulouse - ISAE
10 Avenue Edouard Belin
31055 Toulouse, France
olivier.besson@isae.fr

ABSTRACT

We consider the problem of estimating the scatter matrix in complex elliptically symmetric (CES) distributions using the expected likelihood (EL) approach. The latter, originally derived in the Gaussian case, is based on the fact that the probability density function (p.d.f.) of the likelihood ratio (LR) for the (unknown) actual covariance matrix does not depend on this matrix, and is fully specified by the matrix dimension M and the number of independent training samples T. We extend this result to CES distributions as well as to angular central Gaussian (ACG) distributions. More precisely, we prove that for CES distributions, the p.d.f. of the LR, evaluated at the true scatter matrix $\boldsymbol{\Sigma}_0$, does not depend on the latter but depends on the density generator of the CES distribution. As for the ACG case, we demonstrate that the LR for $\boldsymbol{\Sigma}_0$ is distribution-free. This invariance property paves the way to derivation of regularized covariance matrix estimates, where the regularization parameters are chosen from the EL principle. The relevance of such a choice for the regularization parameters is illustrated on an example with fixed-point diagonally loaded estimates.

Index Terms— Complex elliptically symmetric distributions, covariance matrix estimation, expected likelihood principle, likelihood ratio, regularization.

1. INTRODUCTION

Covariance matrix estimation plays a central role in many array processing applications, including direction of arrival estimation, design of adaptive filters or adaptive detection. Most often, the maximum likelihood (ML) approach is invoked due to its good asymptotic properties. However, in small sample support, adaptive filters based on ML estimation of the disturbance covariance matrix do not perform well. At least, they can be significantly improved over, in terms of output signal to noise ratio (SNR) and measure of effectiveness, by regularization schemes such as diagonal loading [1]. Moreover, the ML estimator yields the ultimate equal to one value for the likelihood ratio (LR) which, as argued in [2, 3], is questionable. Indeed, in the Gaussian case, it was demonstrated in [2, 3] that the probability density function (p.d.f.) of $LR(\boldsymbol{R}_0)$ (where $\boldsymbol{R}_0$ is the true (actual) covariance matrix) does not depend on $\boldsymbol{R}_0$ and is fully specified by matrix dimension M and number T of independent identically distributed (i.i.d.) samples. Moreover, this p.d.f. is concentrated around values of $LR(\boldsymbol{R}_0)$ which are much lower than 1. This quite remarkable invariance property led to the development of the so-called "Expected Likelihood" (EL) approach. The latter postulates that any covariance matrix estimate (CME) should result in a LR which is commensurate with that of the true covariance matrix, and this value of the LR can be pre-calculated, due to the above mentioned invariance property of $LR(\boldsymbol{R}_0)$. The EL principle is therefore a statistically sound method to assess the "quality" of any (possibly parameterized or regularized) covariance matrix estimate $\boldsymbol{R}(\hat{\boldsymbol{\Omega}})$ by comparing its likelihood ratio $LR(\boldsymbol{R}(\hat{\boldsymbol{\Omega}}))$ against the p.d.f. for $LR(\boldsymbol{R}_0)$. Its relevance has been demonstrated e.g. in [2, 4] where the EL principle allows for detection of severely erroneous MUSIC-based DOA estimates in the so-called threshold area (breakdown prediction), and rectification of the latter in order to meet the expected likelihood ratio values (breakdown cure). Similarly, for adaptive detection problems, the EL approach was instrumental in designing regularized CME, whose regularization parameters are chosen such that the so-obtained CME is statistically as likely as the unknown actual covariance matrix [3].

However, in a large number of radar applications, the traditional assumption on training data being a set of i.i.d complex Gaussian random samples is strongly violated due to a significant in-homogeneity of this data. The latter has been often modeled as spherically invariant random vectors (SIRV), i.e., as a product of a positive valued random variable (r.va.) called texture and an independent complex Gaussian random vector (r.v.) called speckle [5]. This so-called compound Gaussian model belongs to the broader class of complex elliptically symmetric (CES) distributions which have recently gained much interest for array processing applications [6] and particularly for covariance matrix estimation problems [7, 8]. Therefore, the main issue addressed in this paper is to extend the EL principle to the class of CES distributions. More precisely, we show that the likelihood ratio $LR_{CES}(\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the scatter matrix, proportional to the covariance matrix, shares the required for EL invariance principle: its p.d.f. for $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ does not depend on $\boldsymbol{\Sigma}_0$. However, it depends on a one-dimensional function $g(t)$ called the density generator of the CES distribution. Since $g(t)$ is unknown in practice, it becomes desirable to have invariance also with respect to the unknown function $g(.)$. We thus also consider complex angular central Gaussian (ACG) distributions, obtained by normalization of the r.v. $\boldsymbol{x} \sim \mathbb{C}\mathcal{E}_M(\boldsymbol{0}, \boldsymbol{\Sigma}, g)$ by its L_2 norm: $\boldsymbol{z} = \boldsymbol{x}/\left\|\boldsymbol{x}\right\|_2$. We prove that $LR_{ACG}(\boldsymbol{\Sigma}_0)$ is now completely invariant with respect to both $\boldsymbol{\Sigma}_0$ and $g(t)$ in the CES distribution that describes r.v. $\boldsymbol{x}$. This invariance property allows for design of regularized CME using the EL principle. We illustrate this fact with fixed-point diagonally loaded scatter matrix estimates.

[*]The work of O. Besson was partly supported by DGA/MRIS under grant no. 2012.60.0012.00.470.75.01.

2. LIKELIHOOD RATIO FOR CES DISTRIBUTIONS

2.1. CES distributions

We refer the reader to [6] for a very comprehensive review along with application of CES distributions to a number of array processing problems. A random vector $\boldsymbol{x} \in \mathbb{C}^M \sim \mathbb{C}\mathcal{E}_M(\mathbf{0}, \boldsymbol{\Sigma}, g)$ if it admits the following stochastic representation

$$\boldsymbol{x} \overset{d}{=} \mathcal{R}\boldsymbol{A}\boldsymbol{u} \tag{1}$$

where the non-negative real random variable $\mathcal{R} \triangleq \sqrt{\mathcal{Q}}$, called modular variate, is independent of the complex random vector $\boldsymbol{u}$, which is uniformly distributed on the complex sphere $\mathbb{C}S^M$, i.e., $\boldsymbol{u} \sim \mathcal{U}(\mathbb{C}S^M)$. $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^H$ is a factorization of $\boldsymbol{\Sigma}$ and the latter is referred to as the scatter matrix. Herein, $\overset{d}{=}$ means "has the same distribution as". Let the p.d.f. $f(\mathcal{Q})$ of the modular variate $\mathcal{Q}$ be given by $f(\mathcal{Q}) = \delta_{M,g}^{-1}\mathcal{Q}^{M-1}g(\mathcal{Q})$ where $g : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is called the density generator. We consider herein the absolutely continuous case where $\boldsymbol{\Sigma}$ is non singular. Then the p.d.f. of $\boldsymbol{x}$ is given by

$$f(\boldsymbol{x}|\boldsymbol{\Sigma}, g) = C_{M,g}|\boldsymbol{\Sigma}|^{-1} g\left(\boldsymbol{x}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right) \tag{2}$$

where $C_{M,g} = 2(S_M\delta_{M,g})^{-1}$ and $S_M = 2\pi^M/\Gamma(M)$ denotes the surface area of $\mathbb{C}S^M$. Under the assumption that $\mathcal{E}\{\mathcal{R}^p\} < \infty$, one has $\mathcal{E}\{\boldsymbol{x}\boldsymbol{x}^H\} = M^{-1}\mathcal{E}\{\mathcal{R}^2\}\boldsymbol{\Sigma}$ and thus the scatter matrix $\boldsymbol{\Sigma}$ is proportional to the covariance matrix. Since the couple $(g, \boldsymbol{\Sigma})$ does not uniquely identify $f(\boldsymbol{x}|\boldsymbol{\Sigma}, g)$, a scale constraint on $g(.)$ or $\boldsymbol{\Sigma}$ is usually enforced: herein, we assume that $\mathcal{E}\{\mathcal{R}^2\} = M$ and hence $\text{cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}$.

We assume that T i.i.d r.v. $\boldsymbol{x}_t \sim \mathbb{C}\mathcal{E}_M(\mathbf{0}, \boldsymbol{\Sigma}, g)$ are available so that the joint p.d.f. of $\boldsymbol{X}_T = \begin{bmatrix}\boldsymbol{x}_1 & \cdots & \boldsymbol{x}_T\end{bmatrix} \in \mathbb{C}^{M\times T}$ can be written as

$$f(\boldsymbol{X}_T|\boldsymbol{\Sigma}, g) = C_{M,g}^T|\boldsymbol{\Sigma}|^{-T}\prod_{t=1}^{T} g(\boldsymbol{x}_t^H\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_t). \tag{3}$$

For $T \geq M$, the maximum likelihood estimator (MLE) of the scatter matrix $\boldsymbol{\Sigma}$ is the solution to the estimating equation [6, 9, 10]

$$\boldsymbol{\Sigma}_{\text{ML}} = \mathcal{T}(\boldsymbol{\Sigma}_{\text{ML}}) = \frac{1}{T}\sum_{t=1}^{T}\phi(\boldsymbol{x}_t^H\boldsymbol{\Sigma}_{\text{ML}}^{-1}\boldsymbol{x}_t)\boldsymbol{x}_t\boldsymbol{x}_t^H \tag{4}$$

where $\phi(t) \triangleq -g'(t)/g(t)$. In [6] based on the results of Kent and Tyler [9, 10] for the real case, the uniqueness and convergence of the fixed point iterations $(\boldsymbol{\Sigma}_{\text{ML}})_{k+1} = \mathcal{T}\left[(\boldsymbol{\Sigma}_{\text{ML}})_k\right]$ to the unique solution of (4), for any initial estimate of $\boldsymbol{\Sigma}$, has been proven under certain technical conditions.

2.2. Likelihood ratio for CES distributions

Our goal is to estimate $\boldsymbol{\Sigma}$ using the EL principle. Towards this end, a first step consists in deriving the likelihood ratio for the true scatter matrix. We start with a parametric scatter matrix model $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$ where $\boldsymbol{\Omega}_\ell$ is a set of ℓ parameters that uniquely specify the scatter matrix model. The $LR_{CES}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{X}_T)$ is obtained from (3) as [11]:

$$\begin{aligned} LR_{CES}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{X}_T, g) &= \frac{f(\boldsymbol{X}_T|\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell), g)}{\max_{\boldsymbol{\Sigma}} f(\boldsymbol{X}_T|\boldsymbol{\Sigma}, g)} \\ &= |\boldsymbol{\Sigma}_{\text{ML}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)|^T\prod_{t=1}^{T}\frac{g(\boldsymbol{x}_t^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)\boldsymbol{x}_t)}{g(\boldsymbol{x}_t^H\boldsymbol{\Sigma}_{\text{ML}}^{-1}\boldsymbol{x}_t)}. \end{aligned} \tag{5}$$

Now, using the stochastic representation of $\boldsymbol{x}_t$ in (1), the expected likelihood, i.e., the LR value for the actual (true) scatter matrix $\boldsymbol{\Sigma}_0$ is given by

$$LR_{CES}(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g) \overset{d}{=} |\boldsymbol{A}|^T\prod_{t=1}^{T}\frac{g(\mathcal{Q}_t\boldsymbol{u}_t^H\boldsymbol{u}_t)}{g(\mathcal{Q}_t\boldsymbol{u}_t^H\boldsymbol{A}^{-1}\boldsymbol{u}_t)} \tag{6}$$

where $\boldsymbol{u}_t \sim \mathcal{U}(\mathbb{C}S^M)$ and $\boldsymbol{A} \triangleq \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma}_{\text{ML}}\boldsymbol{\Sigma}_0^{-1/2}$. Pre and post-multiplying (4) by $\boldsymbol{\Sigma}_0^{-1/2}$ and using (1) we get

$$\boldsymbol{A} = \frac{1}{T}\sum_{t=1}^{T}\phi(\mathcal{Q}_t\boldsymbol{u}_t^H\boldsymbol{A}^{-1}\boldsymbol{u}_t)\mathcal{Q}_t\boldsymbol{u}_t\boldsymbol{u}_t^H \tag{7}$$

whose distribution clearly depends on $g(.)$ but not on $\boldsymbol{\Sigma}_0$. Consequently, the p.d.f. of $LR_{CES}(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g)$ is invariant with respect to (w.r.t) the true scatter matrix $\boldsymbol{\Sigma}_0$, and is specified only by $f(\mathcal{Q})$, M and T.

2.3. Likelihood ratio for ACG distributions

In most practical applications, $f(\mathcal{Q})$ is not precisely known, hence the need to investigate estimation schemes which do not require this knowledge. The usual way to proceed is to normalize the vectors $\boldsymbol{x}_t$, i.e., to use as input data

$$\boldsymbol{z}_t = \frac{\boldsymbol{x}_t}{\|\boldsymbol{x}_t\|_2}. \tag{8}$$

If $\boldsymbol{x} \sim \mathbb{C}\mathcal{E}_M(\mathbf{0}, \boldsymbol{\Sigma}, g)$, then $\boldsymbol{z}$ follows a complex angular central Gaussian (ACG) distribution, which we denote as $\boldsymbol{z} \sim \mathbb{C}\mathcal{AG}(\mathbf{0}, \boldsymbol{\Sigma})$. For non-singular $\boldsymbol{\Sigma}$, the p.d.f. for $\boldsymbol{z}$ is given by [6, 12]

$$f(\boldsymbol{z}|\boldsymbol{\Sigma}) = S_M^{-1}|\boldsymbol{\Sigma}|^{-1}\left(\boldsymbol{z}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{z}\right)^{-M}. \tag{9}$$

Assuming independence of the $\boldsymbol{z}_t$, the joint distribution of $\boldsymbol{Z}_T = \begin{bmatrix}\boldsymbol{z}_1 & \cdots & \boldsymbol{z}_T\end{bmatrix}$ is thus given by

$$f(\boldsymbol{Z}_T|\boldsymbol{\Sigma}) = S_M^{-T}|\boldsymbol{\Sigma}|^{-T}\prod_{t=1}^{T}\left(\boldsymbol{z}_t^H\boldsymbol{\Sigma}^{-1}\boldsymbol{z}_t\right)^{-M}. \tag{10}$$

In [12] the MLE of $\boldsymbol{\Sigma}$ was shown to be the solution to

$$\boldsymbol{\Sigma}_{\text{ML}} = \frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\boldsymbol{\Sigma}_{\text{ML}}^{-1}\boldsymbol{z}_t}. \tag{11}$$

The estimate (11) is also the MLE of $\boldsymbol{\Sigma}$ under the more general assumption $\boldsymbol{x}_t \sim \mathbb{C}\mathcal{E}_M(\mathbf{0}, \tau_t\boldsymbol{\Sigma}, g_t)$ with the functions g_t being given but not necessarily the same [6]. Moreover, the fixed point iterations

$$\boldsymbol{\Sigma}_{k+1} = \frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\boldsymbol{\Sigma}_k^{-1}\boldsymbol{z}_t} \tag{12}$$

converge to $\boldsymbol{\Sigma}_{\text{ML}}$ which exists and is unique up to a positive scalar.

Let us study the EL approach as an alternative to MLE. For a (possibly parameterized) scatter matrix $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$, the likelihood ratio in the ACG case is given by

$$\begin{aligned} LR_{ACG}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T) &= \frac{f(\boldsymbol{Z}_T|\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell))}{\max_{\boldsymbol{\Sigma}} f(\boldsymbol{Z}_T|\boldsymbol{\Sigma})} \\ &= |\boldsymbol{\Sigma}_{\text{ML}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)|^T\prod_{t=1}^{T}\left[\frac{\boldsymbol{z}_t^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)\boldsymbol{z}_t}{\boldsymbol{z}_t^H\boldsymbol{\Sigma}_{\text{ML}}^{-1}\boldsymbol{z}_t}\right]^{-M}. \end{aligned} \tag{13}$$

Using the fact that $\boldsymbol{z}_t = \frac{\boldsymbol{x}_t}{\|\boldsymbol{x}_t\|_2} \overset{d}{=} \frac{\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{u}_t}{\left\|\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{u}_t\right\|_2}$ where $\boldsymbol{u}_t \sim \mathcal{U}\left(\mathbb{C}S^M\right)$ or $\boldsymbol{u}_t \sim \mathbb{C}\mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}\right)$, it follows that

$$\begin{aligned} LR_{ACG}\left(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T\right) &= |\boldsymbol{\Sigma}_{\text{ML}}\boldsymbol{\Sigma}_0^{-1}|^T \prod_{t=1}^{T}\left[\frac{\boldsymbol{z}_t^H\boldsymbol{\Sigma}_0^{-1}\boldsymbol{z}_t}{\boldsymbol{z}_t^H\boldsymbol{\Sigma}_{\text{ML}}^{-1}\boldsymbol{z}_t}\right]^{-M} \\ &\overset{d}{=} |\boldsymbol{A}_{ACG}|^T \prod_{t=1}^{T}\left[\frac{\boldsymbol{u}_t^H\boldsymbol{u}_t}{\boldsymbol{u}_t^H\boldsymbol{A}_{ACG}^{-1}\boldsymbol{u}_t}\right]^{-M} \end{aligned} \tag{14}$$

where $\boldsymbol{A}_{ACG} = \boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{\Sigma}_{\text{ML}}\boldsymbol{\Sigma}_0^{-1/2}$ verifies

$$\begin{aligned} \boldsymbol{A}_{ACG} &= \frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{z}_t\boldsymbol{z}_t^H\boldsymbol{\Sigma}_0^{-1/2}}{\boldsymbol{z}_t^H\boldsymbol{\Sigma}_{\text{ML}}^{-1}\boldsymbol{z}_t} \\ &\overset{d}{=} \frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{u}_t\boldsymbol{u}_t^H}{\boldsymbol{u}_t^H\boldsymbol{A}_{ACG}^{-1}\boldsymbol{u}_t}. \end{aligned} \tag{15}$$

Clearly $\boldsymbol{A}_{ACG}$ is *distribution-free* and therefore, for any given T and M ($T \geq M$) we can pre-calculate the p.d.f. for $LR_{ACG}\left(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T\right)$ with any required accuracy and use it as the expected likelihood p.d.f. for quality assessment of any given scatter matrix model $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$.

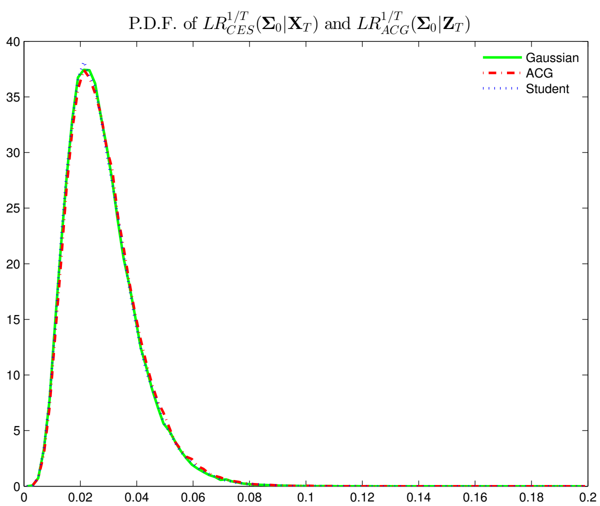

Fig. 1. Probability density function of $LR_{CES}^{1/T}\left(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g\right)$ and $LR_{ACG}^{1/T}\left(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T\right)$. $M = 12$ and $T = 24$.

2.4. Illustrations

We illustrate the above theoretical results about the distributions of $LR_{CES}^{1/T}\left(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g\right)$ and $LR_{ACG}^{1/T}\left(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T\right)$ and we investigate the influence of $g(.)$ onto the p.d.f. of the LR in the CES case. We consider a uniform linear array of $M = 12$ elements with half-wavelength separation. The true scatter matrix was considered to be as per AR(1) process, i.e., $\left[\boldsymbol{\Sigma}_0\right]_{m,n} = \rho_0^{|m-n|}$ and $\rho = 0.99$. As for CES distributions, we consider a Gaussian distribution and a multivariate Student t-distribution with d degrees of freedom, defined herein as

$$f(\boldsymbol{x}|\boldsymbol{\Sigma}_0) \propto |\boldsymbol{\Sigma}_0|^{-1}\left[1 + d^{-1}\boldsymbol{x}^H\boldsymbol{\Sigma}_0^{-1}\boldsymbol{x}\right]^{-(M+d)} \tag{16}$$

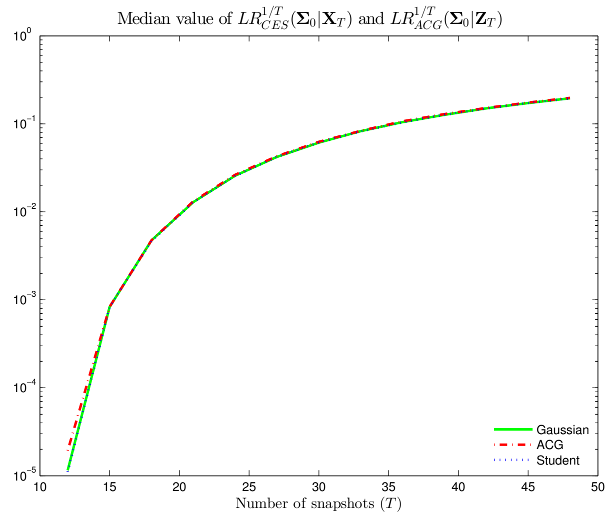

Fig. 2. Median value of $LR_{CES}^{1/T}\left(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g\right)$ and $LR_{ACG}^{1/T}\left(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T\right)$ versus T. $M = 12$.

where $\propto$ means proportional to. In the sequel we set $d = 1$. The p.d.f. of the above likelihood ratios is displayed in Figure 1 for $T = 24$. As can be observed, the p.d.f. of $LR_{CES}^{1/T}\left(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g\right)$ are very close to each other and depend weakly on $g(.)$: they are the same for $g(t) = \exp\left\{-t\right\}$ (Gaussian case) and $g(t) = (1 + t/d)^{-(d+M)}$ (Student case). Moreover, they are very close to the p.d.f. of $LR_{ACG}^{1/T}\left(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T\right)$. Therefore, the LR for the true scatter matrix $\boldsymbol{\Sigma}_0$ shows quite an invariance with respect to the distribution of the data. This is confirmed in Figure 2 where we plot the median value of the LR as a function of T. Clearly, the three different distributions result in almost the same median value. Note that the latter is much lower than 1 which calls for regularization schemes that drive down the LR compared to the MLE whose LR value is 1. This is the object of the next section.

3. REGULARIZED SCATTER MATRIX ESTIMATION USING THE EXPECTED LIKELIHOOD PRINCIPLE

As discussed above, the MLE in (11) obtained from the fixed point iterations in (12) yields the ultimate $LR_{ACG}\left(\boldsymbol{\Sigma}_{ML}|\boldsymbol{Z}_T\right) = 1$ value. However, this estimate may not be that effective for adaptive processing applications, since it is far more likely than the actual scatter matrix $\boldsymbol{\Sigma}_0$. For this reason, regularization based on diagonal loading or shrinkage-to-structure have been proposed in the literature. For instance, [13] considers time-varying autoregressive models TVAR(m). Due to lack of space, we consider here only the following scheme proposed initially in [14] and then in [7, 8]:

$$\boldsymbol{\Sigma}(\beta) = (1-\beta)\frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\boldsymbol{\Sigma}^{-1}(\beta)\boldsymbol{z}_t} + \beta\boldsymbol{I}_M \tag{17}$$

with the following iterative algorithm to obatin it:

$$\mathbf{\Sigma}_{k+1}(\beta) = (1-\beta)\frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\left(\mathbf{\Sigma}_k(\beta)\right)^{-1}\boldsymbol{z}_t} + \beta\boldsymbol{I}_M \quad (18a)$$

$$\mathbf{\Sigma}_{k+1}(\beta) = \frac{M}{\mathrm{Tr}\{\mathbf{\Sigma}_{k+1}(\beta)\}}\mathbf{\Sigma}_{k+1}(\beta). \quad (18b)$$

The proof of convergence of this iterative routine to the unique solution of (17) has been recently introduced in [7] using Perron-Frobenius theory. In [7] the authors suggested to specify the optimal loading factor β as the one which minimizes the Frobenius norm of the error, i.e.,

$$\beta_O = \arg\min_{\beta}\mathcal{E}\left\{\|\mathbf{\Sigma}(\beta)-\mathbf{\Sigma}_0\|_F^2\right\} = \frac{M^2 - M^{-1}\mathrm{Tr}\{\mathbf{\Sigma}_0\mathbf{\Sigma}_0^H\}}{(M^2 - MT - T) + (T + (T-1)/M)\mathrm{Tr}\{\mathbf{\Sigma}_0\mathbf{\Sigma}_0^H\}} \quad (19)$$

where $\mathbf{\Sigma}(\beta)$ is given in (17). This optimal loading factor depends on $\mathbf{\Sigma}_0$ which is unknown but, as suggested in [7], it could be replaced by a consistent estimate. Herein, we are interested in applying the EL principle to selection of β. More precisely, we propose to select the loading factor β such that

$$LR_{ACG}^{1/T}\left(\mathbf{\Sigma}(\beta)|\boldsymbol{Z}_T\right) = |\mathbf{\Sigma}_{\mathrm{ML}}\mathbf{\Sigma}^{-1}(\beta)|\prod_{t=1}^{T}\left[\frac{\boldsymbol{z}_t^H\mathbf{\Sigma}^{-1}(\beta)\boldsymbol{z}_t}{\boldsymbol{z}_t^H\mathbf{\Sigma}_{\mathrm{ML}}^{-1}\boldsymbol{z}_t}\right]^{-M/T} = \mathrm{med}\left[\omega(LR|M,T)\right] \quad (20)$$

where $\omega(LR|M,T)$ is the true p.d.f. of $LR_{ACG}^{1/T}\left(\mathbf{\Sigma}_0|\boldsymbol{Z}_T\right)$, $\mathbf{\Sigma}_{\mathrm{ML}}$ is the complex Tyler's M-estimate (12) and med $[\omega(LR|M,T)]$ stands for the median value. Since strict equality may be difficult to obtain, in practice we evaluate $LR_{ACG}^{1/T}\left(\mathbf{\Sigma}(\beta)|\boldsymbol{Z}_T\right)$ on a grid of values for β and pick the one which results in the LR closest to the median.

We now study whether or not the choice of the loading factor from the EL principle is optimal. In order to evaluate the quality of a CME $\hat{\mathbf{\Sigma}}$, we use the SNR loss

$$SNR_{\mathrm{loss}} = \frac{\left(\boldsymbol{s}_0^H\mathbf{\Sigma}_0^{-1}\boldsymbol{s}_0\right)^2}{\left(\boldsymbol{s}_0^H\hat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}_0\hat{\mathbf{\Sigma}}^{-1}\boldsymbol{s}_0\right)\left(\boldsymbol{s}_0^H\mathbf{\Sigma}_0^{-1}\boldsymbol{s}_0\right)} \quad (21)$$

where $\boldsymbol{s}_0 = \begin{bmatrix}1 & e^{i\pi\sin\theta_0} & \cdots & e^{i\pi(M-1)\sin\theta_0}\end{bmatrix}^T$ stands for the steering vector corresponding to the looked direction θ_0. We consider here the estimate based on shrinkage of the NSCM

$$\hat{\mathbf{\Sigma}} = (1-\beta)\frac{M}{T}\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}_t^H + \beta\boldsymbol{I}_M \quad (22)$$

(referred to as DL in the figures) and its fixed-point version in (17)-(18), referred to as FP-DL in the figures. We compare both of them to the oracle estimator and to the conventional MLE in Figure 3. Interestingly enough, it appears that the oracle loading factor β_0 in (19) results in a matrix $\mathbf{\Sigma}(\beta)$ in (17) whose LR closely matches that of $\mathbf{\Sigma}_0$. As a result, the SNR loss achieved by the oracle estimate is very high. More interesting is the fact that the EL approach yields the same LR value as the oracle estimate. Moreover, the EL and the oracle estimate yields the same output SNR. This shows that selection of β from the EL principle is as good as the oracle choice. Finally, observe that the fixed point diagonal loading performs much better than MLE, especially in low sample support.

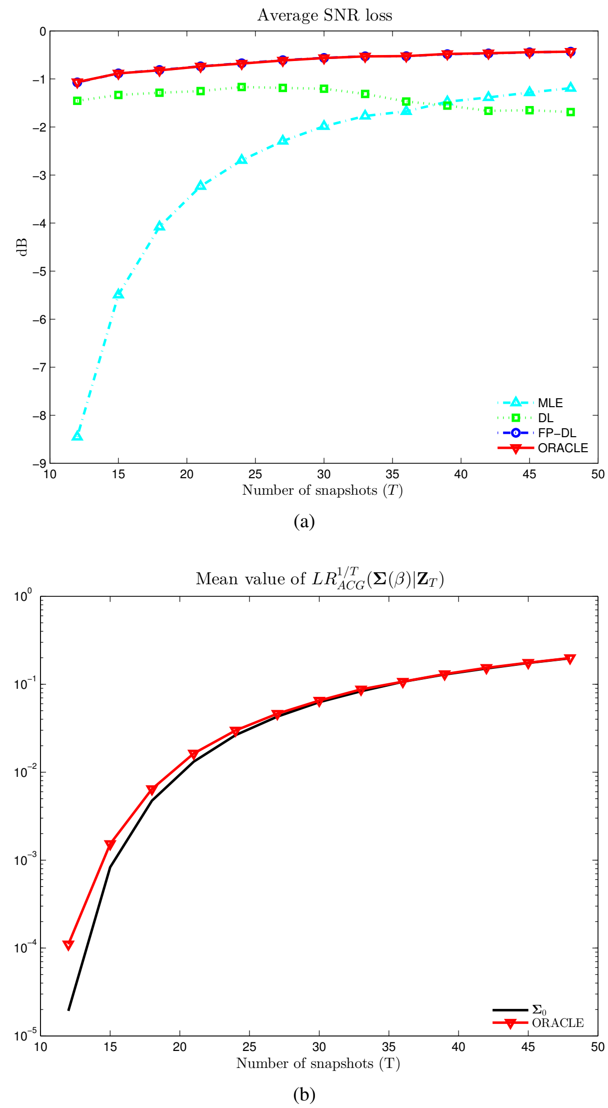

Fig. 3. Performance of diagonally loaded estimates versus number of snapshots T. $M = 12$ and $d = 1$. (a) SNR loss (b) Mean value of $LR_{ACG}^{1/T}\left(\mathbf{\Sigma}(\beta)|\boldsymbol{Z}_T\right)$.

4. SUMMARY AND CONCLUSIONS

In this paper, we extended the expected likelihood methodology introduced in [2, 3] to the class of complex elliptically symmetric distributions and complex angular central Gaussian distributions. We demonstrated that for the true (a priori unknown) scatter matrix $\mathbf{\Sigma}_0$, the p.d.f. of the likelihood ratio does not depend on this matrix, only on the density generator $g(.)$, the sample volume T and matrix dimension M. For ACG distributions, this p.d.f. is fully specified by T and M only. This paved the way to regularized scatter matrix estimation schemes where the regularization parameters are chosen so that the associated LR complies with the support of the p.d.f. of $LR(\mathbf{\Sigma}_0)$.

5. REFERENCES

[1] Y. I. Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum SNR," *Radio Engineering and Electronic Physics*, vol. 26, pp. 87–95, March 1981.

[2] Y.I. Abramovich, N.K. Spencer, and A.Y. Gorokhov, "Bounds on maximum likelihood ratio-Part I: Application to antenna array detection-estimation with perfect wavefront coherence," *IEEE Transactions Signal Processing*, vol. 52, no. 6, pp. 1524–1536, June 2004.

[3] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, "Modified GLRT and AMF framework for adaptive detectors," *IEEE Transactions Aerospace Electronic Systems*, vol. 43, no. 3, pp. 1017–1051, July 2007.

[4] 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 Transactions Signal Processing*, vol. 55, no. 1, pp. 20–31, January 2007.

[5] E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant process," *IEE Proceedings Radar, Sonar and Navigation*, vol. 134, no. 2, pp. 191–197, April 1987.

[6] E. Ollila, D. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions: survey, new results and applications," *IEEE Transactions Signal Processing*, vol. 60, no. 11, pp. 5597–5625, November 2012.

[7] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Transactions Signal Processing*, vol. 59, no. 9, pp. 4097–4107, September 2011.

[8] A. Wiesel, "Unified framework to regularized covariance estimation in scaled Gaussian models," *IEEE Transactions Signal Processing*, vol. 60, no. 1, pp. 29–38, January 2012.

[9] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, March 1987.

[10] J. T. Kent and D. E. Tyler, "Redescending M-Estimates of multivariate location and scatter," *The Annals of Statistics*, vol. 19, no. 4, pp. 2102–2119, December 1991.

[11] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley, 1982.

[12] D. E. Tyler, "Statistical analysis for the angular central Gaussian distribution on the sphere," *Biometrika*, vol. 74, no. 3, pp. 579–589, September 1987.

[13] Y. I. Abramovich, N. K. Spencer, and M. D. E. Turley, "Time-varying autoregressive (TVAR) models for multiple radar observations," *IEEE Transactions Signal Processing*, vol. 55, no. 4, pp. 1298–1311, April 2007.

[14] Y. I. Abramovich and N. K. Spencer, "Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering," in *Proceedings ICASSP*, Honolulu, HI, April 2007, pp. 1105–1108.