Regularized Covariance Matrix Estimation in Complex Elliptically Symmetric Distributions Using the Expected Likelihood Approach - Part 1: The Over-Sampled Case

Yuri Abramovich, Olivier Besson

To link to this article: DOI: 10.1109/TSP.2013.2272924
URL: http://dx.doi.org/10.1109/TSP.2013.2272924

To cite this version: Abramovich, Yuri and Besson, Olivier *Regularized Covariance Matrix Estimation in Complex Elliptically Symmetric Distributions Using the Expected Likelihood Approach - Part 1: The Over-Sampled Case.* (2013) IEEE Transactions on Signal Processing, vol. 61 (n° 23). pp. 5807-5818. ISSN 1053-587X

Regularized Covariance Matrix Estimation in Complex Elliptically Symmetric Distributions Using the Expected Likelihood Approach— Part 1: The Over-Sampled Case

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

***Abstract*—In Abramovich *et al.* ["Bounds on Maximum Likelihood Ratio—Part I: Application to Antenna Array Detection-Estimation With Perfect Wavefront Coherence," *IEEE Trans. Signal Process.*, vol. 52, pp. 1524–1536, June 2004], it was demonstrated, for multivariate complex Gaussian distribution, that the probability density function (p.d.f.) of the likelihood ratio (LR) for the (unknown) actual covariance matrix $\boldsymbol{R}_0$ does not depend on this matrix and is fully specified by the matrix dimension M and the number of independent training samples T. This invariance property hence enables one to compare the LR of any derived covariance matrix estimate against this p.d.f., and eventually get an estimate that is statistically "as likely" as $\boldsymbol{R}_0$. This "expected likelihood" quality assessment allowed significant improvement of MUSIC DOA estimation performance in the so-called "threshold area," and for diagonal loading and TVAR model order selection in adaptive detectors. Recently, the so-called complex elliptically symmetric (CES) distributions have been introduced for description of highly in-homogeneous clutter returns. The aim of this series of two papers is to extend the EL approach to this class of CES distributions as well as to a particularly important derivative, namely the complex angular central distribution (ACG). For both cases, we demonstrate a similar invariance property for the LR associated with the true scatter matrix $\boldsymbol{\Sigma}_0$. Furthermore, we derive fixed point regularized covariance matrix estimates using the generalized expected likelihood methodology. This first part is devoted to the conventional scenario ($T \geq M$) while Part II deals with the undersampled scenario ($T \leq M$).**

***Index Terms*—Covariance matrix estimation, elliptically symmetric distributions, expected likelihood, likelihood ratio, regularization.**

I. Introduction

In a large number of radar applications, the traditional assumption on training data being a set of independent identically distributed (i.i.d) complex Gaussian random samples is strongly violated due to a significant in-homogeneity of this data. Examples from airborne moving target indicator or ship-borne radars with strongly in-homogeneous clutter are well-known [1]. For high-frequency over-the-horizon radars, and specifically for mode-selective multiple input multiple output (MIMO) radars, similar scenario takes place when adaptive MIMO beamformers are trained using Doppler-processed training data [2], [3]. If ignored, significant non-homogeneity of training data has an adverse effect on adaptive processing since it significantly reduces the effective number of training data and more generally, makes the Gaussian model-based inference inaccurate. In most studies, such in-homogeneous set of data is modeled as a set of spherically invariant random vectors (SIRV) [4]–[6]. A SIRV can be viewed as a special case of a broader class, complex elliptically symmetric (CES) distributions which are considered in the sequel. While this model describes in-homogeneous clutter, in-discriminatory application of this model that ignores additive white Gaussian noise, may lead to a number of problems, as demonstrated in [7]. In other words, this approach is suitable in "clutter-limited" applications, where the clutter-only covariance matrix is a full-rank matrix with the minimal eigenvalue that significantly exceeds the additive white noise power. In such a case, the latter may be ignored and the training data that contains energetic clutter may be described as a set of i.i.d SIRV or CES data.

Yet, so far, the Gaussian assumption has been predominating and much attention has focused on the problem of maximum likelihood (ML) covariance matrix estimation and, more generically, on adaptive detection based on ML principles in the Gaussian case. Within this framework, it was demonstrated that for a limited number of i.i.d training data T, a number of adaptive detection-estimation techniques properties, derived under the $(T \rightarrow \infty, M = \mathrm{const.})$ for ML principle asymptotic condition, are not true. Typical example is provided by MUSIC direction of arrival (DOA) estimation technique proven to be asymptotically efficient [8], [9]. However, as demonstrated in [10], for a certain small enough sample support T MUSIC "breaks down" i.e., it starts to generate severely erroneous DOA estimates. Another well-known problem is a relatively poor performance of adaptive filters (antennas) and adaptive detectors that adopt the ML covariance matrix estimate under a limited sample support [11]. It has been evidenced in various studies that regularization ("shrinkage") of the covariance matrix estimate, such as diagonal loading [13], [14] can significantly improve detection performance, if the shrinkage parameters are properly chosen. To address these and similar issues that occur under small sample support, in [10]–[12]

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

O. Besson is with the ISAE, Department Electronics Optronics Signal, University of Toulouse, 31055 Toulouse, France (e-mail: olivier.besson@isae.fr).

the technique called "Expected Likelihood" (EL) has been proposed. This technique is based on the invariance of the likelihood ratio (LR), constructed for the multivariate complex Gaussian data. More specifically, it uses the fact that the p.d.f. of $\mathrm{LR}(\boldsymbol{R}_0)$ (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 the i.i.d sample volume T. This invariance makes it possible to evaluate the "quality" of any (possibly parametric or regularized) covariance matrix estimate $\boldsymbol{R}(\hat{\boldsymbol{\Omega}})$ by comparing its likelihood ratio $\mathrm{LR}(\boldsymbol{R}(\hat{\boldsymbol{\Omega}}))$ against the p.d.f. for $\mathrm{LR}(\boldsymbol{R}_0)$. The estimate $\boldsymbol{R}(\hat{\boldsymbol{\Omega}})$ is then treated as appropriate if $LR(\boldsymbol{R}(\hat{\boldsymbol{\Omega}}))$ is within the support of $\mathrm{LR}(\boldsymbol{R}_0)$ p.d.f., pre-calculated for given M and T. In other words, if $\boldsymbol{R}(\hat{\boldsymbol{\Omega}})$ is statistically as likely as $\boldsymbol{R}_0$, the EL approach deems it properly regularized. Recall that the unrestricted ML covariance matrix estimate produces the ultimate equal to one LR value irrespective of sample support $(T \geq M)$, while the LR value generated by the true covariance matrix $\mathrm{LR}(\boldsymbol{R}_0)$ is significantly smaller for realistic $(T \gtrsim (1\text{–}5)M)$ sample support volumes [10], [11]. The EL approach was shown to be effective in identifying "broken" MUSIC-produced DOA estimates ("breakdown prediction) and rectifying the set of these estimates to meet the expected likelihood ratio values ("breakdown cure") [10], [12]. Accordingly, its ability to improve adaptive filters has been proved in [11].

Obviously, this EL methodology could be quite useful in addressing similar problems when dealing with non-Gaussian data. For this reason, the extension of the EL principles over the broader class of complex elliptically symmetric (CES) multivariate random variables constitutes the focus of this study. CES distributions are parameterized by the scatter matrix $\boldsymbol{\Sigma}$ and a one-dimensional function $g(t)$ called the density generator [15]. Since the latter is usually unknown in practice, we also consider complex angular central Gaussian (ACG) distributions which depend on the scatter matrix $\boldsymbol{\Sigma}$ only.

The paper is organized as follows. In Section II we introduce the discussed above likelihood ratios $\mathrm{LR}_{\mathrm{CES}}(\boldsymbol{\Sigma})$ and $\mathrm{LR}_{\mathrm{ACG}}(\boldsymbol{\Sigma})$ for conventional $(T \geq M)$ training conditions and derive their respective invariance properties. In Section III we derive the fixed point ML $\mathrm{TVAR}(m)$ covariance matrix estimate, while in Section IV we discuss the application of the EL methodology to selection of the loading factor and $\mathrm{TVAR}(m)$ order m in diagonally loaded and $\mathrm{TVAR}(m)$ covariance matrix estimates. In Section V we present the results of Monte-Carlo simulations that demonstrate significant superiority of the regularized fixed point estimates with respect to the unconstrained (fixed point) ML estimates for adaptive filters (antennas) applications. The summary and conclusions are given in Section VI.

II. Likelihood Ratio and Its Invariance for Data With Complex Elliptically Symmetric Distribution

A. Complex Elliptically Symmetric Distributions

Description of CES distributions and their properties can be found e.g., in [16]–[19]. A very comprehensive review along with application of CES distributions to a number of array processing problems can be found in the recent paper [15]. We refer the reader to this paper for details that could be skipped in the short review to be presented now and which is inspired by the presentation in [15]. Herein we consider the special absolutely continuous case with zero mean, when the p.d.f. of the r.v. $\boldsymbol{x} \in \mathbb{C}^M$ is of the form

$$f(\boldsymbol{x}|\boldsymbol{\Sigma}_0, g) = C_{M,g}|\boldsymbol{\Sigma}_0|^{-1} g\left(\boldsymbol{x}^H \boldsymbol{\Sigma}_0^{-1} \boldsymbol{x}\right) \tag{1}$$

for a positive definite Hermitian (PDH) $M \times M$ matrix $\boldsymbol{\Sigma}_0$ called the scatter matrix, and function $g : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ called density generator that satisfies finite moment condition $\delta_{M,g} = \int_0^\infty t^{M-1} g(t) dt < \infty$ to ensure integrability of $f(\boldsymbol{x}|\boldsymbol{\Sigma}_0, g)$. Above $C_{M,g}$ is a normalization constant ensuring that $f(\boldsymbol{x})$ integrates to 1 and is given by $C_{M,g} \triangleq 2(S_M \delta_{M,g})^{-1}$ where $S_M = 2\pi^M/\Gamma(M)$ is the surface area of the unit complex M-sphere $\mathbb{C}S^M = \{\boldsymbol{z} \in \mathbb{C}^M; \|\boldsymbol{z}\| = 1\}$. We adopt the following notation in the following $\boldsymbol{x} \sim \mathbb{C}\mathcal{E}_M(\boldsymbol{0}, \boldsymbol{\Sigma}_0, g)$. Some important properties of CES distributions will be of use in the sequel [15]–[19]. First, $\boldsymbol{x}$ admits the following stochastic representation

$$\boldsymbol{x} \overset{d}{=} \mathcal{R} \boldsymbol{\Sigma}_0^{1/2} \boldsymbol{u} \tag{2}$$

where the non-negative real random variable $\mathcal{R} \triangleq \sqrt{\mathcal{Q}}$, called the modular variate, is independent of the complex random vector $\boldsymbol{u}$ possessing a uniform distribution on $\mathbb{C}S^M$ denoted as $\boldsymbol{u} \sim \mathcal{U}(\mathbb{C}S^M)$. Here, $\overset{d}{=}$ means "has the same distribution as". Second, the p.d.f. $f(\mathcal{Q})$ of the modular variate $\mathcal{Q}$ is given by

$$f(\mathcal{Q}) = \delta_{M,g}^{-1} \mathcal{Q}^{M-1} g(\mathcal{Q}) \tag{3}$$

which is also the p.d.f. of the Hermitian form $\boldsymbol{x}^H \boldsymbol{\Sigma}_0^{-1} \boldsymbol{x}$ as follows from (2). The complex normal distribution $\mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_0)$ is obtained for the particular $g(t) = \exp\{-t\}$, yielding $C_{M,g} = \pi^{-M}$ as the normalizing constant. Note that if $\mathcal{E}\{\mathcal{R}^p\} < \infty$, $\boldsymbol{x}$ admits finite p-th order moments, $\mathcal{E}\{\boldsymbol{x}\} = \boldsymbol{0}$ and $\mathcal{E}\{\boldsymbol{x}\boldsymbol{x}^H\} = \frac{\mathcal{E}\{\mathcal{R}^2\}}{M} \boldsymbol{\Sigma}_0$. Thus, the scatter matrix $\boldsymbol{\Sigma}_0$ is proportional to the covariance matrix under finite 2nd-order moment assumption.

Given (1), for a set of T i.i.d r.v. $\boldsymbol{x}_t \sim \mathbb{C}\mathcal{E}_M(\boldsymbol{0}, \boldsymbol{\Sigma}_0, g)$, we get for $\boldsymbol{X}_T = [\boldsymbol{x}_1 \quad \cdots \quad \boldsymbol{x}_T] \in \mathbb{C}^{M \times T}$

$$f(\boldsymbol{X}_T|\boldsymbol{\Sigma}_0, g) = C_{M,g}^T |\boldsymbol{\Sigma}_0|^{-T} \prod_{t=1}^T g\left({\boldsymbol{x}_t}^H \boldsymbol{\Sigma}_0^{-1} \boldsymbol{x}_t\right). \tag{4}$$

For $T \geq M$, the maximum likelihood estimator (MLE) of the scatter matrix $\boldsymbol{\Sigma}_0$ is the matrix $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ that minimizes over the set of PDH matrices the negative log-likelihood function:

$$L_T(\boldsymbol{\Sigma}) \triangleq T \log|\boldsymbol{\Sigma}| - \sum_{t=1}^T \log g\left({\boldsymbol{x}_t}^H \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_t\right) \tag{5}$$

and hence is the solution (assuming that $g(\cdot)$ is continuously differentiable) to the estimating equation [15], [20], [21]

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}} = \mathcal{T}(\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}) \tag{6a}$$

$$\mathcal{T}(\boldsymbol{\Sigma}) = \frac{1}{T} \sum_{t=1}^T \phi\left({\boldsymbol{x}_t}^H \boldsymbol{\Sigma}^{-1} \boldsymbol{x}_t\right) \boldsymbol{x}_t {\boldsymbol{x}_t}^H \tag{6b}$$

where $\phi(t) \triangleq -g'(t)/g(t)$. For $\mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_0)$ where $g(t) = \exp\{-t\}$, $\phi(t) = 1$ and from (6) follows the well-known sample covariance matrix estimate. In general case, where the weight function $\phi(\cdot)$ is not a constant, the estimation equation is implicit and an algorithm to find its solution is needed. In [[15], Theorem 6] based on the results of Kent and Tyler [22], [23] for the real case, the uniqueness and convergence of the fixed point iterations $(\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}})_{k+1} = \mathcal{T}[(\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}})_k]$ to the unique solution of (6), for any initial estimate of $\boldsymbol{\Sigma}_0$, has been proven under certain technical conditions on $\phi(t)$, see [15] for further details.

For $\mathbb{C}\mathcal{E}_M(\mathbf{0}, \boldsymbol{\Sigma}_0, g)$ distributions that meet these conditions, let us consider the likelihood ratio for any 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. This $\mathrm{LR}_{\mathrm{CES}}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{X}_T)$ may be found as usual [24]:

$$\mathrm{LR}_{\mathrm{CES}}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{X}_T, g) = \frac{f(\boldsymbol{X}_T|\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell), g)}{\max_{\boldsymbol{\Sigma}} f(\boldsymbol{X}_T|\boldsymbol{\Sigma}, g)}. \tag{7}$$

From (4), we get

$$\mathrm{LR}_{\mathrm{CES}}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{X}_T, g) = |\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)|^T \times \prod_{t=1}^{T} \frac{g\left(\boldsymbol{x}_t{}^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)\boldsymbol{x}_t\right)}{g\left(\boldsymbol{x}_t{}^H\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1}\boldsymbol{x}_t\right)}. \tag{8}$$

With respect to (2) the "expected likelihood", i.e., the LR value for the actual (true) scatter matrix $\boldsymbol{\Sigma}_0$ may be presented as

$$\mathrm{LR}_{\mathrm{CES}}(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g) \overset{d}{=} |\boldsymbol{A}|^T \prod_{t=1}^{T} \frac{g\left(\mathcal{Q}_t\boldsymbol{u}_t{}^H\boldsymbol{u}_t\right)}{g\left(\mathcal{Q}_t\boldsymbol{u}_t{}^H\boldsymbol{A}^{-1}\boldsymbol{u}_t\right)} \tag{9}$$

where $\boldsymbol{u}_t \sim \mathcal{U}(\mathbb{C}S^M)$ and $\boldsymbol{A} \triangleq \boldsymbol{\Sigma}_0^{-1/2}\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}\boldsymbol{\Sigma}_0^{-1/2}$. Now, due to the invariance of the MLE under non singular data transformations, $\boldsymbol{A}$ is the MLE of the scatter matrix from a $\mathbb{C}\mathcal{E}_M(\mathbf{0}, \boldsymbol{I}, g)$ distribution. Consequently, $\boldsymbol{A}$ does not depend on $\boldsymbol{\Sigma}_0$, only on $g(\cdot)$. This could also be seen by pre and post-multiplying (6) by $\boldsymbol{\Sigma}_0^{-1/2}$ and using (2) to get

$$\boldsymbol{A} = \frac{1}{T}\sum_{t=1}^{T} \phi\left(\mathcal{Q}_t\boldsymbol{u}_t{}^H\boldsymbol{A}^{-1}\boldsymbol{u}_t\right)\mathcal{Q}_t\boldsymbol{u}_t\boldsymbol{u}_t{}^H. \tag{10}$$

Therefore, the p.d.f. of $\mathrm{LR}_{\mathrm{CES}}(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g)$ is invariant with respect to (w.r.t) the true scatter matrix $\boldsymbol{\Sigma}_0$, and is explicitly specified by $f(\mathcal{Q})$ in (3) and parameters M and T.

B. Angular Central Gaussian Distribution

For all cases where $f(\mathcal{Q})$ is accurately known a priori and only the scatter matrix (or its parameters) is to be estimated, the EL principle can be applied since the p.d.f. for $\mathrm{LR}_{\mathrm{CES}}(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g)$ could be pre-calculated for the given $f(\mathcal{Q}), M, T$, using Monte-Carlo simulations at least. As discussed in the introduction, in many cases the distribution $f(\mathcal{Q})$ is not known a priori, and hence $\sqrt{\mathcal{Q}_t}$, $t = 1, \ldots, T$ are often treated as unknown deterministic parameters. For unknown $f(\mathcal{Q})$, the input vectors $\boldsymbol{x}_t$ are often being transformed to the set of normalized vectors

$$\boldsymbol{z}_t = \frac{\boldsymbol{x}_t}{\|\boldsymbol{x}_t\|_2}. \tag{11}$$

If $\boldsymbol{x} \sim \mathbb{C}\mathcal{E}_M(\mathbf{0}, \boldsymbol{\Sigma}_0, g)$, then the distribution of its projection onto the unit complex M-sphere $\boldsymbol{x}/\|\boldsymbol{x}\|_2$ is said to have a complex angular elliptic distribution. In particular, if the CES distribution is a central complex Gaussian, i.e., $\boldsymbol{x} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$ then the distribution of $\boldsymbol{z} = \boldsymbol{x}/\|\boldsymbol{x}\|_2$ is said to have a complex angular central Gaussian (ACG) distribution, which we denote as $\boldsymbol{z} \sim \mathbb{C}\mathcal{AG}(\mathbf{0}, \boldsymbol{\Sigma}_0)$. For non-singular $\boldsymbol{\Sigma}_0$, the p.d.f. of $\boldsymbol{z}$ is given by [15], [25]

$$f(\boldsymbol{z}|\boldsymbol{\Sigma}_0) = S_M^{-1}|\boldsymbol{\Sigma}_0|^{-1}\left(\boldsymbol{z}^H\boldsymbol{\Sigma}_0^{-1}\boldsymbol{z}\right)^{-M}. \tag{12}$$

Note that the matrix $\boldsymbol{\Sigma}_0$ can be only identified up to a scale, since $\boldsymbol{\Sigma}_0$ and $c\boldsymbol{\Sigma}_0$ yield the same distribution for any $c > 0$. Note also that for a central (zero mean) case, the central Gaussian distribution for $\boldsymbol{x}$ could be replaced by any central CES distribution and the resulting angular distribution would be the same. That is, if $\boldsymbol{x} \sim \mathbb{C}\mathcal{E}_M(\mathbf{0}, \boldsymbol{\Sigma}_0, g)$ then $\boldsymbol{x}/\|\boldsymbol{x}\|_2 \sim \mathbb{C}\mathcal{AG}(\mathbf{0}, \boldsymbol{\Sigma}_0)$. Note that although the density in (12) looks like the generic density of a complex elliptical distribution in (1), it does not have a CES distribution itself and does not possess the characterizing stochastic representation (2) [15]. Yet, the non-singular ACG distribution can be generated using the r.v. $\boldsymbol{u} \sim \mathcal{U}(\mathbb{C}S^M)$ as $\boldsymbol{Bu}/\|\boldsymbol{Bu}\| \sim \mathbb{C}\mathcal{AG}(\mathbf{0}, \boldsymbol{\Sigma}_0)$ for $\boldsymbol{\Sigma}_0 = \boldsymbol{B}\boldsymbol{B}^H$ and non-singular $\boldsymbol{B}$.

Assuming independence of the $\boldsymbol{z}_t$, the joint distribution of $\boldsymbol{Z}_T = [\,\boldsymbol{z}_1 \quad \cdots \quad \boldsymbol{z}_T\,]$ is thus given by

$$f(\boldsymbol{Z}_T|\boldsymbol{\Sigma}_0) = S_M^{-T}|\boldsymbol{\Sigma}_0|^{-T}\prod_{t=1}^{T}\left(\boldsymbol{z}_t^H\boldsymbol{\Sigma}_0^{-1}\boldsymbol{z}_t\right)^{-M}. \tag{13}$$

In [25] it was demonstrated that the MLE for $\boldsymbol{\Sigma}_0$ in this case still corresponds to a solution to (6) with the weight function being simply $\phi_{\mathrm{ACG}}(t) = \frac{M}{T}$. In other words, $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ in the ACG case satisfies

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}} = \frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1}\boldsymbol{z}_t}. \tag{14}$$

Moreover in [26] it was demonstrated that the estimate (14) being the ML estimate of $\boldsymbol{\Sigma}_0$ under assumption $\boldsymbol{x}_t \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \tau_t\boldsymbol{\Sigma}_0)$ for $t = 1, \ldots, T$ is also the ML estimate for a more general case when $\boldsymbol{x}_t \sim \mathbb{C}\mathcal{E}_M(\mathbf{0}, \tau_t\boldsymbol{\Sigma}_0, g_t)$ with the functions g_t being given but not necessarily the same. Clearly this quite a universal property of the complex Tyler's M-estimator, along with the invariance of the likelihood ratio $\mathrm{LR}_{\mathrm{ACG}}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)$ (see below) makes this estimate very attractive. Note that with respect to (14), the fixed point iterations

$$\hat{\boldsymbol{\Sigma}}_{k+1} = \frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\hat{\boldsymbol{\Sigma}}_k^{-1}\boldsymbol{z}_t} \tag{15}$$

converge to $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ which exists and is unique up to a positive scalar [23], [27]–[29]. For uniqueness, one may want to restrict $\boldsymbol{\Sigma}_0$ in a suitable way, e.g., by assuming $\mathrm{Tr}\{\boldsymbol{\Sigma}_0\} = M$ (or $|\boldsymbol{\Sigma}_0| = 1$).

For a (possibly parameterized) scatter matrix $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$, the likelihood ratio in the ACG case is given by

$$\begin{aligned}\mathrm{LR}_{\mathrm{ACG}}&(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T)\\ &= \frac{f(\boldsymbol{Z}_T|\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell))}{\max_{\boldsymbol{\Sigma}} f(\boldsymbol{Z}_T|\boldsymbol{\Sigma})}\\ &= |\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)|^T\prod_{t=1}^{T}\left[\frac{\boldsymbol{z}_t^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)\boldsymbol{z}_t}{\boldsymbol{z}_t^H\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1}\boldsymbol{z}_t}\right]^{-M}.\end{aligned} \tag{16}$$

We can now specify $\mathrm{LR}_{\mathrm{ACG}}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)$. Since $\boldsymbol{z}_t = \frac{\boldsymbol{x}_t}{\|\boldsymbol{x}_t\|_2} \overset{d}{=} \frac{\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{u}_t}{\|\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{u}_t\|_2}$ where $\boldsymbol{u}_t \sim \mathcal{U}(\mathbb{C}S^M)$ or $\boldsymbol{u}_t \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, it follows that

$$\begin{aligned}\mathrm{LR}_{\mathrm{ACG}}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T) &= \left|\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}\boldsymbol{\Sigma}_0^{-1}\right|^T \prod_{t=1}^{T}\left[\frac{\boldsymbol{z}_t^H\boldsymbol{\Sigma}_0^{-1}\boldsymbol{z}_t}{\boldsymbol{z}_t^H\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1}\boldsymbol{z}_t}\right]^{-M} \\ &\overset{d}{=} |\boldsymbol{A}_{\mathrm{ACG}}|^T \prod_{t=1}^{T}\left[\frac{\boldsymbol{u}_t{}^H\boldsymbol{u}_t}{\boldsymbol{u}_t{}^H\boldsymbol{A}_{\mathrm{ACG}}^{-1}\boldsymbol{u}_t}\right]^{-M} \qquad (17)\end{aligned}$$

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

$$\begin{aligned}\boldsymbol{A}_{\mathrm{ACG}} &= \frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{\Sigma}_0^{-1/2}\boldsymbol{z}_t\boldsymbol{z}_t^H\boldsymbol{\Sigma}_0^{-1/2}}{\boldsymbol{z}_t^H\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1}\boldsymbol{z}_t} \\ &\overset{d}{=} \frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{u}_t\boldsymbol{u}_t{}^H}{\boldsymbol{u}_t{}^H\boldsymbol{A}_{\mathrm{ACG}}^{-1}\boldsymbol{u}_t}. \qquad (18)\end{aligned}$$

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

III. ML TVAR(m) Covariance Matrix Estimation for Complex Angular Central Gaussian Distribution

Let us consider a set of T i.i.d M-variate complex angular central Gaussian vectors $\boldsymbol{Z}_T = [\,\boldsymbol{z}_1 \ \cdots \ \boldsymbol{z}_T\,]$ generated by an arbitrary complex central elliptical distribution. Let $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$ be an identified scatter matrix parameterized by a set of parameters $\boldsymbol{\Omega}_\ell$. Then the likelihood function (LF) can be introduced as follows

$$\begin{aligned}&\mathrm{LF}_{\mathrm{ACG}}[\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T] \\ &\quad = S_M^{-T}|\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)|^T\prod_{t=1}^{T}\left[\boldsymbol{z}_t^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)\boldsymbol{z}_t\right]^{-M}. \qquad (19)\end{aligned}$$

For a TVAR(m) model ($m = 0, \ldots, M-2$), we have to find the maximum of this LF over the class of structured positive definite (p.d.) Hermitian matrices with $[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)]_{i,j} = 0$ for$|i-j| > m$ which according to [30], is the only necessary condition for a p.d. matrix $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$ to serve as the scatter matrix of a TVAR(m) process. Let $\boldsymbol{C}^{(m)} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)$ with $\boldsymbol{C}_{i,j}^{(m)} = 0$ for $|i-j| > m$. Then, up to an additive constant,

$$\begin{aligned}&\log \mathrm{LF}_{\mathrm{ACG}}\left[\boldsymbol{C}^{(m)}|\boldsymbol{Z}_T\right] \\ &\quad = T\log|\boldsymbol{C}^{(m)}| - M\sum_{t=1}^{T}\log\left(\boldsymbol{z}_t^H\boldsymbol{C}^{(m)}\boldsymbol{z}_t\right). \qquad (20)\end{aligned}$$

Since only $\boldsymbol{C}_{i,j}^{(m)}$, $|i-j| \leq m$ are subject to optimization, the ML equation may be presented as

$$\frac{\partial \log \mathrm{LF}_{\mathrm{ACG}}\left[\boldsymbol{C}^{(m)}|\boldsymbol{Z}_T\right]}{\partial \boldsymbol{C}_{i,j}^{(m)}} = 0 \quad \text{for } |i-j| \leq m \qquad (21a)$$

$$\boldsymbol{C}_{i,j}^{(m)} = 0 \quad \text{for } |i-j| > m. \qquad (21b)$$

Using the fact that [31]

$$\frac{\partial \log|\boldsymbol{C}^{(m)}|}{\partial \boldsymbol{C}_{i,j}^{(m)}} = \left\{\left[\boldsymbol{C}^{(m)}\right]^{-1}\right\}_{i,j} \qquad (22a)$$

$$\frac{\partial \log \boldsymbol{z}_t^H\boldsymbol{C}^{(m)}\boldsymbol{z}_t}{\partial \boldsymbol{C}_{i,j}^{(m)}} = \left\{\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\boldsymbol{C}^{(m)}\boldsymbol{z}_t}\right\}_{i,j} \qquad (22b)$$

it follows that the MLE $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{(m)}$ of $\boldsymbol{\Sigma}_0$ in the TVAR(m) model satisfies

$$\left[\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{(m)}\right]_{i,j} = \left[\frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\left(\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{(m)}\right)^{-1}\boldsymbol{z}_t}\right]_{i,j} \quad |i-j| \leq m \qquad (23a)$$

$$\left\{\left[\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{(m)}\right]^{-1}\right\}_{i,j} = 0 \quad |i-j| > m. \qquad (23b)$$

The latter means that the ML TVAR(m) estimate of the scatter matrix satisfies the estimation equation

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{(m)} = \mathsf{DG}^{(m)}\left[\frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\left(\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{(m)}\right)^{-1}\boldsymbol{z}_t}\right] \qquad (24)$$

where $\mathsf{DG}^{(m)}[\cdot]$ is the Dym-Gohberg band-inverse transformation of a Hermitian non negative definite matrix, defined as [30]

$$\left\{\mathsf{DG}^{(m)}[\boldsymbol{R}]\right\}_{i,j} = \boldsymbol{R}_{i,j} \quad |i-j| \leq m \qquad (25a)$$

$$\left\{\left(\mathsf{DG}^{(m)}[\boldsymbol{R}]\right)^{-1}\right\}_{i,j} = 0 \quad |i-j| > m. \qquad (25b)$$

Note that $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{(m)}$ is invariant to scaling since $\mathsf{DG}^{(m)}[c\boldsymbol{R}] = c\mathsf{DG}^{(m)}[\boldsymbol{R}]$. In order to obtain $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{(m)}$ in (24), we propose to resort to the fixed point iterations

$$\check{\boldsymbol{\Sigma}}_{k+1}^{(m)} = \mathsf{DG}^{(m)}\left[\frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\left(\hat{\boldsymbol{\Sigma}}_{k}^{(m)}\right)^{-1}\boldsymbol{z}_t}\right] \qquad (26a)$$

$$\hat{\boldsymbol{\Sigma}}_{k+1}^{(m)} = \frac{M}{\mathrm{Tr}\left\{\check{\boldsymbol{\Sigma}}_{k+1}^{(m)}\right\}}\check{\boldsymbol{\Sigma}}_{k+1}^{(m)}. \qquad (26b)$$

At this stage, we were unable to prove convergence of the iterative scheme (26) to a unique solution: therefore, this is still an open issue to be solved.

IV. Application of the Expected Likelihood Approach for Scatter Matrix Estimation

The unrestricted (unstructured) MLE Tyler's M-estimator (fixed point solution) for $T \geq M$ provides the globally optimal solution that yields the ultimate value $\mathrm{LR}_{\mathrm{ACG}}(\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}|\boldsymbol{Z}_T) = 1 \gg \mathrm{LR}_{\mathrm{ACG}}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)$. Hence, even for conventional $(T \geq M)$ training conditions this estimate, $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ may not be that effective for adaptive processing applications. For this reason, initially in [32] and then in [33], [34] the "shrinkage" fixed point (diagonally loaded)

$\hat{\mathbf{\Sigma}}(\beta) = \lim_{k\to\infty} \hat{\mathbf{\Sigma}}_k(\beta)$ estimator has been proposed, where $\hat{\mathbf{\Sigma}}_k(\beta)$ is obtained from the following iterative procedure:

$$\breve{\mathbf{\Sigma}}_{k+1}(\beta) = (1-\beta)\frac{M}{T}\sum_{t=1}^{T}\frac{\mathbf{z}_t\mathbf{z}_t^H}{\mathbf{z}_t^H(\hat{\mathbf{\Sigma}}_k(\beta))^{-1}\mathbf{z}_t} + \beta\mathbf{I}_M \tag{27a}$$

$$\hat{\mathbf{\Sigma}}_{k+1}(\beta) = \frac{M}{\mathrm{Tr}\{\breve{\mathbf{\Sigma}}_{k+1}(\beta)\}}\breve{\mathbf{\Sigma}}_{k+1}(\beta). \tag{27b}$$

The proof of convergence of this iterative routine to a unique solution $\hat{\mathbf{\Sigma}}(\beta)$ has been recently introduced in [33] based on Perron-Frobenius theory. We refer to $\hat{\mathbf{\Sigma}}(\beta)$ as FP-DL in the sequel. Yet, the problem of selecting the shrinkage (loading factor) β is open and crucial. In [33] the authors suggested to specify the optimal loading factor β as the stochastic approximation of the Oracle (clairvoyant) scatter matrix $\mathbf{\Sigma}_0$, found as the minimum of the Frobenius norm of the error, i.e.,

$$\beta_O = \arg\min_{\beta}\mathcal{E}\left\{\left\|\tilde{\mathbf{\Sigma}}(\beta) - \mathbf{\Sigma}_0\right\|_F^2\right\} \tag{28}$$

where

$$\tilde{\mathbf{\Sigma}}(\beta) = (1-\beta)\frac{M}{T}\sum_{t=1}^{T}\frac{\mathbf{z}_t\mathbf{z}_t^H}{\mathbf{z}_t^H\mathbf{\Sigma}_0^{-1}\mathbf{z}_t} + \beta\mathbf{I}_M. \tag{29}$$

We would like to investigate how this Oracle estimator compares with the EL approach for selecting β: for conventional scenario ($T \geq M$) the EL approach selects the loading factor β such that

$$\begin{aligned}\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\mathbf{\Sigma}}(\beta)|\mathbf{Z}_T) &= |\hat{\mathbf{\Sigma}}_{\mathrm{ML}}\hat{\mathbf{\Sigma}}^{-1}(\beta)|\prod_{t=1}^{T}\left[\frac{\mathbf{z}_t^H\hat{\mathbf{\Sigma}}^{-1}(\beta)\mathbf{z}_t}{\mathbf{z}_t^H\hat{\mathbf{\Sigma}}_{\mathrm{ML}}^{-1}\mathbf{z}_t}\right]^{-M/T}\\ &\simeq \mathrm{med}[\omega(\mathrm{LR}|M,T)]\end{aligned} \tag{30}$$

where $\omega(\mathrm{LR}|M,T)$ is the true p.d.f. of the $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\mathbf{\Sigma}_0|\mathbf{Z}_T)$, $\hat{\mathbf{\Sigma}}_{\mathrm{ML}}$ is the complex Tyler's M-estimate (15) and $\mathrm{med}[\omega(\mathrm{LR}|M,T)]$ stands for the median value. Comparative analysis of the loading factor selection rules (28), (30) is introduced in the next section.

Let us now consider our fixed point $\mathrm{TVAR}(m)$ solution (26). Similarly to (27), we may introduce the diagonally loaded $\mathrm{TVAR}(m)$ fixed point solution as $\hat{\mathbf{\Sigma}}^{(m)}(\beta) = \lim_{k\to\infty}\hat{\mathbf{\Sigma}}_k^{(m)}(\beta)$ (provided that this limit exists and is unique, which so far is still an open problem) where $\hat{\mathbf{\Sigma}}_k^{(m)}(\beta)$ is obtained from

$$\begin{aligned}&\breve{\mathbf{\Sigma}}_{k+1}^{(m)}(\beta)\\ &= \mathsf{DG}^{(m)}\left[(1-\beta)\frac{M}{T}\sum_{t=1}^{T}\frac{\mathbf{z}_t\mathbf{z}_t^H}{\mathbf{z}_t^H\left(\hat{\mathbf{\Sigma}}_k^{(m)}(\beta)\right)^{-1}\mathbf{z}_t} + \beta\mathbf{I}_M\right]\end{aligned} \tag{31a}$$

$$\hat{\mathbf{\Sigma}}_{k+1}^{(m)}(\beta) = \frac{M}{\mathrm{Tr}\left\{\breve{\mathbf{\Sigma}}_{k+1}^{(m)}(\beta)\right\}}\breve{\mathbf{\Sigma}}_{k+1}^{(m)}(\beta). \tag{31b}$$

In the sequel, we refer to $\hat{\mathbf{\Sigma}}^{(m)}(\beta)$ as FP-DG-DL. It is noteworthy that for conventional Gaussian model $\mathbf{x}_t \sim \mathbb{C}\mathcal{N}(\mathbf{0},\mathbf{R}_0)$, the loaded $\mathrm{TVAR}(m)$ covariance matrix estimate proved quite an impressive improvement when applied to realistic data [35]. There are all reasons to expect similar improvement delivered by diagonal loading in (31). Yet, for this model two parameters (m,β) should be properly selected. Similarly to (30), parameters m or (m,β) may be treated as being properly selected if the likelihood ratio of the scatter matrices $\hat{\mathbf{\Sigma}}_{\mathrm{ML}}^{(m)}$ in (24) and $\hat{\mathbf{\Sigma}}^{(m)}(\beta)$ in (31) meet the expected likelihood condition. Finally, observe that while convergence of the fixed point iterations is an important theoretical issue, practically though, the EL criterion (30) may be used as a "stopping rule" for iterates that approach the EL threshold. Actual improvement in adaptive processing performance delivered by the suggested EL-supported regularized estimators is analyzed in the next section.

V. Performance Analysis. Simulation Results

Let us consider the case of data distributed according to a multivariate Student t-distribution with d degrees of freedom, defined herein as

$$f(\mathbf{x}_t|\mathbf{\Sigma}_0) \propto |\mathbf{\Sigma}_0|^{-1}\left[1 + d^{-1}\mathbf{x}_t{}^H\mathbf{\Sigma}_0^{-1}\mathbf{x}_t\right]^{-(M+d)}. \tag{32}$$

The r.v. $\mathbf{x}_t \overset{d}{=} \frac{\mathbb{C}\mathcal{N}(\mathbf{0},\Sigma_0)}{\sqrt{\mathbb{C}\chi_d^2/d}}$ where $\mathbb{C}\chi_d^2$ stands for the complex chi-square distribution with d degrees of freedom, whose p.d.f. is defined as $p(u) \propto u^{d-1}e^{-u}$. In all simulations below, we set $d = 1$. All algorithms will use the normalized data $\mathbf{z}_t = \mathbf{x}_t/\|\mathbf{x}_t\|$. Dimension M of uniform linear array (ULA) with half wavelength spacing was chosen to be $M = 12$ and the true scatter matrix was considered to be as per $\mathrm{AR}(1)$ process

$$[\mathbf{\Sigma}_0]_{m,n} = \rho_0^{|m-n|}.$$

Instead of mean-square error in covariance matrix estimation, we assess the quality of our estimates by analyzing the statistical properties of the SNR loss factor defined as [36], [37]

$$\mathrm{SNR}_{\mathrm{loss}} = \frac{\left(\mathbf{s}_0^H\hat{\mathbf{\Sigma}}^{-1}\mathbf{s}_0\right)^2}{\left(\mathbf{s}_0^H\hat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma}_0\hat{\mathbf{\Sigma}}^{-1}\mathbf{s}_0\right)\left(\mathbf{s}_0^H\mathbf{\Sigma}_0^{-1}\mathbf{s}_0\right)} \tag{33}$$

where $\mathbf{s}_0 = [1 \quad e^{i\pi\sin\theta_0} \quad \ldots \quad e^{i\pi(M-1)\sin\theta_0}]^T$ stands for the steering vector corresponding to the looked direction θ_0. In our simulations, we choose $\theta_0 = 5°$ and $\rho_0 = 0.99$ so that the SNR gain provided by the optimal Wiener filter $\mathbf{w}_{\mathrm{opt}} \propto \mathbf{\Sigma}_0^{-1}\mathbf{s}_0$ compared to a conventional beamformer $\mathbf{w}_{\mathrm{cbf}} \propto \mathbf{s}_0$ is about 12 dB. In (33), $\hat{\mathbf{\Sigma}}$ is a notation for a generic covariance matrix estimate considered in the sequel.

A. Distribution of the Likelihood Ratio

Let us first illustrate the theoretical results about the distributions of $\mathrm{LR}_{\mathrm{CES}}^{1/T}(\mathbf{\Sigma}_0|\mathbf{X}_T,g)$ and $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\mathbf{\Sigma}_0|\mathbf{Z}_T)$: both of them are independent of $\mathbf{\Sigma}_0$. The latter depends on M and T only while the former also depends on $g(\cdot)$. The median value of $\mathrm{LR}_{\mathrm{CES}}^{1/T}(\mathbf{\Sigma}_0|\mathbf{X}_T,g)$, for both a Gaussian and a Student t-distribution with $d = 1$ degree of freedom, as well as the median value of $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\mathbf{\Sigma}_0|\mathbf{Z}_T)$ are plotted in Fig. 1. Additionally, the p.d.f. of the above likelihood ratios is displayed in Fig. 2 for $T = 16$. The following comments can be made:

- The p.d.f. of $\mathrm{LR}_{\mathrm{CES}}^{1/T}(\mathbf{\Sigma}_0|\mathbf{X}_T,g)$ are seen to be nearly identical and seem to depend weakly on $g(\cdot)$: they are the same for $g(t) = \exp\{-t\}$ (Gaussian case) and $g(t) =$

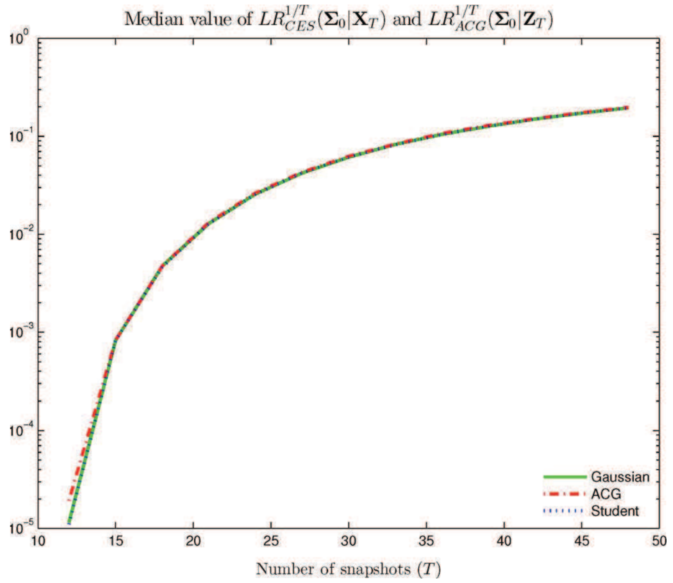

Fig. 1. Median value of $\mathrm{LR}_{\mathrm{CES}}^{1/T}(\mathbf{\Sigma}_0|\boldsymbol{X}_T, g)$ and $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\mathbf{\Sigma}_0|\boldsymbol{Z}_T)$ versus T. $M = 12$.

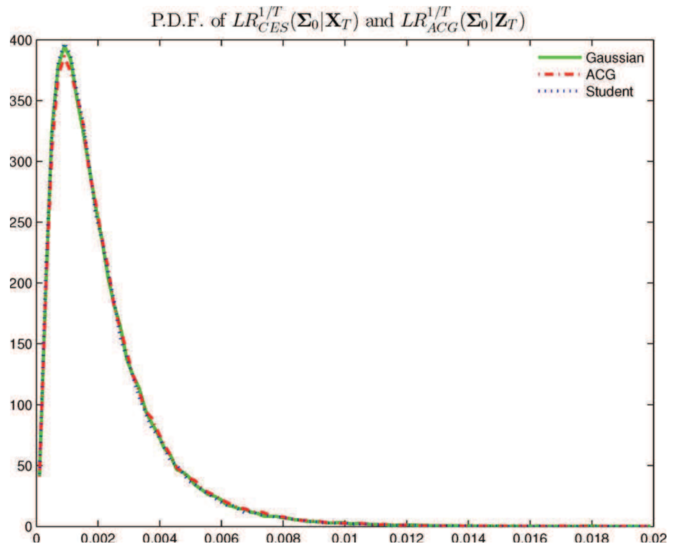

Fig. 2. Probability density function of $\mathrm{LR}_{\mathrm{CES}}^{1/T}(\mathbf{\Sigma}_0|\boldsymbol{X}_T, g)$ and $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\mathbf{\Sigma}_0|\boldsymbol{Z}_T)$. $M = 12$ and $T = 16$.

$(1 + t/d)^{-(d+M)}$ (Student case). Moreover, they are very close to the p.d.f. of $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\mathbf{\Sigma}_0|\boldsymbol{Z}_T)$. Therefore, the p.d.f. of the LR for the true scatter matrix $\mathbf{\Sigma}_0$ shows quite an invariance with respect to the distribution of the data. Note that asymptotically, it is known that $-2\ln \mathrm{LR}(\mathbf{\Sigma}_0|\boldsymbol{X}_T)$ converges to a $\chi^2_{M^2}$ distribution, which obviously does not depend on the data distribution: therefore, as $T \to \infty$, the distribution of the log likelihood ratio should not depend on $g(\,\cdot\,)$. It turns out that this is also approximately true in finite sample, although the finite sample distribution is not close to the asymptotic one.

- The median values are seen to be much inferior to 1, the value obtained with the MLE. These median values increase when T increases (for a fixed M) and when M decreases (for a fixed T). For large values of M the LR take very small values.

Acronyms	Estimators
FP-DL	(27)
FP-DG-DL	(31)
DL	(34)
DG-DL	(35)

B. Diagonally Loaded Estimates

We now study diagonally loaded regularized estimates, and more particularly the influence of the shrinkage factor β on both the LR values and the SNR loss. We consider here the estimate based on shrinkage of the normalized sample covariance matrix (NSCM) $\hat{\mathbf{\Sigma}}_{\mathrm{NSCM}} = \frac{M}{T}\sum_{t=1}^{T} \boldsymbol{z}_t\boldsymbol{z}_t^H$, i.e.,

$$\hat{\mathbf{\Sigma}}_{\mathrm{DL}}(\beta) = (1-\beta)\hat{\mathbf{\Sigma}}_{\mathrm{NSCM}} + \beta\boldsymbol{I}_M \tag{34}$$

(referred to as DL in the figures), its fixed-point version in (27) and their Dym-Gohberg regularization

$$\hat{\mathbf{\Sigma}}_{\mathrm{DG-DL}}(\beta) = \mathsf{DG}^{(m)}\left[(1-\beta)\hat{\mathbf{\Sigma}}_{\mathrm{NSCM}} + \beta\boldsymbol{I}_M\right] \tag{35}$$

(referred to as DG-DL in the figures) and the fixed-point TVAR(1) estimate (31). The value of m is set to $m = 1$. For the sake of convenience the following table relates the acronyms used in the figures with their corresponding estimators:

In Figs. 3–4 we investigate the influence of β and the relation between LR and SNR loss. The solid line there represents $\mathrm{med}[\omega(\mathrm{LR}|M,T)]$. These figures illustrate the fact that selecting the loading factor β from the EL principle results in a SNR very close to that of the optimal (clairvoyant) filter. Therefore, this validates selection of the loading factor using the EL approach. Observe that selecting β is a crucial issue for some estimates which are very sensitive to variations in β: this is particularly so for $\mathrm{TVAR}(1)$ estimates. In such cases, EL principle offers a quite efficient solution to the problem of selecting β. On the other hand, FP-DL is seen to be less sensitive to variations of β in terms of SNR loss: but this is also the case for the corresponding LR. Finally, note (and this will be observed in all simulations) that fixed-point estimates always outperform their non-iterative counterparts.

We now turn to performance analysis versus T. As before, we consider the shrinkage estimate (34) and its fixed-point iterative version in (27). For both of them, the loading factor β is chosen according to the EL principle in (30), viz

$$\beta_{\mathrm{EL}} = \arg\min_{\beta} \left| \mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\mathbf{\Sigma}}(\beta)|\boldsymbol{Z}_T) - \mathrm{med}[\omega(\mathrm{LR}|M,T)] \right| \tag{36}$$

that is the value of β for which $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\mathbf{\Sigma}}(\beta)|\boldsymbol{Z}_T)$ is closest to the median value of $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\mathbf{\Sigma}_0|\boldsymbol{Z}_T)$. For comparison purposes, we compare the EL-based estimates with the estimate of [33]. The latter corresponds to the FP-DL estimate of (31) where the loading factor β_O is chosen as in (28), and is given by

$$\beta_O = \frac{M^2 - M^{-1}\mathrm{Tr}\left\{\mathbf{\Sigma}_0\mathbf{\Sigma}_0^H\right\}}{(M^2 - MT - T) + (T + (T-1)/M)\mathrm{Tr}\left\{\mathbf{\Sigma}_0\mathbf{\Sigma}_0^H\right\}}. \tag{37}$$

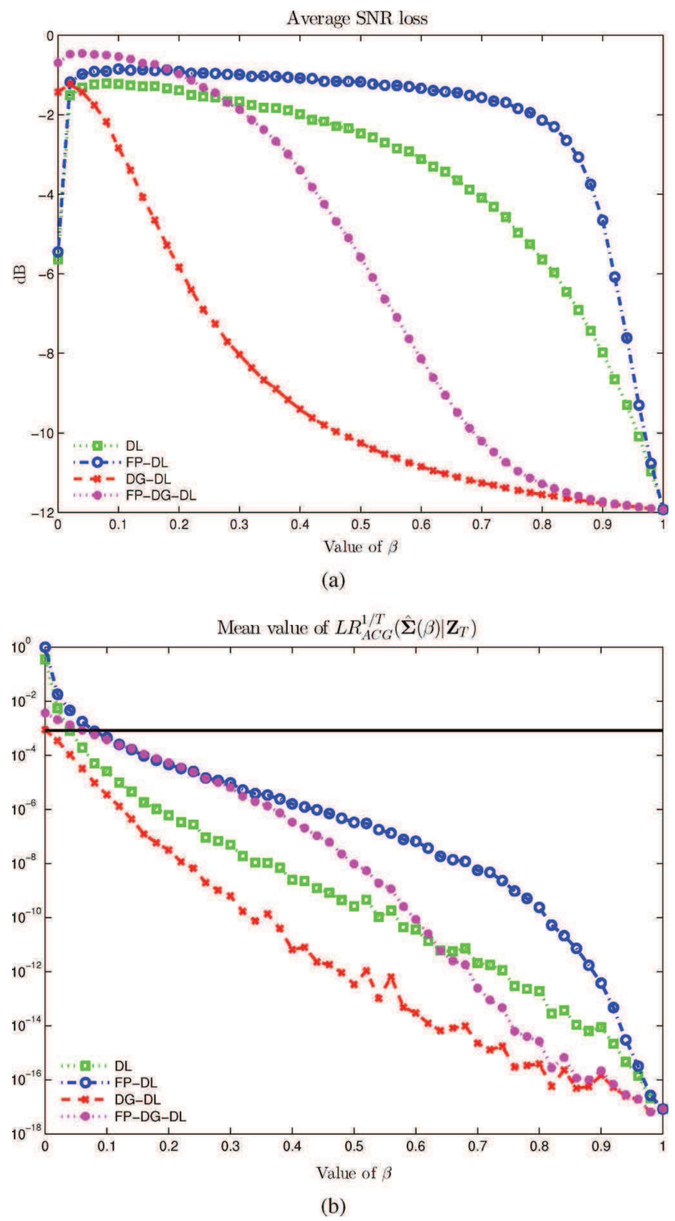

Fig. 3. Performance of diagonally loaded estimates versus β. $M = 12$, $d = 1$ and $T = 15$. (a) SNR loss (b) Mean value of $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\boldsymbol{\Sigma}}(\beta)|\boldsymbol{Z}_T)$.

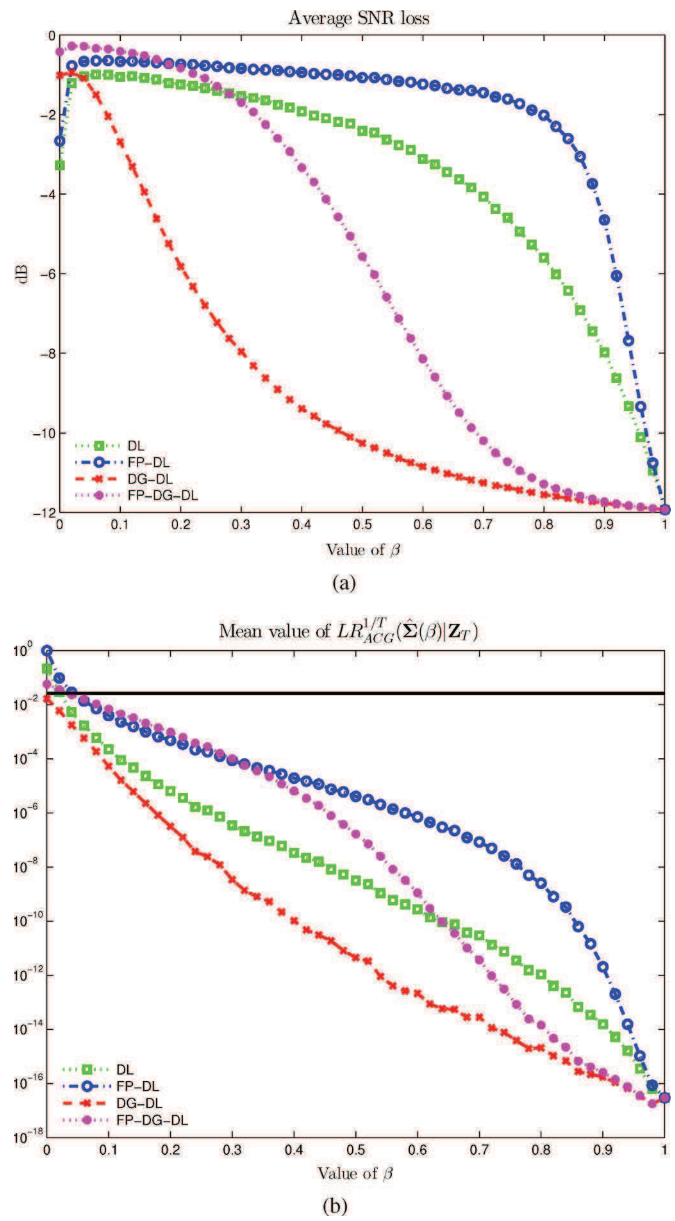

Fig. 4. Performance of diagonally loaded estimates versus β. $M = 12$, $d = 1$ and $T = 24$. (a) SNR loss (b) Mean value of $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\boldsymbol{\Sigma}}(\beta)|\boldsymbol{Z}_T)$.

We refer to $\hat{\boldsymbol{\Sigma}}(\beta_O)$ as the Oracle estimate. In Fig. 5 we display the average SNR loss, the mean value of the LR and the mean value of the loading factor selected by each method (DL and FP-DL correspond to the choice (36) of β). Interestingly enough, it appears that the Oracle loading factor β_O in (37) results in a matrix $\hat{\boldsymbol{\Sigma}}(\beta_O)$ whose LR closely matches that of $\boldsymbol{\Sigma}_0$. 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*, but slightly different values of the loading factor β. Yet, the EL and the Oracle estimate yields the same output SNR. This is because, as illustrated in Fig. 3(a), the FP-DL estimate is not very sensitive to variations in β. To summarize, this simulation shows that the Oracle estimate results in a LR value which matches $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)$. Since the EL approach selects the loading factor so that the resulting LR is also $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)$, the EL approach performs as well as the Oracle. It should also be stressed that FP-DL significantly outperforms the MLE, especially in low sample support, demonstrating the interest of regularization in this regime.

C. $\mathrm{TVAR}(m)$ *Regularization*

We now consider simulations with $\mathrm{TVAR}(m)$ estimates. In Fig. 6, we compare the estimates $\mathsf{DG}^{(m)}[\hat{\boldsymbol{\Sigma}}_{\mathrm{NSCM}}]$, $\mathsf{DG}^{(m)}[\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}]$, and the estimates in (31) and (35). When shrinkage is used in conjunction with Dym-Gohberg approximation, the value of β is selected according to the EL principle, i.e.,

$$\beta_{\mathrm{EL}}{}^{(m)} = \arg\min_{\beta} \left| \mathrm{LR}_{\mathrm{ACG}}^{1/T}\left(\hat{\boldsymbol{\Sigma}}^{(m)}(\beta)|\boldsymbol{Z}_T\right) - \mathrm{med}[\omega(\mathrm{LR}|M,T)] \right|. \quad (38)$$

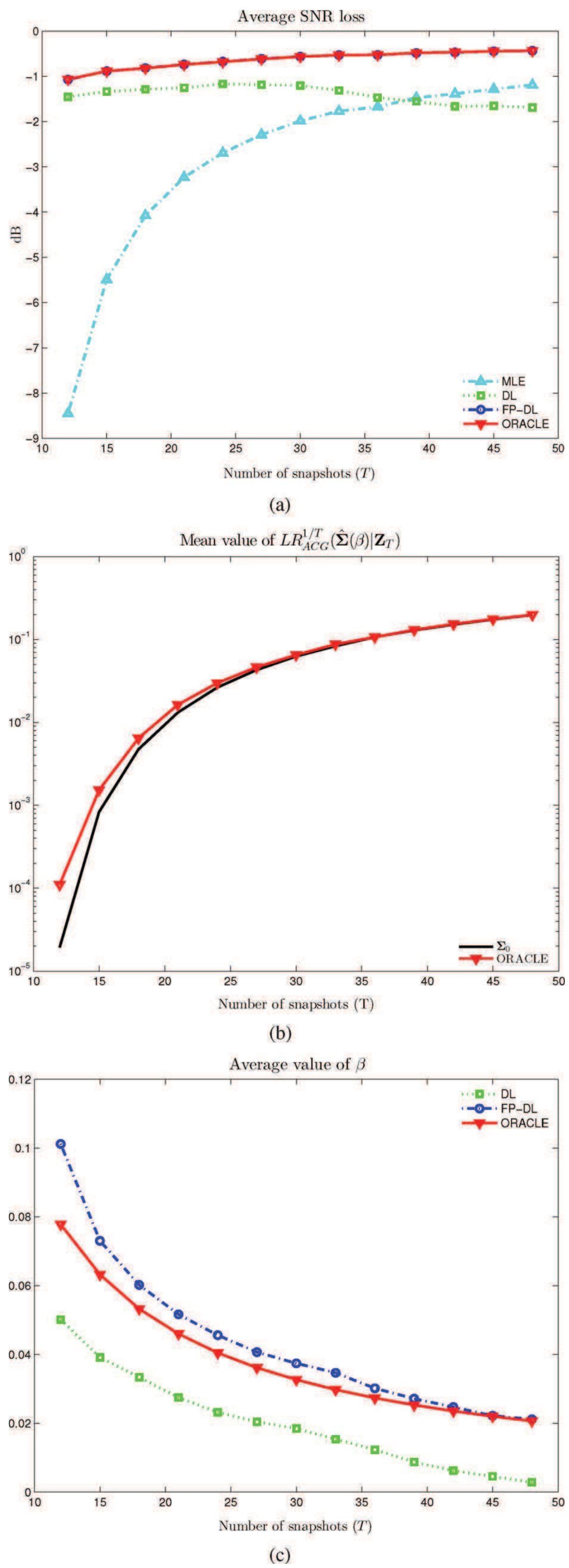

Fig. 5. Performance of diagonally loaded estimates versus number of snapshots T. $M = 12$ and $d = 1$. (a) SNR loss (b) Mean value of $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\boldsymbol{\Sigma}}(\beta)|\boldsymbol{Z}_T)$ (c) Mean value of loading factor.

The value of m is set to $m = 1$ in this simulation. As can be observed, the fixed-point diagonally loaded TVAR estimate offers the highest output SNR (average SNR loss of about -0.8 dB for $T = M$), followed by a Dym-Gohberg approximation of Tyler's MLE. It appears that shrinkage (or diagonal loading) associated with $\mathrm{TVAR}(1)$ modeling is not useful. This is further investigated now.

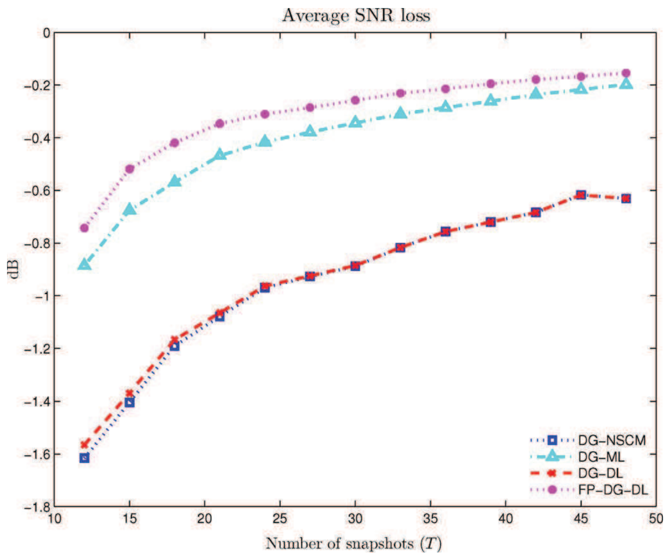

Fig. 6. Performance of $\mathrm{TVAR}(1)$ estimates versus number of snapshots T. $M = 12$, $d = 1$ and $m = 1$.

D. Comparison Between DL and $\mathrm{TVAR}(m)$ Estimates

Our next simulation explores the influence of the true underlying model for $\boldsymbol{\Sigma}_0$ onto regularized schemes which are based on a model $\boldsymbol{\Sigma}(\boldsymbol{\Omega}(\ell))$. More precisely, we study the respective performance of "shrinkage to the structure" (i.e., $\mathrm{TVAR}(m)$ only without diagonal loading), diagonal loading, and their combination, i.e., fixed-point diagonally loaded $\mathrm{TVAR}(m)$ estimates. We still consider the case of an $\mathrm{AR}(1)$ scatter matrix $\boldsymbol{\Sigma}_0(k, \ell) = \rho^{|k-\ell|}$: in this case, we wish to study if $\mathrm{TVAR}(m)$ only is better than FP-DL, and if diagonal loading can improve $\mathrm{TVAR}(m)$ estimation. We also consider a case where the (k, ℓ) element of $\boldsymbol{\Sigma}_0$ corresponds to the $|k - \ell|$-th correlation lag of an $\mathrm{ARMA}(2, 2)$ process whose spectrum (correlation) is close to but different from that of the $\mathrm{AR}(1)$ process considered so far. In any case, $\boldsymbol{\Sigma}_0^{-1}$ is not longer a banded matrix and $\boldsymbol{\Sigma}_0$ does not correspond to the covariance matrix of a $\mathrm{TVAR}(m)$ process. The fixed-point diagonal loading will be tested with two different choices of the loading factor β: either β is selected according to (36) or it is chosen so that $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\boldsymbol{\Sigma}}(\beta)|\boldsymbol{Z}_T) \simeq \mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\boldsymbol{\Sigma}}^{(m)}(0)|\boldsymbol{Z}_T)$. In the latter case, we thus compare $\mathrm{TVAR}(m)$ only and diagonal loading with the same likelihood ratio. Figs. 7–8 consider $m = 1$ while $m = 2$ in Figs. 9–10. The following conclusions can be drawn from observation of these figures. First, note that if the true scatter matrix $\boldsymbol{\Sigma}_0$ belongs to the class $\boldsymbol{\Sigma}(\boldsymbol{\Omega}(\ell))$, in the instance $\mathrm{TVAR}(1)$, shrinkage to the $\mathrm{TVAR}(1)$ structure alone (i.e., without DL) performs better than FP-DL even if the two estimates have the same LR, see Fig. 7. However, even in this case, a further reduction of LR to the median value leads to additional gains, i.e., $\mathrm{TVAR}(1) + \mathrm{DL}$ is found to be better than $\mathrm{TVAR}(1)$ alone. In contrast, in the case of an $\mathrm{ARMA}(2, 2)$

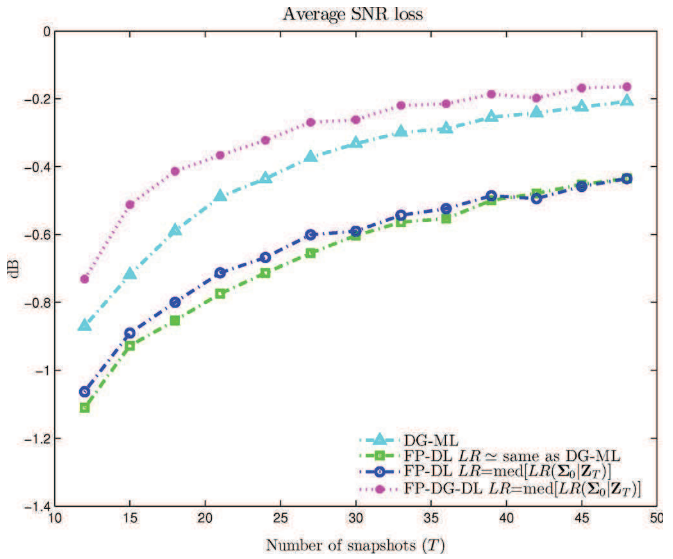

Fig. 7. Comparison between $\mathrm{TVAR}(1)$, fixed point diagonal loading and fixed point diagonally loaded $\mathrm{TVAR}(1)$ estimates in the $\mathrm{AR}(1)$ case. $M = 12$, $d = 1$ and $m = 1$.

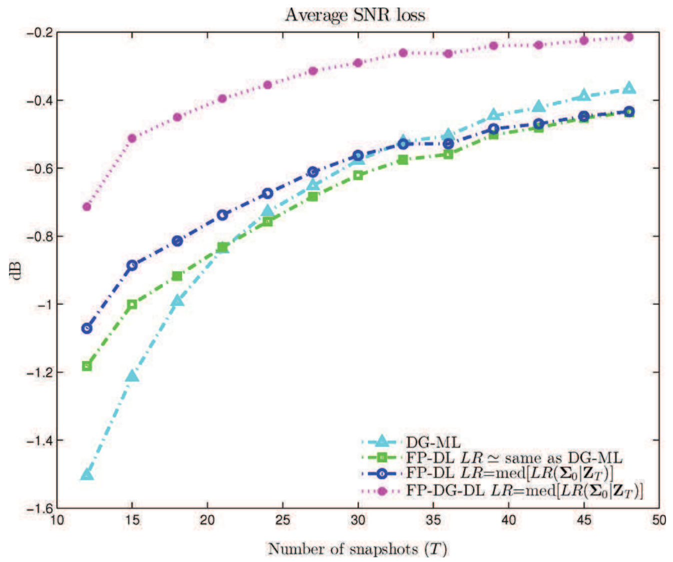

Fig. 9. Comparison between $\mathrm{TVAR}(2)$, fixed point diagonal loading and fixed point diagonally loaded $\mathrm{TVAR}(2)$ estimates in the $\mathrm{AR}(1)$ case. $M = 12$, $d = 1$ and $m = 2$.

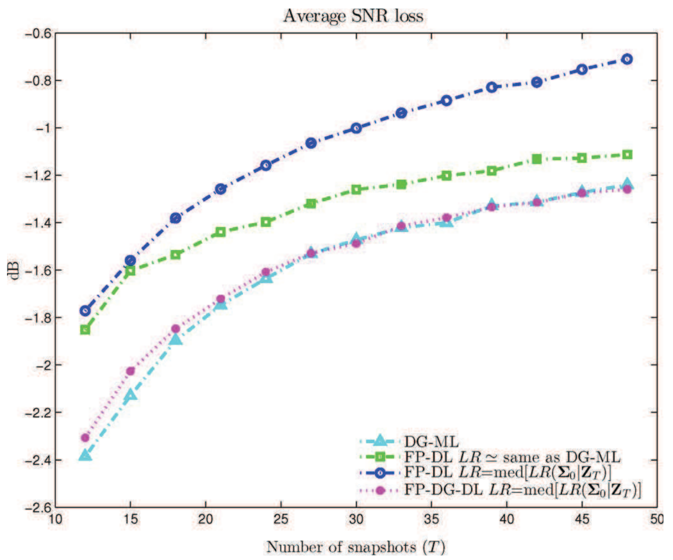

Fig. 8. Comparison between $\mathrm{TVAR}(1)$, fixed point diagonal loading and fixed point diagonally loaded $\mathrm{TVAR}(1)$ estimates in the $\mathrm{ARMA}(2,2)$ case. $M = 12$, $d = 1$ and $m = 1$.

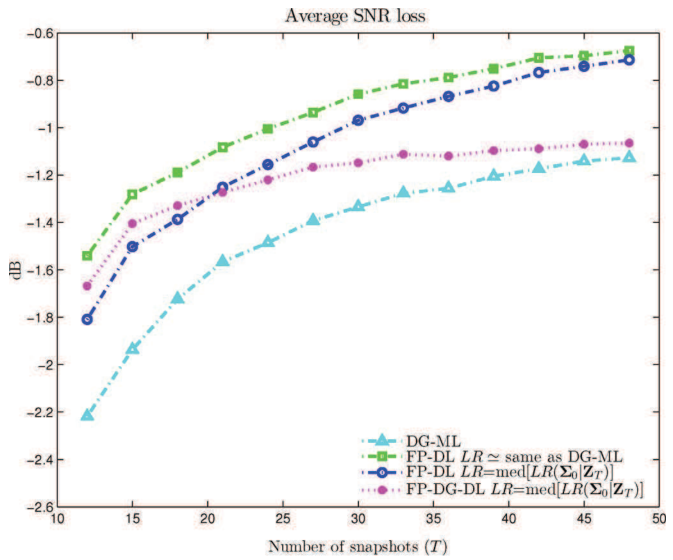

Fig. 10. Comparison between $\mathrm{TVAR}(2)$, fixed point diagonal loading and fixed point diagonally loaded $\mathrm{TVAR}(2)$ estimates in the $\mathrm{ARMA}(2,2)$ case. $M = 12$, $d = 1$ and $m = 2$.

scatter matrix, when $\mathbf{\Sigma}_0$ is not as per a $\mathrm{TVAR}(1)$ model, diagonal loading performs better than $\mathrm{TVAR}(m)$. It even performs better than diagonally loaded $\mathrm{TVAR}(m)$, as if when the two are used jointly, shrinkage to the structure is predominant. Therefore, there is no universally "best" regularization scheme: all depends on how close is the selected model to the true one. If we know or are lucky to select such one that the true matrix belongs to the restricted set, we get best results. If the restricted class does not include the true matrix, this "shrinkage to the structure" may be less efficient, and another shrinkage (actually FP-DL) may be more efficient.

So far, the order m of our $\mathrm{TVAR}(m)$ estimates was fixed. We now consider joint estimation of m and β according to the EL principle. When estimating β for fixed m, we followed the rule in (38), i.e., we looked for the matrix $\hat{\mathbf{\Sigma}}^{(m)}(\beta)$ whose LR is closest to the median LR. If the same strategy is adopted for estimation of both m and β, i.e., if we select the couple (m, β) so that $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\mathbf{\Sigma}}^{(m)}(\beta)|\boldsymbol{Z}_T)$ is closest to $\mathrm{med}[\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\mathbf{\Sigma}_0|\boldsymbol{Z}_T)]$, then high orders are likely to be chosen. In order to favor models with minimal order, we estimate m as the minimal order for which $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\mathbf{\Sigma}}^{(m)}(\beta_{\mathrm{EL}}{}^{(m)})|\boldsymbol{Z}_T)$ complies with $\omega(\mathrm{LR}|M,T)$ [11]. More precisely, m is estimated as

$$\hat{m} = \min\left\{m/\mathrm{LR}_{\mathrm{ACG}}^{1/T}\left(\hat{\mathbf{\Sigma}}^{(m)}\left(\beta_{\mathrm{EL}}{}^{(m)}\right)|\boldsymbol{Z}_T\right) > \eta_L\right\} \tag{39}$$

where η_L is the 10% quantile of $\omega(\mathrm{LR}|M,T)$, i.e., $\int_{\eta_L}^{1}\omega(\mathrm{LR}|M,T)dLR = 0.9$.

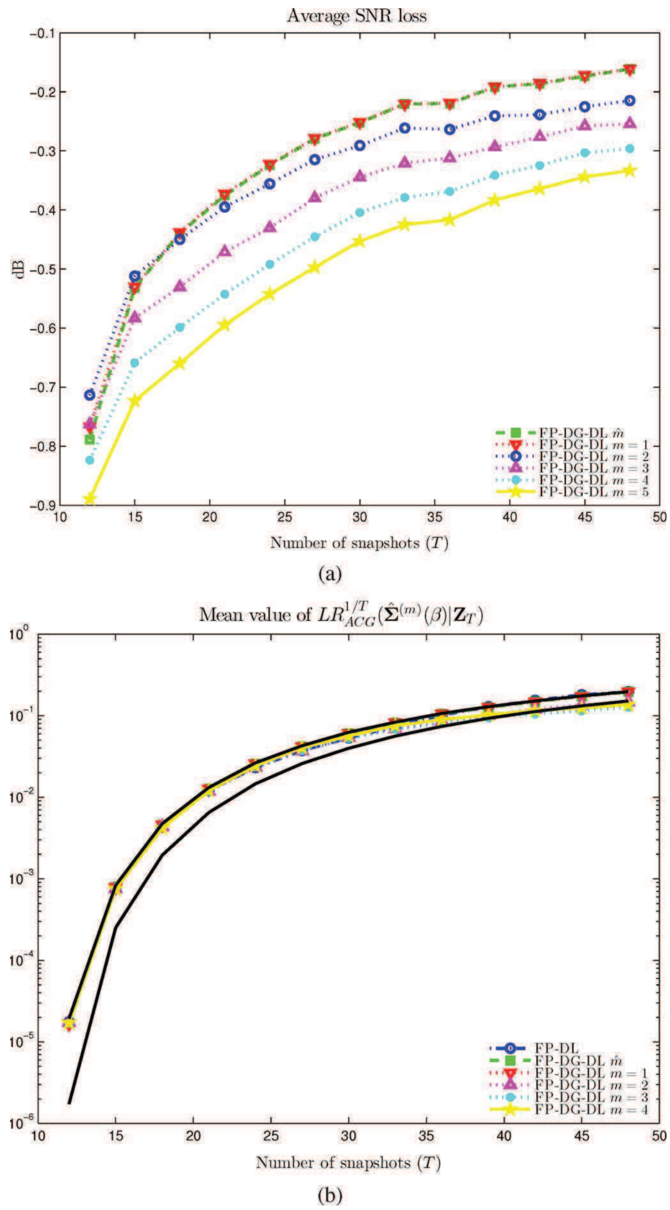

Fig. 11. Performance of $\mathrm{TVAR}(m)$ estimates versus number of snapshots T with an $\mathrm{AR}(1)$-type scatter matrix. $M = 12$ and $d = 1$. (a) SNR loss (b) Mean value of $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\mathbf{\Sigma}}^{(m)}(\beta)|\boldsymbol{Z}_T)$.

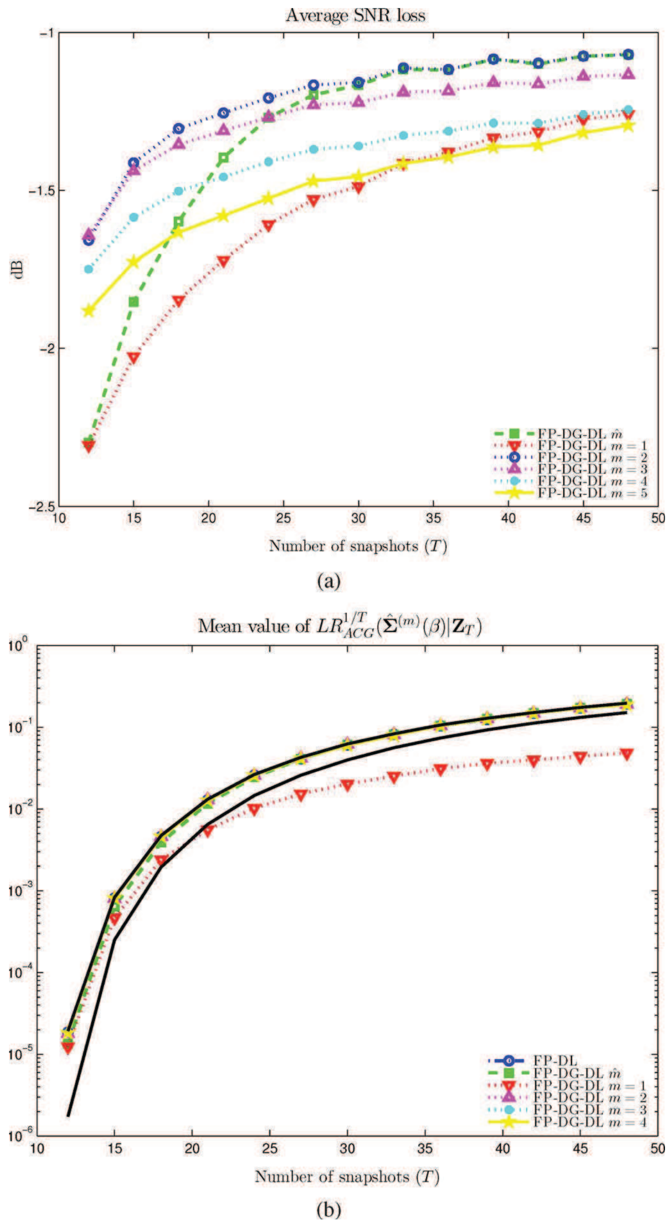

Fig. 12. Performance of $\mathrm{TVAR}(m)$ estimates versus number of snapshots T with an $\mathrm{ARMA}(2,2)$-type scatter matrix. $M = 12$ and $d = 1$. (a) SNR loss (b) Mean value of $\mathrm{LR}_{\mathrm{ACG}}^{1/T}(\hat{\mathbf{\Sigma}}^{(m)}(\beta)|\boldsymbol{Z}_T)$.

For comparison purposes, we also display the performance achieved when m is fixed to some value. The results are reported in Figs. 11–12 where we compare the performance of $\hat{\mathbf{\Sigma}}^{(m)}(\beta_{\mathrm{EL}}{}^{(m)})$ and $\hat{\mathbf{\Sigma}}^{(\hat{m})}(\beta_{\mathrm{EL}}{}^{(\hat{m})})$ in terms of SNR loss. In these figures, the black solid lines represent the threshold η_L and the median value. Additionally, we display in Tables I–II an histogram of the values of $\hat{m}$. As can be seen, in the $\mathrm{AR}(1)$-type scatter matrix $\mathbf{\Sigma}_0$, the estimated order is nearly always $\hat{m} = 1$, which is the true underlying model order. However for this case all $\mathrm{TVAR}(m)$ models yield LR values compliant with that of the true scatter matrix, i.e., at least above the threshold η_L. In the $\mathrm{ARMA}(2,2)$ case, as T increases, one has $\hat{m} = 2$ with high probability: this appears to be the best choice as least for T large enough. In fact the $\mathrm{TVAR}(1)$ estimate results in a LR which is below the threshold, which explains why one has to go to at least $m = 2$. These two simulations show that selecting m according to the EL principle yields a close to optimal solution. The $\mathrm{TVAR}(\hat{m})$ is shown to perform quite well, at least it does not penalize too much performance compared to fixing m. In this case, the optimal value of m is $m = M - 1$ but the SNR loss of $\mathrm{TVAR}(\hat{m})$ is within 0.4 dB. To summarize, whatever the case, in practice one does not know which value of m is optimal, and hence the latter must be set somehow arbitrarily. The EL principle offers an automatic way of estimating m which, in most situations, is very efficient. Accordingly, selection of β only for the FP-DL estimate according to the EL principle is very efficient.

VI. Summary and Conclusions

In this paper, we extended the expected likelihood methodology introduced in [10], [11] over the i.i.d. training samples

TABLE I
HISTOGRAMS OF TVAR MODEL ORDER ESTIMATED FROM THE EL PRINCIPLE VERSUS NUMBER OF SNAPSHOTS T IN THE $\mathrm{AR}(1)$ CASE. $M = 12$ AND $d = 1$

$\hat{m}$	1	2	3	4	$m \geq 5$
$T = 12$	97.4%	1.6%	0.6%	0.2%	0.2%
$T = 15$	99.2%	0.8%	0%	0%	0%
$T = 18$	99.4%	0.6%	0%	0%	0%
$T = 21$	99.4%	0.6%	0%	0%	0%
$T = 24$	99.2%	0.8%	0%	0%	0%
$T = 27$	99.8%	0.2%	0%	0%	0%
$T = 30$	99.8%	0.2%	0%	0%	0%
$T = 33$	100%	0%	0%	0%	0%
$T = 36$	99.6%	0.2%	0%	0%	0.2%
$T = 39$	100%	0%	0%	0%	0%
$T = 42$	99.8%	0.2%	0%	0%	0%
$T = 45$	99.4%	0.6%	0%	0%	0%
$T = 48$	100%	0%	0%	0%	0%

TABLE II
HISTOGRAMS OF TVAR MODEL ORDER ESTIMATED FROM THE EL PRINCIPLE VERSUS NUMBER OF SNAPSHOTS T IN THE $\mathrm{ARMA}(2,2)$ CASE. $M = 12$ AND $d = 1$

$\hat{m}$	1	2	3	4	$m \geq 5$
$T = 12$	88.4%	9.6%	1.4%	0.2%	0.4%
$T = 15$	71.6%	28.2%	0.2%	0%	0%
$T = 18$	55.4%	44.4%	0.2%	0%	0%
$T = 21$	32%	67.8%	0.2%	0%	0%
$T = 24$	17.6%	82.2%	0.2%	0%	0%
$T = 27$	9%	91%	0%	0%	0%
$T = 30$	2.8%	97.2%	0%	0%	0%
$T = 33$	1.2%	98.2%	0.6%	0%	0%
$T = 36$	0.8%	99%	0.2%	0%	0.2%
$T = 39$	0.2%	99.2%	0.6%	0%	0%
$T = 42$	0%	99.2%	0.8%	0%	0%
$T = 45$	0%	99.8%	0.2%	0%	0%
$T = 48$	0%	98.2%	1.8%	0%	0%

with complex elliptically symmetric distributions, and particularly over the class of samples with complex angular central Gaussian distribution. These distributions are appropriate for non-homogeneous clutter description when the covariance (scatter) matrix of this clutter is of full rank and the additive Gaussian internal noise may be ignored. In this first part, for conventional (over-sampled) training conditions, we demonstrated that for the true (a priori unknown) scatter matrix, the p.d.f. of the likelihood ratio does not depend on this matrix. For angular central Gaussian complex data, this p.d.f. is fully specified by the sample volume T and matrix dimension M, and does not depend on the density generator as per complex elliptically symmetric data. In those cases where the density generator is not accurately known a priori, it is therefore more appropriate to operate with the normalized training data that are described by the complex ACG distribution. While closed-form analytical formulas for the scenario-invariant p.d.f. have not been derived, Monte-Carlo simulations with i.i.d. white noise Gaussian random vectors could be used to pre-calculate these p.d.f. with any required accuracy. The particular quantiles of these p.d.f., such as median value, are then used as thresholds or target value for appropriate selection of shrinkage parameters in fixed-point scatter matrix estimation.

In particular, the EL approach was proposed for diagonal loading factor selection in the fixed-point regularized scatter matrix estimation scheme of [32]–[34]. Interestingly enough, we observed that the Oracle estimator (which minimizes the MSE) yields a value of the likelihood ratio which is very close to the median LR for the true scatter matrix. Since the latter is the target value for the EL approach, it demonstrates that the EL approach is statistically sound. Furthermore, we explored in this paper another type of regularization, different from diagonal loading, often referred to as shrinkage to the structure. Specifically, we introduced the fixed-point ML $\mathrm{TVAR}(m)$ scatter matrix estimate, along with a diagonally loaded version of it. We showed that for autoregressive experimental data, $\mathrm{TVAR}(1)$ estimates perform better than fixed-point diagonal loading: yet, introduction of DL in conjunction with $\mathrm{TVAR}(1)$ shows improvement with respect to $\mathrm{TVAR}(1)$ only. When the true scatter matrix does not belong to the $\mathrm{TVAR}(m)$ class, then fixed-point diagonal loading was shown to outperform $\mathrm{TVAR}(m)$-based estimates, while the difference is not large. It was also demonstrated that the EL approach allows for an accurate estimation of the best model $\mathrm{TVAR}(m)$ order.

Hence, the EL approach offers a systematic, statistically sound and efficient way of fixing the regularization parameters in regularized covariance matrix estimation schemes. Moreover, the extension to CES and ACG distributions presented in this paper expand our ability to address problems with severe in-homogeneity of training data in adaptive processing applications. Regularized covariance matrix estimates, well developed and proven to be highly effective in adaptive antenna (filter) applications with multivariate complex Gaussian data, now got extended over a broader class of CES and ACG distributions.

REFERENCES

[1] M. Rangaswamy, “Spherically invariant random processes for modeling non-Gaussian radar clutter,” in *Proc. 27th Asilomar Conf.*, Pacific Grove, CA, Nov. 1–3, 1993, pp. 1106–1110.

[2] G. Frazer, D. Meehan, Y. Abramovich, and B. Johnson, “Mode-selective OTHR: A new cost-effective sensor for maritime domain awareness,” in *Proc. IEEE Radar Conf.*, May 10–14, 2010, pp. 935–940.

[3] G. Frazer, Y. Abramovich, and B. Johnson, “Mode-selective OTH radar: Experimental results for one-way transmission via the ionosphere,” in *Proc. IEEE Radar Conf.*, May 23–27, 2011, pp. 397–402.

[4] K. Yao, “A representation theorem and its application to spherically invariant processes,” *IEEE Trans. Inf. Theory*, vol. 19, no. 5, pp. 600–608, Sep. 1973.

[5] E. Conte and M. Longo, “Characterisation of radar clutter as a spherically invariant process,” in *Proc. IEE Radar, Sonar Navig.*, Apr. 1987, vol. 134, no. 2, pp. 191–197.

[6] E. Conte, M. Lops, and G. Ricci, “Asymptotically optimum radar detection in compound-Gaussian clutter,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 2, pp. 617–625, Apr. 1995.

[7] J. H. Michels, M. Rangaswamy, and B. Himed, “Performance of parametric and covariance based STAP tests in compound-Gaussian clutter,” *Digit. Signal Process.*, vol. 12, no. 2–3, pp. 307–328, Apr.–Jul. 2002.

[8] P. Stoica and A. Nehorai, “MUSIC, maximum likelihood and Cramér-Rao bound,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[9] P. Stoica and A. Nehorai, “MUSIC, maximum likelihood and Cramér-Rao bound: Further results and comparisons,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.

[10] 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.

[11] Y. I. Abramovich, N. K. Spencer, and A. Y. Gorokhov, “Modified GLRT and AMF framework for adaptive detectors,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 1017–1051, Jul. 2007.

[12] 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.

[13] 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.

[14] O. P. Cheremisin, “Efficiency of adaptive algorithms with regularised sample covariance matrix,” *Radio Eng. Electron. Phys.*, vol. 27, no. 10, pp. 69–77, 1982.

[15] 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.

[16] P. R. Krishnaiah and J. Lin, “Complex elliptically symmetric dsitributions,” *Commun. Statist.—Theory Methods*, vol. 15, no. 12, pp. 3693–3718, 1986.

[17] K. T. Fang and Y. T. Zhang*, Generalized Multivariate Analysis*. Berlin: Springer Verlag, 1990.

[18] T. W. Anderson and K.-T. Fang, Theory and Applications of Elliptically Contoured and Related Distributions Dep. Statistics, Stanford Univ., , Tech. Rep. 24, Sep. 1990.

[19] A. C. Micheas, D. K. Dey, and K. V. Mardia, “Complex elliptical distributions with application to shape analysis,” *J. Statist. Plan. Interf.*, vol. 136, no. 9, pp. 2961–2982, Sep. 2006.

[20] E. Ollila and V. Koivunen, “Robust antenna array processing using m-estimators of pseudo-covariance,” in *Proc. 14th IEEE Int. Symp. Pers., Indoor Mobile Radio Commun.*, Beijing, China, Sep. 7–10, 2003, pp. 2659–2663.

[21] E. Ollila and V. Koivunen, “Influence function and asymptotic efficiency of scatter matrix based array processors,” *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 247–259, Jan. 2009.

[22] D. E. Tyler, “A distribution-free M-estimator of multivariate scatter,” *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, Mar. 1987.

[23] J. T. Kent and D. E. Tyler, “Redescending M-estimates of multivariate location and scatter,” *Ann. Statist.*, vol. 19, no. 4, pp. 2102–2119, Dec. 1991.

[24] R. J. Muirhead*, Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.

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

[26] E. Ollila and D. E. Tyler, “Distribution-free detection under complex elliptically symmetric clutter distribution,” in *Proc. 7th SAM Workshop*, Hoboken, NJ, Jun. 17–20, 2012, pp. 421–424.

[27] J. T. Kent and D. E. Tyler, “Maximum likelihood estimation for the wrapped Cauchy distribution,” *J. Appl. Statist.*, vol. 15, no. 2, pp. 247–2549, 1988.

[28] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, “Covariance structure maximum-likelihood estimates in compound Gaussian noise: Existence and algorithm analysis,” *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 34–48, Jan. 2008.

[29] Y. Chitour and F. Pascal, “Exact maximum likelihood estimates for SIRV covariance matrix: Existence and algorithm analysis,” *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4563–4573, Oct. 2008.

[30] Y. I. Abramovich, N. K. Spencer, and M. D. E. Turley, “Time-varying autoregressive (TVAR) models for multiple radar observations,” *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1298–1311, Apr. 2007.

[31] A. Hjorugnes*, Complex-Valued Matrix Derivatives With Applications in Signal Processing and Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

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

[33] 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.

[34] A. Wiesel, “Unified framework to regularized covariance estimation in scaled Gaussian models,” *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 29–38, Jan. 2012.

[35] Y. Abramovich, M. Rangaswamy, B. Johnson, P. Corbell, and N. Spencer, “Performance analysis of two-dimensional parametric STAP for airborne radar using KASSPER data,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 1, pp. 118–139, Jan. 2011.

[36] 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.

[37] J. Ward, Space-Time Adaptive Processing for Airborne Radar Lincoln Lab., Mass. Inst. Technol., Lexington, MA, Tech. Rep. 1015, Dec. 1994.

Yuri I. Abramovich (M’96–SM’06–F’08) received the Dipl. Eng. (Honors) degree in radio electronics in 1967 and the Cand. Sci. degree (Ph.D. equivalent) in theoretical radio techniques in 1971, both from the Odessa Polytechnic University, Odessa (Ukraine), U.S.S.R., and in 1981, he received the D.Sc. degree in radar and navigation from the Leningrad Institute for Avionics, Leningrad (Russia), U.S.S.R.

From 1968 to 1994, he was with the Odessa State Polytechnic University, Odessa, as a Research Fellow, Professor, and ultimately as Vice-Chancellor of Science and Research. From 1994 to 2006, he was at the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP), Adelaide, Australia. He joined the Australian Defence Science and Technology Organisation (DSTO), Adelaide, in 2000, as principal research scientist, seconded to CSSIP until its closure. As of January 2012, he has been with W.R. Systems Ltd., Fairfax, VA. His research interests are in signal processing (particularly spatio-temporal adaptive processing, beamforming, signal detection and estimation), its application to radar (particularly over-the-horizon radar), electronic warfare, and communications.

Dr. Abramovich was Associate Editor of IEEE Transactions on Signal Processing from 2002 to 2005. Since 2007, he has served as Associate Editor of IEEE Transactions on Aerospace and Electronic Systems and is currently a member of the IEEE AESS Board of Governors.

Olivier Besson (SM’04) received the M. S. and Ph.D. degrees in signal processing in 1988 and 1992, respectively, both from the Institut National Polytechnique, Toulouse.

He is currently a Professor with the Department of Electronics, Optronics and Signal of the Institut Supérieur de l’Aéronautique et de l’Espace (ISAE), Toulouse. His research interests are in the area of robust adaptive array processing, mainly for radar applications.

Dr. Besson is a member of the Sensor Array and Multichannel technical committee (SAM TC) of the IEEE Signal Processing Society.