Regularized Covariance Matrix Estimation in Complex Elliptically Symmetric Distributions Using the Expected Likelihood Approach - Part 2: The Under-Sampled Case

Olivier Besson, Yuri Abramovich

To link to this article: DOI: 10.1109/TSP.2013.2285511
URL: http://dx.doi.org/10.1109/TSP.2013.2285511

To cite this version: Besson, Olivier and Abramovich, Yuri *Regularized Covariance Matrix Estimation in Complex Elliptically Symmetric Distributions Using the Expected Likelihood Approach - Part 2: The Under-Sampled Case.* (2013) IEEE Transactions on Signal Processing, vol. 61 (n° 23). ISSN 1053-587X

Regularized Covariance Matrix Estimation in Complex Elliptically Symmetric Distributions Using the Expected Likelihood Approach—Part 2: The Under-Sampled Case

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

***Abstract*—In the first part of these two papers, we extended the expected likelihood approach originally developed in the Gaussian case, to the broader class of complex elliptically symmetric (CES) distributions and complex angular central Gaussian (ACG) distributions. More precisely, we demonstrated that the probability density function (p.d.f.) of the likelihood ratio (LR) for the (unknown) actual scatter matrix $\boldsymbol{\Sigma}_0$ does not depend on the latter: it only depends on the density generator for the CES distribution and is distribution-free in the case of ACG distributed data, i.e., it only depends on the matrix dimension M and the number of independent training samples T, assuming that $T \geq M$. Additionally, regularized scatter matrix estimates based on the EL methodology were derived. In this second part, we consider the under-sampled scenario ($T \leq M$) which deserves specific treatment since conventional maximum likelihood estimates do not exist. Indeed, inference about the scatter matrix can only be made in the T-dimensional subspace spanned by the columns of the data matrix. We extend the results derived under the Gaussian assumption to the CES and ACG class of distributions. Invariance properties of the under-sampled likelihood ratio evaluated at $\boldsymbol{\Sigma}_0$ are presented. Remarkably enough, in the ACG case, the p.d.f. of this LR can be written in a rather simple form as a product of beta distributed random variables. The regularized schemes derived in the first part, based on the EL principle, are extended to the under-sampled scenario and assessed through numerical simulations.**

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

I. Introduction

The Gaussian assumption has been historically the dominating framework for adaptive radar detection problems, partly because of the richness of statistical tools available to derive detection/estimation schemes and to assess their performance in finite sample problems. The most famous examples include the celebrated Reed Mallet Brennan rule for characterization of the signal to noise ratio loss of adaptive filters [1] or, for detection problems, the now classic papers by Kelly [2]–[4] about generalized likelihood ratio test (GLRT) in unknown Gaussian noise or the adaptive subspace detectors of [5]–[7] in partially homogeneous noise environments. All of them highly benefit from the beautiful and rich theory of multivariate Gaussian distributions and Wishart matrices [8]–[10] and have served as references for decades. At the core of adaptive filtering or adaptive detection is the problem of estimating the disturbance covariance matrix. It is usually addressed through the maximum likelihood (ML) principle, mainly because ML estimates have the desirable property of being asymptotically efficient [11], [12]. However, in low sample support, their performance may degrade and they can be significantly improved upon using regularized covariance matrix estimates (CME), such as diagonal loading [13], [14]. Moreover, the ML estimator results in the ultimate equal to one likelihood ratio (LR), a property that is questionable, as argued in [15]–[17]. In the latter references, it is proved that the LR, evaluated at the true covariance matrix $\boldsymbol{R}_0$, has a probability density function (p.d.f.) that does not depend on $\boldsymbol{R}_0$ but only on the sample volume T and the dimension M of the observation space, i.e., number of antennas or pulses. More importantly, with high probability the LR takes values much lower than one and, therefore, one may wonder if an estimate whose LR significantly exceeds that of the true covariance matrix is reasonable. Based on these results, the expected likelihood (EL) principle was developed in [15]–[17] with successful application to adaptive detection or direction of arrival (DoA) estimation. In the former case, regularized estimation schemes were investigated with a view to drive down the LR to values that are compliant with those for $\boldsymbol{R}_0$, the true underlying covariance matrix. As for DoA estimation, the EL approach was instrumental in identifying severely erroneous MUSIC DoA estimates (breakdown prediction) and rectifying the set of these estimates to meet the expected likelihood ratio values (breakdown cure) [15], [16].

However, in a number of applications, the Gaussian assumption may be violated and detection/estimation schemes based on this assumption may suffer from a certain lack of robustness, resulting in significant performance degradation. Therefore, many studies have focused on more accurate radar data modeling along with corresponding detection/estimation schemes. In this respect, the class of compound-Gaussian models, see e.g., [18]–[20], has been extensively studied. The radar return is here modeled as the product of a positive valued

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

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

random variable (r.va.) called texture and an independent complex Gaussian random vector (r.v.) called speckle, and is referred to as a spherically invariant random vector (SIRV). Since exact knowledge of the p.d.f. of the texture is seldom available, the usual way is to treat the textures as unknown deterministic quantities and to carry out ML estimation of the speckle covariance matrix [21]–[24]. This approach results in an implicit equation which is solved through an iterative procedure. SIRV belong to a broader class, namely complex elliptically symmetric (CES) distributions [25], [26] which have recently been studied for array processing applications, see [27] and references therein. A CES distributed r.v. $\boldsymbol{x}$ has a stochastic representation of the form $\boldsymbol{x} \overset{d}{=} \sqrt{\mathcal{Q}}\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{u}$ where $\boldsymbol{\Sigma}_0$ is the scatter matrix, $\sqrt{\mathcal{Q}}$ is called the modular variate and is independent of the complex random vector $\boldsymbol{u}$ which is uniformly distributed on the complex M-sphere. In most practical situations, the p.d.f. of $\mathcal{Q}$ is not known, and therefore there is an interest to estimate $\boldsymbol{\Sigma}_0$ irrespective of it. A mechanism to achieve this goal is to normalize $\boldsymbol{x}$ as $\boldsymbol{z} = \boldsymbol{x}/\left\|\boldsymbol{x}\right\|_2$ whose p.d.f. is described by the complex angular central Gaussian (ACG) distribution and is specified by the scatter matrix only. There is thus a growing interest in deriving scatter matrix estimates (SME) within the framework of CES or ACG distributions, see the comprehensive reviews of Esa Ollila *et al.* in [27] and Ami Wiesel in [28]. In the first part [29] of this series of papers, we addressed this problem using the EL approach. More precisely, we extended the EL principle from the Gaussian framework to the CES and ACG distributions, and proved invariance properties of the LR for the true scatter matrix $\boldsymbol{\Sigma}_0$. The over-sampled scenario $(T \geq M)$ only was considered in [29]. However, in some applications the number of antenna elements M exceeds the number of i.i.d. training samples T and therefore the under-sampled scenario $(T \leq M)$ is of utmost importance. This case deserves a special treatment as MLE do not longer exist and inference about the scatter matrix is possible only in the T-dimensional subspace spanned by the columns of the data matrix [30]. The goal of this paper is thus to extend the results of [30], which deals with Gaussian data, to CES and ACG distributed data and to complement [29] by considering $T \leq M$. Accordingly, the regularized estimation schemes developed in [29] will be adapted to this new case. As we hinted at above, CES distributions rely on the knowledge of the p.d.f. of the modular variate while ACG distributions do not. Therefore, in the sequel, we will concentrate on the ACG case.

More precisely, in Section II, we derive the LR for ACG distributions in the under-sampled case. We demonstrate its invariance properties and show that, for $T = M$, it coincides with the over-sampled LR of [29]. The case of CES distributions is addressed in the Appendix. In Section III we briefly review the regularized estimates of [29] and indicate how their regularization parameters are chosen in the under-sampled case. Numerical simulations are presented in Section IV and our conclusions are drawn in Section V.

II. Likelihood Ratio for Complex ACG Distributions in the Under-Sampled Case

As said previously, the likelihood ratio (and its p.d.f. when evaluated at the true (covariance) scatter matrix) is the fundamental quantity for the EL approach. In this section, we derive this likelihood ratio for under-sampled training conditions $(T \leq M)$ in the case of complex ACG distributed data. For Gaussian distributed data, the under-sampled scenario has been studied in [30], [31] where the EL approach was used to detect outliers produced by MUSIC DoA estimation, and in [32], [33] for adaptive detection using regularized covariance matrix estimates. As explained in [30], this scenario requires a specific analysis since (unstructured) maximum likelihood estimates do not longer exist, and information about the covariance matrix can be retrieved only in the T-dimensional subspace spanned by the data matrix. Moreover, in deriving an under-sampled likelihood ratio LR^u, some requirements are in force. Of course, LR^u should lie in the interval $[0, 1]$ and maximization of the likelihood ratio should be associated to maximization of the likelihood function, at least over a restricted set. Additionally, the p.d.f. of LR^u, when evaluated at the true covariance matrix, should depend only on M and T, so as to implement an EL approach. Finally, when $T = M$, LR^u should coincide with its over-sampled counterpart. In the sequel, we build upon the theory developed in [30] and extend it to the case of ACG distributions.

A vector $\boldsymbol{z}$ is said to have a complex angular central Gaussian (ACG) distribution, which we denote as $\boldsymbol{z} \sim \mathbb{C}\mathcal{AG}(\mathbf{0}, \boldsymbol{\Sigma}_0)$, if it can be written as $\boldsymbol{z} = \boldsymbol{x}/\left\|\boldsymbol{x}\right\|_2$ where $\boldsymbol{x}$ follows a complex central Gaussian distribution, i.e., $\boldsymbol{x} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_0)$. For non-singular $\boldsymbol{\Sigma}_0$, the p.d.f. of $\boldsymbol{z}$ is given by [27], [34], [35]

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

where $\propto$ means proportional to. In fact, for any vector $\boldsymbol{x} \sim \mathbb{C}\mathcal{E}_M(\mathbf{0}, \boldsymbol{\Sigma}_0, g)$ which follows a central CES distribution with scatter matrix $\boldsymbol{\Sigma}_0$ and density generator $g(.)$, the p.d.f. of $\boldsymbol{z} = \boldsymbol{x}/\left\|\boldsymbol{x}\right\|_2$ is still given by (1), and therefore (1) is the density for a large class of scaled vectors. Note that $\boldsymbol{\Sigma}_0$ in (1) is identifiable only up to a scaling factor and can be seen as a shape matrix. Let us assume that we have a set of T i.i.d. samples $\boldsymbol{z}_t$ drawn from the p.d.f. in (1). Then, the joint distribution of $\boldsymbol{Z}_T = [\boldsymbol{z}_1 \cdots \boldsymbol{z}_T]$ can be written as

$$f(\boldsymbol{Z}_T|\boldsymbol{\Sigma}_0) \propto |\boldsymbol{\Sigma}_0|^{-T}\prod_{t=1}^{T}\left(\boldsymbol{z}_t^H\boldsymbol{\Sigma}_0^{-1}\boldsymbol{z}_t\right)^{-M}. \tag{2}$$

Let us then consider the likelihood ratio for testing a parametric scatter (or shape) matrix model $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$ where $\boldsymbol{\Omega}_\ell$ is a set of ℓ parameters that uniquely specify the scatter matrix model. In [29], we derived the LR for over-sampled training conditions $(T \geq M)$ and showed that

$$\begin{aligned} LR_{ACG}\left(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T\right) &= \frac{f(\boldsymbol{Z}_T|\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell))}{\max_{\boldsymbol{\Sigma}} f(\boldsymbol{Z}_T|\boldsymbol{\Sigma})} \\ &= |\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)|^T\prod_{t=1}^{T}\left[\frac{\boldsymbol{z}_t^H\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)\boldsymbol{z}_t}{\boldsymbol{z}_t^H\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1}\boldsymbol{z}_t}\right]^{-M} \end{aligned} \tag{3}$$

where $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ is the maximum likelihood estimate of $\boldsymbol{\Sigma}$, and is the unique (up to a scaling factor) solution [36] to

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}} = \frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1}\boldsymbol{z}_t}. \tag{4}$$

Let us now turn to the under-sampled scenario with $T < M$. Obviously, with $T < M$ training samples, any inference regarding the scatter matrix $\boldsymbol{\Sigma}_0$ may be provided only regarding its projection onto the T-dimensional subspace spanned by the columns of $\boldsymbol{Z}_T = \hat{\boldsymbol{U}}_T \hat{\boldsymbol{\Lambda}}_T^{1/2} \hat{\boldsymbol{V}}_T^H$, or equivalently by the columns of the $M \times T$-variate matrix of eigenvectors $\hat{\boldsymbol{U}}_T$ associated with the T non-zero eigenvalues of the sample matrix $\boldsymbol{Z}_T \boldsymbol{Z}_T^H = \hat{\boldsymbol{U}}_T \hat{\boldsymbol{\Lambda}}_T \hat{\boldsymbol{U}}_T^H$, where $\hat{\boldsymbol{\Lambda}}_T$ stands for the diagonal matrix of the eigenvalues. As already noted, whether $\boldsymbol{x}_t \sim \mathbb{C}\mathcal{E}_M(\boldsymbol{0}, \boldsymbol{\Sigma}_0, g)$ or $\boldsymbol{x}_t \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_0)$, we still have the normalized vectors $\boldsymbol{z}_t \sim \mathbb{C}\mathcal{AG}(\boldsymbol{0}, \boldsymbol{\Sigma}_0)$. Therefore, without loss of generality, we may consider the vectors $\boldsymbol{z}_t$ as being generated by complex Gaussian random vectors $\boldsymbol{x}_t$. For any given candidate $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$, we need to find the full rank Hermitian matrix $\boldsymbol{D}_T(\boldsymbol{\Omega}_\ell)$[1], such that the construct $\boldsymbol{\Sigma}_T(\boldsymbol{\Omega}_\ell) = \hat{\boldsymbol{U}}_T \boldsymbol{D}_T(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T^H$ is "closest" to the model $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$. In [30] it was demonstrated that $\boldsymbol{D}_T(\boldsymbol{\Omega}_\ell)$ may be specified by the condition that the generalized non-zero eigenvalues of the matrix pencil $\hat{\boldsymbol{U}}_T \boldsymbol{D}_T(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T^H - \mu \boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$ are all equal to one, i.e., $\mu_1 = \mu_2 = \cdots = \mu_T = 1$. Since $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell) > 0$, the generalized eigenvalues μ_t, $t = 1, \ldots, T$ are the same as the non-zero eigenvalues of the M-variate Hermitian matrix $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T \boldsymbol{D}_T(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\Omega}_\ell)$ or, since $\boldsymbol{D}_T(\boldsymbol{\Omega}_\ell) > 0$, the non-zero eigenvalues of the T-variate Hermitian matrix $\boldsymbol{D}_T^{1/2}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T \boldsymbol{D}_T^{1/2}(\boldsymbol{\Omega}_\ell)$, which immediately leads to the solution [30]

$$\boldsymbol{D}_T(\boldsymbol{\Omega}_\ell) = \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T\right]^{-1} \tag{5}$$

and

$$\boldsymbol{\Sigma}_T(\boldsymbol{\Omega}_\ell) = \hat{\boldsymbol{U}}_T \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T\right]^{-1} \hat{\boldsymbol{U}}_T^H. \tag{6}$$

Note that for any (arbitrary) matrix $\boldsymbol{\Sigma}$, we might construct the corresponding $\boldsymbol{D}_T = \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{U}}_T\right]^{-1}$ and $\boldsymbol{\Sigma}_T = \hat{\boldsymbol{U}}_T \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{U}}_T\right]^{-1} \hat{\boldsymbol{U}}_T^H$: the latter gathers what can be inferred of $\boldsymbol{\Sigma}$ from the observation of $T < M$ snapshots. It is important to note that for the given generating set of $T < M$ i.i.d Gaussian data $\boldsymbol{x}_t$, $t = 1, \ldots, T$, the scatter matrix $\boldsymbol{\Sigma}_T(\boldsymbol{\Omega}_\ell)$ may be treated as an admissible *singular* covariance matrix model.

At this stage, we need to define ACG distributions with singular covariance matrices and we will follow the lines of Siotani *et al.* [37] who considered singular Gaussian distributions. Let $\boldsymbol{x} \in \mathbb{C}^M$ be Gaussian distributed with a rank-deficient covariance matrix $\boldsymbol{R} = \boldsymbol{U}_r \boldsymbol{D}_r \boldsymbol{U}_r^H$ where $\boldsymbol{U}_r$ is a $M \times r$ orthonormal matrix whose columns span the range space of $\boldsymbol{R}$ and $\boldsymbol{D}_r$ is a positive definite Hermitian (PDH) $r \times r$ matrix. Note that $\boldsymbol{x}$ fully resides in the subspace spanned by $\boldsymbol{U}_r$ with probability one [10], [37]. Let $\boldsymbol{U}_r^\perp$ denote an orthonormal basis for the complement of $\boldsymbol{U}_r$, i.e., $\left(\boldsymbol{U}_r^\perp\right)^H \boldsymbol{U}_r^\perp = \boldsymbol{I}_{M-r}$ and $\left(\boldsymbol{U}_r^\perp\right)^H \boldsymbol{U}_r = \boldsymbol{0}$. Let $\boldsymbol{U} = \left[\boldsymbol{U}_r \quad \boldsymbol{U}_r^\perp\right]$ and let us make the change of variables

$$\tilde{\boldsymbol{x}} = \boldsymbol{U}^H \boldsymbol{x} = \begin{bmatrix} \boldsymbol{U}_r^H \boldsymbol{x} \\ \left(\boldsymbol{U}_r^\perp\right)^H \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{x}}_r \\ \boldsymbol{0} \end{bmatrix}.$$

[1]We should have denoted $\boldsymbol{D}_T\left[\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)\right]$ and $\boldsymbol{\Sigma}_T\left[\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)\right]$ to emphasize that these matrices are constructed from $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$ but, for the ease of notation, we simplify to $\boldsymbol{D}_T(\boldsymbol{\Omega}_\ell)$ and $\boldsymbol{\Sigma}_T(\boldsymbol{\Omega}_\ell)$.

Then $\tilde{\boldsymbol{x}}_r \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{D}_r)$ and its p.d.f. is given by

$$\begin{aligned} f(\tilde{\boldsymbol{x}}_r | \boldsymbol{D}_r) &\propto |\boldsymbol{D}_r|^{-1} \mathrm{etr}\{-\tilde{\boldsymbol{x}}_r^H \boldsymbol{D}_r^{-1} \tilde{\boldsymbol{x}}_r\} \\ &\propto |\boldsymbol{D}_r|^{-1} \mathrm{etr}\{-\boldsymbol{x}^H \boldsymbol{U}_r \boldsymbol{D}_r^{-1} \boldsymbol{U}_r^H \boldsymbol{x}\} \end{aligned}$$

where $\mathrm{etr}\{.\}$ stands for the exponential of the trace of the matrix between braces. Since the Jacobian from $\boldsymbol{x}$ to $\tilde{\boldsymbol{x}}$ is 1, [37] defines a singular Gaussian density as

$$f(\boldsymbol{x} | \boldsymbol{U}_r, \boldsymbol{D}_r) \propto |\boldsymbol{D}_r|^{-1} \mathrm{etr}\{-\boldsymbol{x}^H \boldsymbol{U}_r \boldsymbol{D}_r^{-1} \boldsymbol{U}_r^H \boldsymbol{x}\} \tag{7}$$

for those vectors $\boldsymbol{x}$ such that $\boldsymbol{U}_r^{\perp H} \boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{U}_r^{\perp H} \boldsymbol{R} = \boldsymbol{0}$. Let us now consider $\boldsymbol{z} = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ and define

$$\tilde{\boldsymbol{z}} = \boldsymbol{U}^H \boldsymbol{z} = \frac{\boldsymbol{U}^H \boldsymbol{x}}{\|\boldsymbol{x}\|} = \begin{bmatrix} \tilde{\boldsymbol{x}}_r / \|\tilde{\boldsymbol{x}}_r\| \\ \boldsymbol{0} \end{bmatrix} \triangleq \begin{bmatrix} \tilde{\boldsymbol{z}}_r \\ \boldsymbol{0} \end{bmatrix}.$$

Then $\tilde{\boldsymbol{z}}_r \sim \mathbb{C}\mathcal{AG}(\boldsymbol{0}, \boldsymbol{D}_r)$ follows an ACG distribution with p.d.f.

$$\begin{aligned} f(\tilde{\boldsymbol{z}}_r | \boldsymbol{D}_r) &\propto |\boldsymbol{D}_r|^{-1} \left(\tilde{\boldsymbol{z}}_r^H \boldsymbol{D}_r^{-1} \tilde{\boldsymbol{z}}_r\right)^{-r} \\ &\propto |\boldsymbol{D}_r|^{-1} \left(\boldsymbol{z}^H \boldsymbol{U}_r \boldsymbol{D}_r^{-1} \boldsymbol{U}_r^H \boldsymbol{z}\right)^{-r}. \end{aligned}$$

Following Siotani *et al.*, one can thus define a singular ACG density as

$$f(\boldsymbol{z} | \boldsymbol{U}_r, \boldsymbol{D}_r) \propto |\boldsymbol{D}_r|^{-1} \left(\boldsymbol{z}^H \boldsymbol{U}_r \boldsymbol{D}_r^{-1} \boldsymbol{U}_r^H \boldsymbol{z}\right)^{-r} \tag{8}$$

for $\boldsymbol{z} \in \mathbb{C}S^M = \left\{\boldsymbol{z} \in \mathbb{C}^M / \|\boldsymbol{z}\| = 1\right\}$ and $\boldsymbol{U}_r^{\perp H} \boldsymbol{z} = \boldsymbol{0}$. A set of T independent snapshots $\boldsymbol{Z}_T$ can thus be represented as $\boldsymbol{Z}_T = \boldsymbol{U}_r \tilde{\boldsymbol{Z}}_T$, with density

$$f(\boldsymbol{Z}_T | \boldsymbol{U}_r, \boldsymbol{D}_r) \propto |\boldsymbol{D}_r|^{-T} \prod_{t=1}^{T} \left(\boldsymbol{z}_t^H \boldsymbol{U}_r \boldsymbol{D}_r^{-1} \boldsymbol{U}_r^H \boldsymbol{z}_t\right)^{-r}. \tag{9}$$

Let us assume that $\boldsymbol{U}_r$ *is known*. For $T \geq r$, differentiating (9) with respect to $\boldsymbol{D}_r$ (for fixed $\boldsymbol{U}_r$), it follows that the MLE of $\boldsymbol{D}_r$ satisfies, see also (4)

$$\hat{\boldsymbol{D}}_r^{\mathrm{ML}} = \frac{r}{T} \sum_{t=1}^{T} \frac{\boldsymbol{U}_r^H \boldsymbol{z}_t \boldsymbol{z}_t^H \boldsymbol{U}_r}{\boldsymbol{z}_t^H \boldsymbol{U}_r (\hat{\boldsymbol{D}}_r^{\mathrm{ML}})^{-1} \boldsymbol{U}_r^H \boldsymbol{z}_t}. \tag{10}$$

Furthermore, let us consider the *specific case where the rank of* $\boldsymbol{R}$ *equals the number of available observations*, i.e., $T = r$. Assuming that the $r \times r$ matrix $\boldsymbol{U}_r^H \boldsymbol{Z}_T$ is non-singular, one has

$$\hat{\boldsymbol{D}}_r^{\mathrm{ML}} = \boldsymbol{U}_r^H \boldsymbol{Z}_T \boldsymbol{Z}_T^H \boldsymbol{U}_r \qquad (\text{for } T = r). \tag{11}$$

Indeed, in this case, one has $\boldsymbol{Z}_T^H \boldsymbol{U}_r \times \left(\boldsymbol{U}_r^H \boldsymbol{Z}_T \boldsymbol{Z}_T^H \boldsymbol{U}_r\right)^{-1} \boldsymbol{U}_r^H \boldsymbol{Z}_T = \boldsymbol{I}_r$. Hence

$$\frac{r}{r} \sum_{t=1}^{r} \frac{\boldsymbol{U}_r^H \boldsymbol{z}_t \boldsymbol{z}_t^H \boldsymbol{U}_r}{\boldsymbol{z}_t^H \boldsymbol{U}_r \left(\boldsymbol{U}_r^H \boldsymbol{Z}_T \boldsymbol{Z}_T^H \boldsymbol{U}_r\right)^{-1} \boldsymbol{U}_r^H \boldsymbol{z}_t} = \boldsymbol{U}_r^H \boldsymbol{Z}_T \boldsymbol{Z}_T^H \boldsymbol{U}_r \tag{12}$$

which proves that $\boldsymbol{U}_r^H \boldsymbol{Z}_T \boldsymbol{Z}_T^H \boldsymbol{U}_r$ verifies (10) for $T = r$, and hence is the MLE in this case. This observation is of utmost importance when we consider the under-sampled case.

Indeed, for our specific application with $\boldsymbol{\Sigma}_T(\boldsymbol{\Omega}_\ell)$ in (6) being an admissible singular covariance matrix, we get

$$\begin{aligned} f(\boldsymbol{Z}_T|\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)) &\propto |\boldsymbol{D}_T(\boldsymbol{\Omega}_\ell)|^{-T} \\ &\times \prod_{t=1}^{T} \left(\boldsymbol{z}_t^H \hat{\boldsymbol{U}}_T \boldsymbol{D}_T(\boldsymbol{\Omega}_\ell)^{-1} \hat{\boldsymbol{U}}_T^H \boldsymbol{z}_t\right)^{-T} \\ &\propto |\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T|^T \\ &\times \prod_{t=1}^{T} \left(\boldsymbol{z}_t^H \hat{\boldsymbol{U}}_T \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T\right] \hat{\boldsymbol{U}}_T^H \boldsymbol{z}_t\right)^{-T}. \end{aligned} \quad (13)$$

The previous equation provides the likelihood function for the parameterized scatter matrix $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$. In order to obtain the likelihood ratio $LR^u_{ACG}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T) = f(\boldsymbol{Z}_T|\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell))/\max_{\boldsymbol{\Sigma}} f(\boldsymbol{Z}_T|\boldsymbol{\Sigma})$ in under-sampled conditions $T < M$, we need to find the global ML maximum of $f(\boldsymbol{Z}_T|\boldsymbol{\Sigma})$ over the $T \times T$ PDH matrix $\boldsymbol{D}_T = \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{U}}_T\right]^{-1}$. As proved in (11), this MLE is simply

$$\hat{\boldsymbol{D}}_T^{\mathrm{ML}} = \hat{\boldsymbol{U}}_T^H \boldsymbol{Z}_T \boldsymbol{Z}_T^H \hat{\boldsymbol{U}}_T = \hat{\boldsymbol{\Lambda}}_T \quad (14)$$

where $\hat{\boldsymbol{\Lambda}}_T$ is the diagonal matrix of the eigenvalues of $\boldsymbol{Z}_T\boldsymbol{Z}_T^H$. Therefore, for an under-sampled ($T < M$) scenario, we may use the under-sampled likelihood ratio which can be written in the following equivalent forms:

$$\begin{aligned} LR^u_{ACG}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T) &= |\hat{\boldsymbol{\Lambda}}_T \hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T|^T \\ &\times \prod_{t=1}^{T} \left[\boldsymbol{z}_t^H \hat{\boldsymbol{U}}_T \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T\right] \hat{\boldsymbol{U}}_T^H \boldsymbol{z}_t\right]^{-T} \\ &= |\hat{\boldsymbol{\Lambda}}_T \hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T|^T \prod_{t=1}^{T} \left[\boldsymbol{z}_t^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \boldsymbol{z}_t\right]^{-T} \\ &= |\boldsymbol{Z}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \boldsymbol{Z}_T|^T \prod_{t=1}^{T} \left[\boldsymbol{z}_t^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \boldsymbol{z}_t\right]^{-T} \\ &= \left[\prod_{\mathrm{eig}_j > 0} \mathrm{eig}_j \left(\boldsymbol{Z}_T \boldsymbol{Z}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)\right)\right]^T \\ &\times \prod_{t=1}^{T} \left[\boldsymbol{z}_t^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \boldsymbol{z}_t\right]^{-T}. \end{aligned} \quad (15)$$

It is noteworthy that $LR^u_{ACG}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T)$ is invariant to scaling of $\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T$. Let us now investigate the properties of this under-sampled likelihood ratio.

Let us first prove that, *for $T = M$, the under-sampled LR* (15) *coincides with its over-sampled counterpart* $LR_{ACG}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T)$ in (3). To do so, one needs to derive an expression for Tyler's MLE $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ in (4). In fact, using derivations similar to those which led to (10), one can show that, for $T = M$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}} = \boldsymbol{Z}_T\boldsymbol{Z}_T^H$ since $\boldsymbol{Z}_T^H\left(\boldsymbol{Z}_T\boldsymbol{Z}_T^H\right)^{-1}\boldsymbol{Z}_T = \boldsymbol{I}_T$ and hence

$$\frac{T}{T}\sum_{t=1}^{T} \frac{\boldsymbol{z}_t \boldsymbol{z}_t^H}{\boldsymbol{z}_t^H \left(\boldsymbol{Z}_T \boldsymbol{Z}_T^H\right)^{-1} \boldsymbol{z}_t} = \boldsymbol{Z}_T \boldsymbol{Z}_T^H. \quad (16)$$

Reporting this value in (3) yields, for $T = M$

$$\begin{aligned} LR_{ACG}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T) &= \left|\boldsymbol{Z}_T \boldsymbol{Z}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)\right|^T \\ &\times \prod_{t=1}^{T} \left[\boldsymbol{z}_t^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \boldsymbol{z}_t\right]^{-T} \end{aligned} \quad (17)$$

which coincides with $LR^u_{ACG}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T)$ in (15) when $T = M$.

Let us now prove that similarly to the over-sampled case, for the true scatter matrix $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell) = \boldsymbol{\Sigma}_0$, *the p.d.f. for* $LR^u_{ACG}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)$ *does not depend on* $\boldsymbol{\Sigma}_0$. Observing that $\boldsymbol{z}_t = \frac{\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{n}_t}{\left\|\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{n}_t\right\|_2}$ where $\boldsymbol{n}_t \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ or $\boldsymbol{n}_t \sim \mathcal{U}\left(\mathbb{C}S^M\right)$, it ensues that

$$\begin{aligned} LR^u_{ACG}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T) &= |\boldsymbol{Z}_T^H \boldsymbol{\Sigma}_0^{-1} \boldsymbol{Z}_T|^T \prod_{t=1}^{T} \left[\boldsymbol{z}_t^H \boldsymbol{\Sigma}_0^{-1} \boldsymbol{z}_t\right]^{-T} \\ &\overset{d}{=} |\boldsymbol{N}_T^H \boldsymbol{N}_T|^T \prod_{t=1}^{T} \left[\boldsymbol{n}_t^H \boldsymbol{n}_t\right]^{-T} \end{aligned} \quad (18)$$

which is obviously distribution-free. More insights into the distribution of $LR^u_{ACG}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)$ can be gained by noticing that the matrix $\boldsymbol{W} = \boldsymbol{N}_T^H\boldsymbol{N}_T$ has a complex Wishart distribution with M degrees of freedom, i.e., $\boldsymbol{W} \sim \mathbb{C}\mathcal{W}(M, \boldsymbol{I}_T)$. Let us consider the Bartlett decomposition $\boldsymbol{W} = \boldsymbol{C}^H\boldsymbol{C}$ where $\boldsymbol{C}$ is an upper-triangular matrix and all random variables $\boldsymbol{C}_{ij}$ are independent [8]. Moreover, $|\boldsymbol{C}_{ii}|^2 \sim \mathbb{C}\chi^2_{M-i+1}$ where $\mathbb{C}\chi^2_n$ stands for the complex central chi-square distribution with n degrees of freedom, whose p.d.f. is given by $f_{\mathbb{C}\chi^2_n}(x) = \Gamma^{-1}(n)x^{n-1}\exp\{-x\}$. Additionally, one has $\boldsymbol{C}_{ij} \sim \mathbb{C}\mathcal{N}(0,1)$ for $i \neq j$. It then ensues that

$$\begin{aligned} &[LR^u_{ACG}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)]^{1/T} = |\boldsymbol{W}| \prod_{t=1}^{T} \boldsymbol{W}_{tt}^{-1} \\ &= \frac{|\boldsymbol{C}_{11}|^2}{|\boldsymbol{C}_{11}|^2} \times \frac{|\boldsymbol{C}_{22}|^2}{|\boldsymbol{C}_{12}|^2 + |\boldsymbol{C}_{22}|^2} \times \cdots \times \frac{|\boldsymbol{C}_{TT}|^2}{|\boldsymbol{C}_{1T}|^2 + \cdots + |\boldsymbol{C}_{TT}|^2} \\ &\overset{d}{=} \prod_{t=2}^{T} \frac{\mathbb{C}\chi^2_{M-t+1}}{\mathbb{C}\chi^2_{t-1} + \mathbb{C}\chi^2_{M-t+1}} \overset{d}{=} \prod_{t=1}^{T-1} \mathrm{Beta}(M-t, t) \end{aligned} \quad (19)$$

where $\mathrm{Beta}(a,b)$ stands for the beta distribution. The representation (19) makes it very simple to estimate the distribution of $[LR^u_{ACG}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)]^{1/T}$. Additionally, the average value of $[LR^u_{ACG}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)]^{1/T}$ can be obtained in a straightforward manner as

$$\mathcal{E}\left\{[LR^u_{ACG}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)]^{1/T}\right\} = \prod_{t=1}^{T-1} \frac{M-t}{M}. \quad (20)$$

This average value (or the median value) can serve as a target value for the likelihood ratio associated with any scatter matrix estimate.

To summarize, for under-sampled ($T < N$) training conditions and ACG data $\boldsymbol{z}_t = \boldsymbol{x}_t/\left\|\boldsymbol{x}_t\right\|_2$ with $\boldsymbol{x}_t \sim \mathbb{C}\mathcal{E}_M(\boldsymbol{0}, \boldsymbol{\Sigma}_0, g)$, we introduced the likelihood ratio $LR^u_{ACG}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{Z}_T) \le 1$ that for the true scatter matrix $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell) = \boldsymbol{\Sigma}_0$ is described by a scenario-invariant p.d.f. fully specified by parameters M and T. While an analytical expression for the above mentioned p.d.f. is not available, it can be pre-calculated for some given M and T by Monte-Carlo simulations, using either simulated i.i.d Gaussian r.v. $\boldsymbol{n}_t \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_M)$, cf. Equation (18) or beta distributed random variables, cf. Equation (19). In the Appendix, we derive the under-sampled likelihood ratio $LR^u_{CES}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{X}_T, g)$ for CES distributed samples $\boldsymbol{x}_t \sim \mathbb{C}\mathcal{E}_M(\boldsymbol{0}, \boldsymbol{\Sigma}_0, g)$. We show that, when evaluated at $\boldsymbol{\Sigma}_0$, its p.d.f. does not depend on $\boldsymbol{\Sigma}_0$ but still depends on the density generator $g(.)$, similarly to what was observed in the over-sampled case [29].

III. Regularized Scatter Matrix Estimation Using the Expected Likelihood Approach

For the sake of clarity, we here briefly review the regularized scatter matrix estimates (SME) which were introduced and studied in part 1 for $T \geq M$. More precisely, we focus on the schemes which were shown to achieve the best performance. The first estimate is the conventional diagonal loading estimate

$$\hat{\boldsymbol{\Sigma}}_{\text{DL}}(\beta) = (1-\beta)\frac{M}{T}\sum_{t=1}^{T} \boldsymbol{z}_t \boldsymbol{z}_t^H + \beta \boldsymbol{I}_M. \tag{21}$$

We also consider the fixed point diagonally loaded estimator [28], [38], [39] $\hat{\boldsymbol{\Sigma}}(\beta) = \lim_{k\to\infty} \hat{\boldsymbol{\Sigma}}_k(\beta)$ where $\hat{\boldsymbol{\Sigma}}_k(\beta)$ is obtained from the following iterative algorithm

$$\breve{\boldsymbol{\Sigma}}_{k+1}(\beta) = (1-\beta)\frac{M}{T}\sum_{t=1}^{T} \frac{\boldsymbol{z}_t \boldsymbol{z}_t^H}{\boldsymbol{z}_t^H \left(\hat{\boldsymbol{\Sigma}}_k(\beta)\right)^{-1} \boldsymbol{z}_t} + \beta \boldsymbol{I}_M \tag{22a}$$

$$\hat{\boldsymbol{\Sigma}}_{k+1}(\beta) = \frac{M}{\text{Tr}\{\breve{\boldsymbol{\Sigma}}_{k+1}(\beta)\}}\breve{\boldsymbol{\Sigma}}_{k+1}(\beta). \tag{22b}$$

We refer to $\hat{\boldsymbol{\Sigma}}(\beta)$ as FP-DL in what follows. Both estimates are governed by the loading factor β which is chosen according to the EL principle, i.e.,

$$\beta_{\text{EL}} = \arg\min_{\beta} \left| \left[LR^u_{ACG}\left(\hat{\boldsymbol{\Sigma}}(\beta)|\boldsymbol{Z}_T\right)\right]^{1/T} - \text{med}\left[\omega^u(LR|M,T)\right] \right| \tag{23}$$

where $\omega^u(LR|M,T)$ is the scenario-invariant p.d.f. of the T-th root of $LR^u_{ACG}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)$ in (19), $\text{med}[\omega^u(LR|M,T)]$ stands for the median value and $LR^u_{ACG}\left(\hat{\boldsymbol{\Sigma}}(\beta)|\boldsymbol{Z}_T\right)$ is the under-sampled LR of (15):

$$\left[LR^u_{ACG}\left(\hat{\boldsymbol{\Sigma}}(\beta)|\boldsymbol{Z}_T\right)\right]^{1/T} = |\hat{\boldsymbol{\Lambda}}_T \hat{\boldsymbol{U}}_T^H \hat{\boldsymbol{\Sigma}}^{-1}(\beta)\hat{\boldsymbol{U}}_T| \times \prod_{t=1}^{T}\left[\boldsymbol{z}_t^H \hat{\boldsymbol{\Sigma}}^{-1}(\beta)\boldsymbol{z}_t\right]^{-1}. \tag{24}$$

In other words, the loading factor is such that $\left[LR^u_{ACG}\left(\hat{\boldsymbol{\Sigma}}(\beta)|\boldsymbol{Z}_T\right)\right]^{1/T}$ is closest to the median value of $[LR^u_{ACG}(\boldsymbol{\Sigma}_0|\boldsymbol{Z}_T)]^{1/T}$. For comparison purposes, we will consider the Oracle estimator of [39] defined through the following choice of β:

$$\beta_O = \arg\min_{\beta} \mathcal{E}\left\{\left\|\tilde{\boldsymbol{\Sigma}}(\beta) - \boldsymbol{\Sigma}_0\right\|_F^2\right\} = \frac{M^2 - M^{-1}\text{Tr}\{\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_0^H\}}{(M^2 - MT - T) + (T + (T-1)/M)\text{Tr}\{\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_0^H\}} \tag{25}$$

where $\tilde{\boldsymbol{\Sigma}}(\beta)$ is given by

$$\tilde{\boldsymbol{\Sigma}}(\beta) = (1-\beta)\frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\boldsymbol{\Sigma}_0^{-1}\boldsymbol{z}_t} + \beta\boldsymbol{I}_M. \tag{26}$$

We will also consider regularized TVAR(m) estimates, namely the Dym-Gohberg regularization of (21)

$$\hat{\boldsymbol{\Sigma}}^{(m)}_{\text{DG}-\text{DL}}(\beta) = \mathsf{DG}^{(m)}\left[\hat{\boldsymbol{\Sigma}}_{\text{DL}}(\beta)\right] \tag{27}$$

where $\mathsf{DG}^{(m)}[.]$ is the Dym-Gohberg band-inverse transformation of a Hermitian non negative definite matrix, defined as [40]

$$\left\{\mathsf{DG}^{(m)}[\boldsymbol{R}]\right\}_{i,j} = \boldsymbol{R}_{i,j} \qquad |i-j| \leq m \tag{28a}$$

$$\left\{\left(\mathsf{DG}^{(m)}[\boldsymbol{R}]\right)^{-1}\right\}_{i,j} = 0 \qquad |i-j| > m. \tag{28b}$$

Accordingly, we investigate the fixed point diagonally loaded TVAR(m) estimate [29] defined as $\hat{\boldsymbol{\Sigma}}^{(m)}(\beta) = \lim_{k\to\infty}\hat{\boldsymbol{\Sigma}}^{(m)}_k(\beta)$ (a formal proof of convergence of this iterative scheme is still an open issue) where

$$\breve{\boldsymbol{\Sigma}}^{(m)}_{k+1}(\beta) = \mathsf{DG}^{(m)}\left[(1-\beta)\frac{M}{T}\sum_{t=1}^{T}\frac{\boldsymbol{z}_t\boldsymbol{z}_t^H}{\boldsymbol{z}_t^H\left(\hat{\boldsymbol{\Sigma}}^{(m)}_k(\beta)\right)^{-1}\boldsymbol{z}_t} + \beta\boldsymbol{I}_M\right] \tag{29a}$$

$$\hat{\boldsymbol{\Sigma}}^{(m)}_{k+1}(\beta) = \frac{M}{\text{Tr}\{\breve{\boldsymbol{\Sigma}}^{(m)}_{k+1}(\beta)\}}\breve{\boldsymbol{\Sigma}}^{(m)}_{k+1}(\beta) \tag{29b}$$

$\hat{\boldsymbol{\Sigma}}^{(m)}(\beta)$ will be referred to as FP-DG-DL in the sequel. For those (fixed-point) diagonally loaded TVAR(m) estimates, the value of β is also selected according to the EL principle, i.e.,

$$\beta^{(m)}_{\text{EL}} = \arg\min_{\beta}\left|\left[LR^u_{ACG}\left(\hat{\boldsymbol{\Sigma}}^{(m)}(\beta)|\boldsymbol{Z}_T\right)\right]^{1/T} - \text{med}\left[\omega^u(LR|M,T)\right]\right|. \tag{30}$$

IV. Numerical Simulations

Similarly to [29], we consider the case of data distributed according to a multivariate Student t-distribution with d degrees of freedom:

$$f(\boldsymbol{x}_t|\boldsymbol{\Sigma}_0) \propto |\boldsymbol{\Sigma}_0|^{-1}\left[1 + d^{-1}\boldsymbol{x}_t^H\boldsymbol{\Sigma}_0^{-1}\boldsymbol{x}_t\right]^{-(M+d)}. \tag{31}$$

In all simulations below, we use $d = 1$. We consider a ULA with $M = 64$ elements. The true scatter matrix was considered to be as per AR(1) process

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

with $\rho_0 = 0.998$. The SNR loss factor

$$SNR_{\text{loss}} = \frac{\left(\boldsymbol{s}_0^H\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{s}_0\right)^2}{\left(\boldsymbol{s}_0^H\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}_0\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{s}_0\right)\left(\boldsymbol{s}_0^H\boldsymbol{\Sigma}_0^{-1}\boldsymbol{s}_0\right)} \tag{32}$$

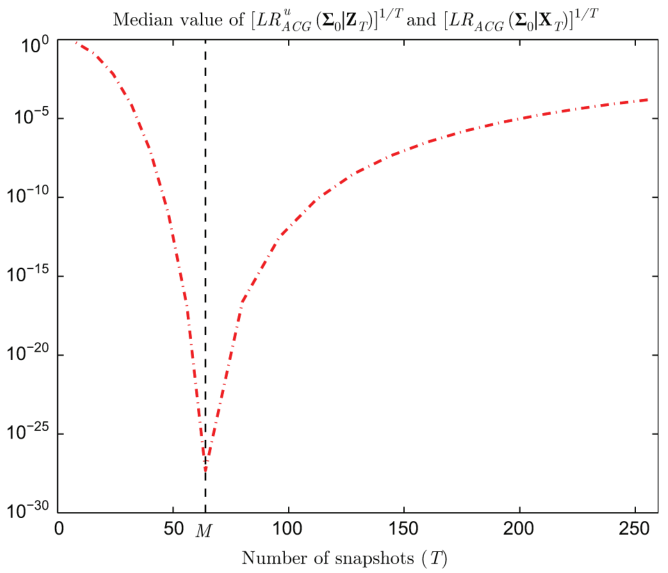

Fig. 1. Median value of $[LR^u_{ACG}(\mathbf{\Sigma}_0|\boldsymbol{Z}_T)]^{1/T}$ and $[LR_{ACG}(\mathbf{\Sigma}_0|\boldsymbol{X}_T)]^{1/T}$ versus T. $M = 64$.

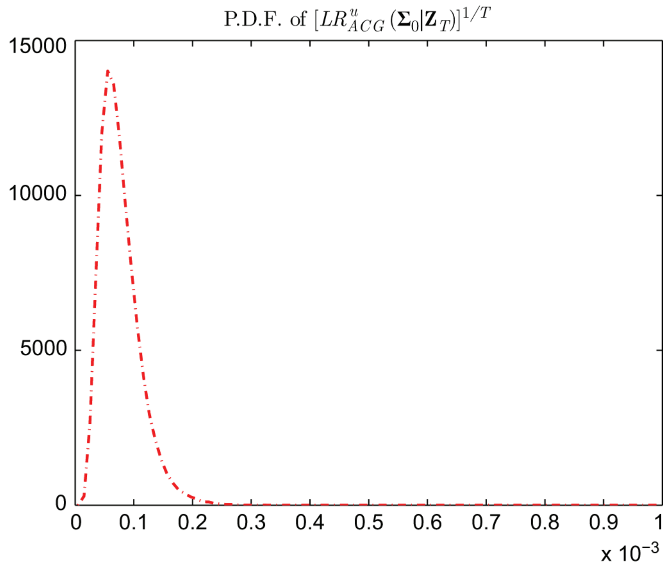

Fig. 3. Probability density function of $[LR^u_{ACG}(\mathbf{\Sigma}_0|\boldsymbol{Z}_T)]^{1/T}$. $M = 64$ and $T = 32$.

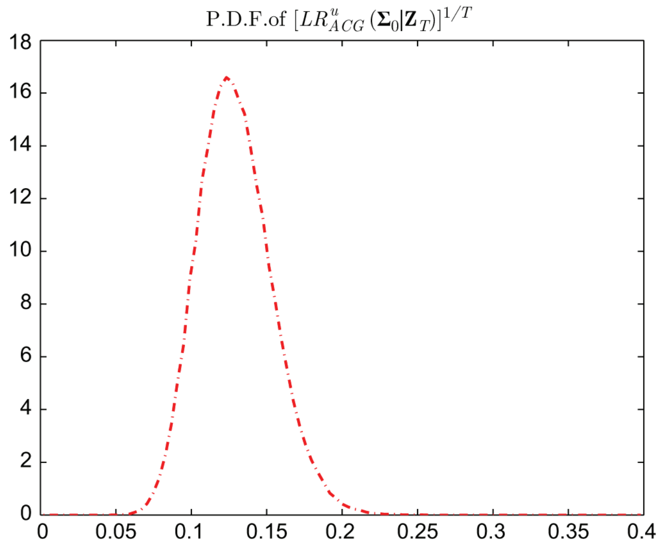

Fig. 2. Probability density function of $[LR^u_{ACG}(\mathbf{\Sigma}_0|\boldsymbol{Z}_T)]^{1/T}$. $M = 64$ and $T = 16$.

will serve as the figure of merit for quality assessment of the estimators. Above, $\hat{\mathbf{\Sigma}}$ is a generic SME and $\boldsymbol{s}_0 = \begin{bmatrix} 1 & e^{i\pi \sin\theta_0} & \cdots & e^{i\pi(M-1)\sin\theta_0}\end{bmatrix}^T$ stands for the steering vector corresponding to the looked direction θ_0 which is set to $\theta_0 = 1^\circ$.

We first examine the distribution $\omega^u(LR|M,T)$ of $[LR^u_{ACG}(\mathbf{\Sigma}_0|\boldsymbol{Z}_T)]^{1/T}$. Fig. 1 displays the median value of $\omega^u(LR|M,T)$ versus T: we also plot in this figure the mean value of $[LR_{ACG}(\mathbf{\Sigma}_0|\boldsymbol{Z}_T)]^{1/T}$ in the over-sampled case. This figure confirms that for $T = M$, the under-sampled and over-sampled median values coincide. As can be observed, the median value of $\omega^u(LR|M,T)$ decreases when T increases, achieves a minimal value for $T = M$ and then increases when $T \geq M$ increases. Figs. 2–3 display the p.d.f. $\omega^u(LR|M,T)$ for $T = 16$ and $T = 32$ respectively. As can be seen, $\omega^u(LR|M,T)$ can take very small values and, as T increases, the support of this p.d.f. is smaller.

Our second simulation deals with the influence of the loading factor β on the SNR loss as well as on the LR, see Fig. 4. As can be observed, the diagonally loaded estimates are not very sensitive to variations in β, at least when the SNR loss is concerned. Their LR however is seen to vary. In contrast, TVAR(m) estimates (especially DG-DL) have a SNR loss which exhibits large variations when β is varied: the latter should be chosen rather small in order to have a good SNR loss. One can also observe a correlation between SNR loss and LR: when β increases, both of them decrease. Whatever the estimate, it appears that choosing β according to the EL principle (23)–(30) results in negligible SNR losses, although the LR could be quite far from the median without penalizing too much SNR for the diagonally loaded estimates.

Fig. 5 displays SNR loss versus number of snapshots. The average value of the loading factor selected from the EL principle is also plotted, as is the average value of $[LR^u_{ACG}(\mathbf{\Sigma}(\beta_O)|\boldsymbol{Z}_T)]^{1/T}$ for the Oracle estimator. A few remarks are in order here. First, it can be seen that the LR for the Oracle estimator is close but slightly different from $\mathrm{med}[\omega^u(LR|M,T)]$: at least, it is not as close as in the over-sampled case. More important is the fact that the FP-DL with the EL principle for choosing β outperforms the Oracle estimator: this is due to the fact that EL selects a higher loading level, i.e., $\beta_{\mathrm{EL}} \geq \beta_O$, in order to have a lower LR. This is a quite remarkable result which shows that the minimization of the MSE between $\hat{\mathbf{\Sigma}}(\beta)$ and $\mathbf{\Sigma}_0$ does not result in the highest SNR in low sample support. As a second observation, notice that the FP diagonally loaded TVAR(1) estimate provides the highest SNR, which was also observed in the over-sampled case.

Similarly to Part 1, we now consider estimation of both m and β for FP $-$ DLTVAR(m) estimates. We use the same procedure as in [29]. For fixed m, we follow the rule in (30) to select β. Then, we estimate m as the minimal order for which

$$\left[LR^u_{ACG}\left(\hat{\mathbf{\Sigma}}^{(m)}(\beta^{(m)}_{\mathrm{EL}})|\boldsymbol{Z}_T\right)\right]^{1/T}$$

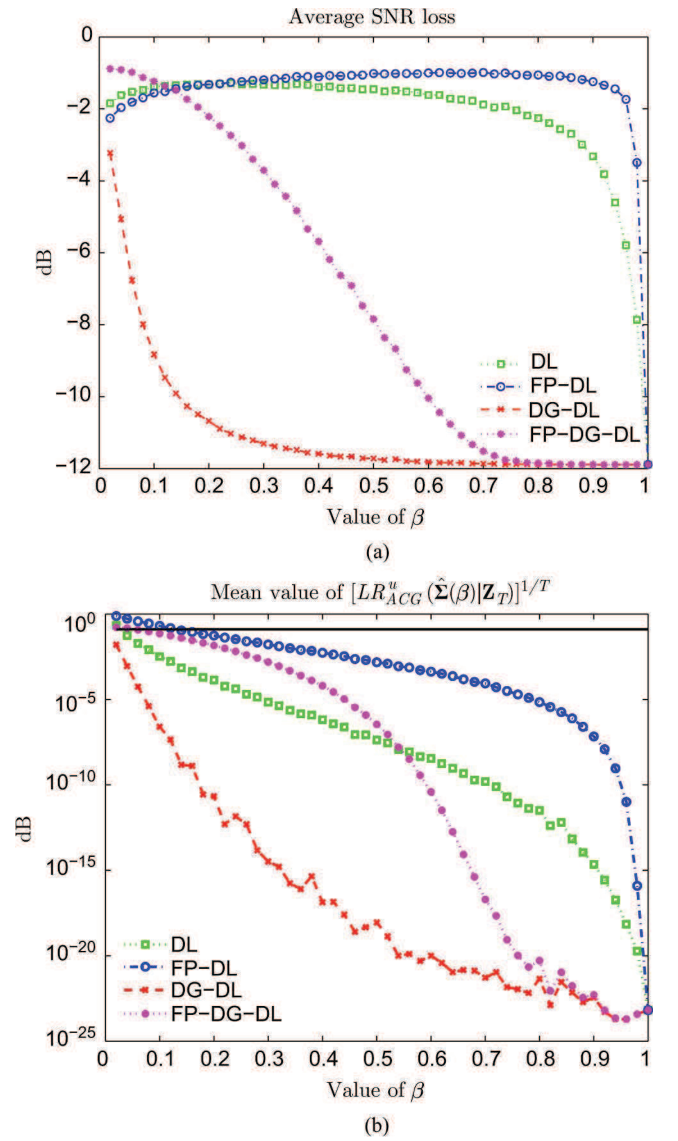

Fig. 4. Performance of diagonally loaded estimates versus β. $M = 64$ and $T = 16$. (a) SNR loss. (b) Mean value of $\left[LR^u_{ACG}\left(\hat{\mathbf{\Sigma}}(\beta)|\boldsymbol{Z}_T\right)\right]^{1/T}$.

is above a threshold:

$$\hat{m} = \min\left\{m/\left[LR^u_{ACG}\left(\hat{\mathbf{\Sigma}}^{(m)}(\beta^{(m)}_{\text{EL}})|\boldsymbol{Z}_T\right)\right]^{1/T} > \eta_L\right\} \quad (33)$$

where η_L is the 10% quantile of $\omega^u(LR|M,T)$, i.e., $\int_{\eta_L}^1 \omega(LR|M,T)dLR = 0.9$.. Since the minimum value of T is $T = 8$, m is necessarily in the interval $[1,7]$. If none of the orders yield a LR which exceeds the threshold, then we select the model order m which results in the LR closest to the median. As in Part 1, we still consider the case of an AR(1) scatter matrix $\mathbf{\Sigma}_0(k,\ell) = \rho_0^{|k-\ell|}$ and we also consider a case where the (k,ℓ) element of $\mathbf{\Sigma}_0$ corresponds to the $|k-\ell|$-th correlation lag of an ARMA(2,2) process whose spectrum (correlation) is close to but different from that of the AR(1) process. The SNR loss and average LR for the FP-DL, $\text{FP}-\text{DL}-\text{TVAR}(\hat{m})$ and $\text{FP}-\text{DL}-\text{TVAR}(m)$ are displayed in Fig. 6 for the AR(1) case and Fig. 7 for the

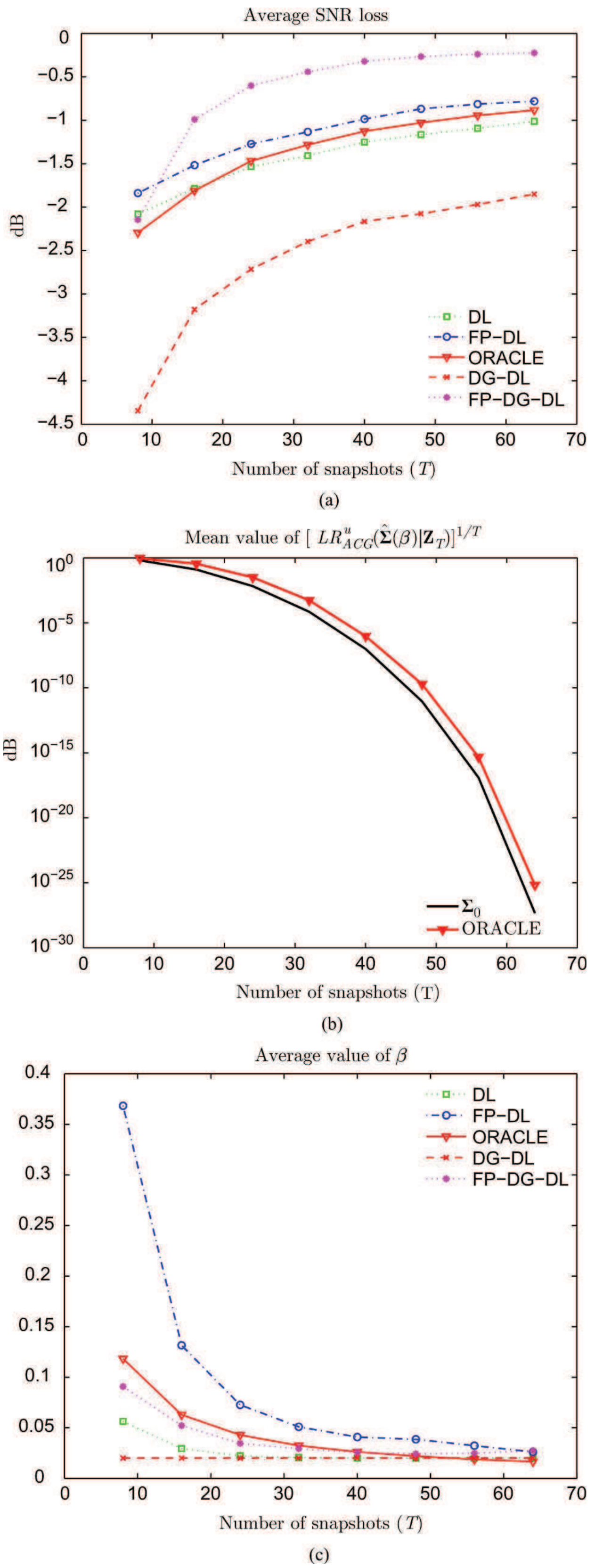

Fig. 5. Performance of regularized estimates versus number of snapshots T. $M = 64$. (a) SNR loss (b) Mean value of $\left[LR^u_{ACG}\left(\hat{\mathbf{\Sigma}}(\beta)|\boldsymbol{Z}_T\right)\right]^{1/T}$ (c) Mean value of loading factor.

$\text{ARMA}(2,2)$ case. In these figures, the two solid black lines represent the threshold η_L and $\text{med}\left[\omega^u(LR|M,T)\right]$. First, it is

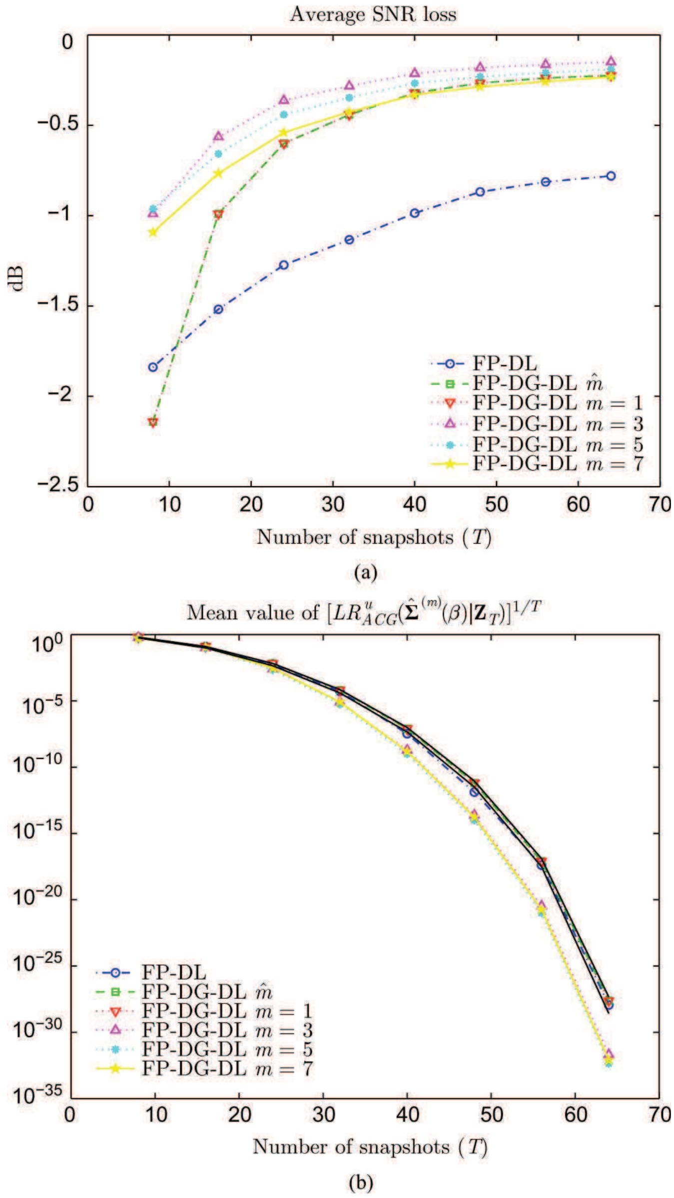

Fig. 6. Performance of fixed-point diagonally loaded $\mathrm{TVAR}(m)$ estimates versus number of snapshots T in the AR(1) case. $M = 64$. (a) SNR loss. (b) Mean value of $\left[LR^u_{ACG}\left(\hat{\mathbf{\Sigma}}^{(m)}(\beta)|\mathbf{Z}_T\right)\right]^{1/T}$.

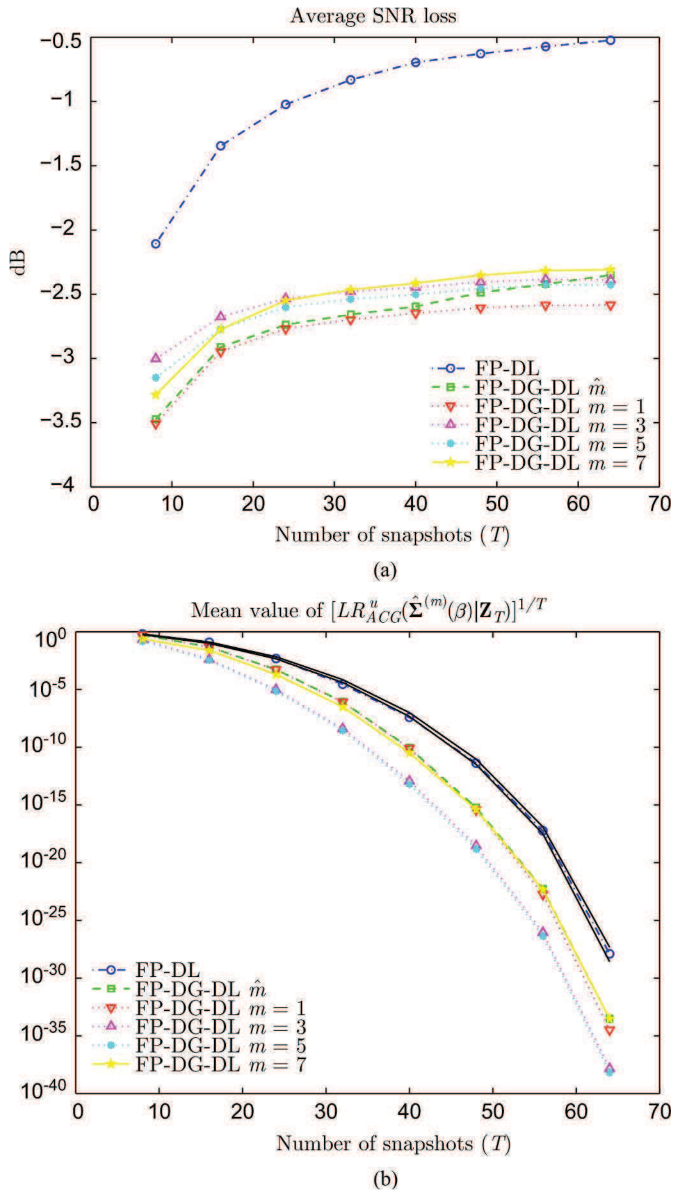

Fig. 7. Performance of fixed-point diagonally loaded $\mathrm{TVAR}(m)$ estimates versus number of snapshots T in the $\mathrm{ARMA}(2,2)$ case. $M = 64$. (a) SNR loss (b) Mean value of $\left[LR^u_{ACG}\left(\hat{\mathbf{\Sigma}}^{(m)}(\beta)|\mathbf{Z}_T\right)\right]^{1/T}$.

noteworthy that in the AR(1) case, the EL principle selects in the vast majority of cases $\hat{m} = 1$ which corresponds to the true model order. However, in contrast to the over-sampled case, this may not be the best choice as orders $m > 1$ results in better SNR at the price of lower LR. For instance, it seems that $m = 3$ yields the highest SNR but the corresponding LR is below the threshold η_L. Next, note that FP-DG-DL outperforms FP-DL, which is reasonable since $\mathbf{\Sigma}_0$ belongs to the class of $\mathrm{TVAR}(m)$ matrices. The $\mathrm{ARMA}(2,2)$ case yields different results. As noted in [29], FP-DL is now better than fixed-point diagonally loaded $\mathrm{TVAR}(m)$ estimates: the latter have lower SNR and LR which are below the threshold, yielding matrices that are not admissible. These two simulations confirm that FP-DL is an ubiquitous estimate which can accommodate various types of scatter matrices.

V. Conclusions

In this paper, we extended the EL approach of [30] to the class of CES and ACG distributions in the under-sampled case, where the number of samples T is less than the dimension M of the observation space. Together with the over-sampled case treated in Part 1 [29], this offers a general methodology to regularized scatter matrix estimation for a large and practically important class of distributions. We demonstrated that the LR evaluated at the true scatter matrix $\mathbf{\Sigma}_0$ still enjoys the same type of invariance properties that were found in the Gaussian case. This invariance makes it possible to assess the quality of any scatter matrix estimate, and a useful tool to tune the regularization parameters of regularized SME. This was demonstrated in the case of fixed-point diagonally loaded estimates, where the Oracle estimator was shown to achieve a LR very close to the median value of $LR(\mathbf{\Sigma}_0)$ which also corresponds to the target LR of the EL-based estimate. Accordingly, we developed regularized estimation schemes based on $\mathrm{TVAR}(m)$ modeling and investigated their use in conjunction with diagonal loading. For this shrinkage to the structure methodology, the EL approach was also efficient in providing estimates of both the model order and the loading factor that yields SNR values very close to that of the optimal (clairvoyant) filter. The framework and methodology of this two-part paper has been demonstrated for adaptive filtering, but it can also serve as a useful framework for other problems that call for fitting of a parametrically-controlled covariance or scatter matrix to under-sampled data.

Appendix
Likelihood Ratio for CES Distributions in the Under-Sampled Case

In this appendix, we derive the likelihood ratio for under-sampled training conditions in the case of CES distributions. Let us start with a r.v. $\boldsymbol{x} \sim \mathbb{C}\mathcal{E}_M(\mathbf{0}, \boldsymbol{\Sigma}, g)$ where $\boldsymbol{\Sigma}$ is a rank-r matrix which can be decomposed as $\boldsymbol{\Sigma} = \boldsymbol{U}_r \boldsymbol{D}_r \boldsymbol{U}_r^H$ where $\boldsymbol{U}_r$ is a $M \times r$ orthonormal matrix whose columns span the range space of $\boldsymbol{\Sigma}$ and $\boldsymbol{D}_r$ is a positive definite Hermitian $r \times r$ matrix. $\boldsymbol{x}$ can be represented as [27]

$$\boldsymbol{x} \overset{d}{=} \mathcal{R}\boldsymbol{U}_r \boldsymbol{B}_r \boldsymbol{u} \tag{34}$$

where $\boldsymbol{B}_r \boldsymbol{B}_r^H = \boldsymbol{D}_r$. Let $\boldsymbol{U}_r^\perp$ denote an orthonormal basis for the complement of $\boldsymbol{U}_r$ and let $\boldsymbol{U} = [\boldsymbol{U}_r \quad \boldsymbol{U}_r^\perp]$. Let us make the change of variables

$$\tilde{\boldsymbol{x}} = \boldsymbol{U}^H \boldsymbol{x} = \begin{bmatrix} \boldsymbol{U}_r^H \boldsymbol{x} \\ \left(\boldsymbol{U}_r^\perp\right)^H \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{x}}_r \\ \mathbf{0} \end{bmatrix} \overset{d}{=} \begin{bmatrix} \mathcal{R}\boldsymbol{B}_r \boldsymbol{u} \\ \mathbf{0} \end{bmatrix}.$$

Then $\tilde{\boldsymbol{x}}_r \sim \mathbb{C}\mathcal{E}_r(\mathbf{0}, \boldsymbol{D}_r, g)$ and its p.d.f. is given by

$$\begin{aligned} f(\tilde{\boldsymbol{x}}_r|\boldsymbol{D}_r) &\propto |\boldsymbol{D}_r|^{-1} g(\tilde{\boldsymbol{x}}_r^H \boldsymbol{D}_r^{-1} \tilde{\boldsymbol{x}}_r) \\ &\propto |\boldsymbol{D}_r|^{-1} g(\boldsymbol{x}^H \boldsymbol{U}_r \boldsymbol{D}_r^{-1} \boldsymbol{U}_r^H \boldsymbol{x}). \end{aligned}$$

Since the Jacobian from $\boldsymbol{x}$ to $\tilde{\boldsymbol{x}}$ is 1, one may define a singular CES density as

$$f(\boldsymbol{x}|\boldsymbol{U}_r, \boldsymbol{D}_r) \propto |\boldsymbol{D}_r|^{-1} g(\boldsymbol{x}^H \boldsymbol{U}_r \boldsymbol{D}_r^{-1} \boldsymbol{U}_r^H \boldsymbol{x}) \tag{35}$$

for vectors $\boldsymbol{x}$ such that $\boldsymbol{U}_r^{\perp H}\boldsymbol{x} = \mathbf{0}$. The joint density of a set of T independent snapshots $\boldsymbol{X}_T$ can thus be written as

$$f(\boldsymbol{X}_T|\boldsymbol{U}_r, \boldsymbol{D}_r) \propto |\boldsymbol{D}_r|^{-T} \prod_{t=1}^{T} g(\boldsymbol{x}_t^H \boldsymbol{U}_r \boldsymbol{D}_r^{-1} \boldsymbol{U}_r^H \boldsymbol{x}_t). \tag{36}$$

Assuming that $\boldsymbol{U}_r$ is known, for $T \geq r$, the MLE of $\boldsymbol{D}_r$ satisfies, see [27],

$$\hat{\boldsymbol{D}}_r^{\mathrm{ML}} = \frac{1}{T}\sum_{t=1}^{T} \phi(\boldsymbol{x}_t^H \boldsymbol{U}_r (\hat{\boldsymbol{D}}_r^{\mathrm{ML}})^{-1} \boldsymbol{U}_r^H \boldsymbol{x}_t) \boldsymbol{U}_r^H \boldsymbol{x}_t \boldsymbol{x}_t^H \boldsymbol{U}_r. \tag{37}$$

Let us now consider $T < M$ snapshots $\boldsymbol{x}_t \sim \mathbb{C}\mathcal{E}_M(\mathbf{0}, \boldsymbol{\Sigma}_0, g)$. As noted in the ACG case, inference about the scatter matrix $\boldsymbol{\Sigma}_0$ is possible only in the T-dimensional subspace spanned by the columns of the $M \times T$-variate matrix of eigenvectors $\hat{\boldsymbol{U}}_T$ associated with the T non-zero eigenvalues of the sample matrix $\boldsymbol{X}_T \boldsymbol{X}_T^H = \hat{\boldsymbol{U}}_T \hat{\boldsymbol{\Lambda}}_T \hat{\boldsymbol{U}}_T^H$. Again, for any given $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$, we need to find the rank-T Hermitian matrix $\boldsymbol{D}_T(\boldsymbol{\Omega}_\ell)$, such that the construct $\boldsymbol{\Sigma}_T(\boldsymbol{\Omega}_\ell) = \hat{\boldsymbol{U}}_T \boldsymbol{D}_T(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T^H$ is "closest" to the model $\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)$ which yields $\boldsymbol{D}_T(\boldsymbol{\Omega}_\ell) = \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T\right]^{-1}$ and $\boldsymbol{\Sigma}_T(\boldsymbol{\Omega}_\ell) = \hat{\boldsymbol{U}}_T \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T\right]^{-1} \hat{\boldsymbol{U}}_T^H$. From the previous definition of singular CES distributions, we may write the joint p.d.f. of $\boldsymbol{X}_T$ as

$$\begin{aligned} f(\boldsymbol{X}_T|\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell), g) &\propto |\boldsymbol{D}_T(\boldsymbol{\Omega}_\ell)|^{-T} \\ &\times \prod_{t=1}^{T} g(\boldsymbol{x}_t^H \hat{\boldsymbol{U}}_T \boldsymbol{D}_T(\boldsymbol{\Omega}_\ell)^{-1} \hat{\boldsymbol{U}}_T^H \boldsymbol{x}_t) \\ &\propto |\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T|^T \\ &\times \prod_{t=1}^{T} g(\boldsymbol{x}_t^H \hat{\boldsymbol{U}}_T \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T\right] \hat{\boldsymbol{U}}_T^H \boldsymbol{x}_t). \end{aligned} \tag{38}$$

In order to obtain the LR, we need to maximize $f(\boldsymbol{X}_T|\boldsymbol{\Sigma}, g)$ over the PDH matrix $\boldsymbol{D}_T = \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1} \hat{\boldsymbol{U}}_T\right]^{-1}$. As argued in (37), the MLE of $\boldsymbol{D}_T$ is the solution to

$$\hat{\boldsymbol{D}}_T^{\mathrm{ML}} = \frac{1}{T}\sum_{t=1}^{T} \phi(\boldsymbol{x}_t^H \hat{\boldsymbol{U}}_T (\hat{\boldsymbol{D}}_T^{\mathrm{ML}})^{-1} \hat{\boldsymbol{U}}_T^H \boldsymbol{x}_t) \hat{\boldsymbol{U}}_T^H \boldsymbol{x}_t \boldsymbol{x}_t^H \hat{\boldsymbol{U}}_T. \tag{39}$$

It follows that, for $T < M$, the under-sampled likelihood ratio is given by

$$\begin{aligned} LR_{CES}^{u}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{X}_T, g) &= |\hat{\boldsymbol{D}}_T^{\mathrm{ML}} \hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T|^T \\ &\times \prod_{t=1}^{T} \frac{g(\boldsymbol{x}_t^H \hat{\boldsymbol{U}}_T \left[\hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T\right] \hat{\boldsymbol{U}}_T^H \boldsymbol{x}_t)}{g(\boldsymbol{x}_t^H \hat{\boldsymbol{U}}_T (\hat{\boldsymbol{D}}_T^{\mathrm{ML}})^{-1} \hat{\boldsymbol{U}}_T^H \boldsymbol{x}_t)} \\ &= |\hat{\boldsymbol{D}}_T^{\mathrm{ML}} \hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \hat{\boldsymbol{U}}_T|^T \\ &\times \prod_{t=1}^{T} \frac{g(\boldsymbol{x}_t^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \boldsymbol{x}_t)}{g(\boldsymbol{x}_t^H \hat{\boldsymbol{U}}_T (\hat{\boldsymbol{D}}_T^{\mathrm{ML}})^{-1} \hat{\boldsymbol{U}}_T^H \boldsymbol{x}_t)} \\ &= \left[\prod_{\mathrm{eig}_j > 0} \mathrm{eig}_j\left(\hat{\boldsymbol{U}}_T \hat{\boldsymbol{D}}_T^{\mathrm{ML}} \hat{\boldsymbol{U}}_T^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)\right)\right]^T \\ &\times \prod_{t=1}^{T} \frac{g(\boldsymbol{x}_t^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \boldsymbol{x}_t)}{g(\boldsymbol{x}_t^H \hat{\boldsymbol{U}}_T (\hat{\boldsymbol{D}}_T^{\mathrm{ML}})^{-1} \hat{\boldsymbol{U}}_T^H \boldsymbol{x}_t)}. \end{aligned} \tag{40}$$

Let us now prove that, for $T = M$, the under-sampled LR (40) coincides with its over-sampled counterpart $LR_{CES}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{X}_T, g)$, which is given by [29]

$$\begin{aligned} LR_{CES}(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{X}_T, g) &= |\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)|^T \\ &\times \prod_{t=1}^{T} \frac{g(\boldsymbol{x}_t^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell) \boldsymbol{x}_t)}{g(\boldsymbol{x}_t^H \hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1} \boldsymbol{x}_t)} \end{aligned} \tag{41}$$

where $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ corresponds to the MLE of $\boldsymbol{\Sigma}$ and satisfies

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}} = \frac{1}{T}\sum_{t=1}^{T} \phi(\boldsymbol{x}_t^H \hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1} \boldsymbol{x}_t) \boldsymbol{x}_t \boldsymbol{x}_t^H \tag{42}$$

with $\phi(t) = -g'(t)/g(t)$. Similarly to the ACG case, we need to obtain the MLE $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ in this special case $T = M$. Let us then prove that for $T = M$

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}} = \hat{\boldsymbol{U}}_T \hat{\boldsymbol{D}}_T^{\mathrm{ML}} \hat{\boldsymbol{U}}_T^H \tag{43}$$

where $\hat{\boldsymbol{D}}_T^{\mathrm{ML}}$ is given in (39). First, observe that $\hat{\boldsymbol{U}}_T$ is a square $M \times M$ non-singular matrix so that $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ in (43) is non singular and its inverse is $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1} = \hat{\boldsymbol{U}}_T \left(\hat{\boldsymbol{D}}_T^{\mathrm{ML}}\right)^{-1} \hat{\boldsymbol{U}}_T^H$. Now, let us pre-multiply (39) by $\hat{\boldsymbol{U}}_T$ and post-multiply it by $\hat{\boldsymbol{U}}_T^H$ to obtain

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}} = \frac{1}{T}\sum_{t=1}^{T} \phi(\boldsymbol{x}_t^H \hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1}\boldsymbol{x}_t)\boldsymbol{x}_t\boldsymbol{x}_t^H \tag{44}$$

which coincides with (42). Using this expression in (40), we have that for $T = M$

$$LR_{CES}^{u}\left(\boldsymbol{\Sigma}(\boldsymbol{\Omega}_\ell)|\boldsymbol{X}_T, g\right) = |\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)|^T \prod_{t=1}^{T} \frac{g(\boldsymbol{x}_t^H \boldsymbol{\Sigma}^{-1}(\boldsymbol{\Omega}_\ell)\boldsymbol{x}_t)}{g(\boldsymbol{x}_t^H \hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}^{-1}\boldsymbol{x}_t)} \tag{45}$$

which is exactly the over-sampled likelihood ratio of (41).

Finally, let us investigate the distribution of $LR_{CES}^{u}\left(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g\right)$. Since $\boldsymbol{x}_t \overset{d}{=} \mathcal{Q}_t\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{u}_t$, it follows from (40) that

$$LR_{CES}^{u}\left(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g\right) \overset{d}{=} \left[\prod_{\mathrm{eig}_j>0} \mathrm{eig}_j\left(\boldsymbol{A}\right)\right]^T \prod_{t=1}^{T} \frac{g(\mathcal{Q}_t\boldsymbol{u}_t^H\boldsymbol{u}_t)}{g(\mathcal{Q}_t\boldsymbol{u}_t^H\boldsymbol{A}^-\boldsymbol{u}_t)} \tag{46}$$

where $\boldsymbol{A} \triangleq \boldsymbol{\Sigma}_0^{-1/2}\hat{\boldsymbol{U}}_T\hat{\boldsymbol{D}}_T^{\mathrm{ML}}\hat{\boldsymbol{U}}_T^H\boldsymbol{\Sigma}_0^{-1/2}$ and $\boldsymbol{A}^- \triangleq \boldsymbol{\Sigma}_0^{1/2}\hat{\boldsymbol{U}}_T(\hat{\boldsymbol{D}}_T^{\mathrm{ML}})^{-1}\hat{\boldsymbol{U}}_T^H\boldsymbol{\Sigma}_0^{1/2}$. Moreover, pre-multiplying (39) by $\boldsymbol{\Sigma}_0^{-1/2}\hat{\boldsymbol{U}}_T$ and post-multiply it by $\hat{\boldsymbol{U}}_T^H\boldsymbol{\Sigma}_0^{-1/2}$, it ensues that

$$\boldsymbol{A} = \frac{1}{T}\sum_{t=1}^{T}\phi(\mathcal{Q}_t\boldsymbol{u}_t^H\boldsymbol{A}^-\boldsymbol{u}_t)\mathcal{Q}_t\boldsymbol{u}_t\boldsymbol{u}_t^H. \tag{47}$$

Note that $\boldsymbol{A}^-$ is a generalized inverse of $\boldsymbol{A}$, i.e., $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$ and $\boldsymbol{A}^-\boldsymbol{A}\boldsymbol{A}^- = \boldsymbol{A}^-$. Unlike the Moore-Penrose pseudo-inverse, the generalized inverse is not unique. In this regard, note that $\hat{\boldsymbol{U}}_T(\hat{\boldsymbol{D}}_T^{\mathrm{ML}})^{-1}\hat{\boldsymbol{U}}_T^H$ is the unique Moore-Penrose pseudo-inverse of the matrix $\hat{\boldsymbol{U}}_T\hat{\boldsymbol{D}}_T^{\mathrm{ML}}\hat{\boldsymbol{U}}_T^H$. Therefore, by specifying a particular (Hermitian say) square root $\boldsymbol{\Sigma}_0^{1/2}$ of $\boldsymbol{\Sigma}_0$ we uniquely specify the matrices $\boldsymbol{A}$ and $\boldsymbol{A}^-$. Finally, from (47), the properties of the matrices $\boldsymbol{A}$ and $\boldsymbol{A}^-$ are entirely specified by a set of i.i.d complex uniform vectors $\boldsymbol{u}_t \sim \mathcal{U}\left(\mathbb{C}S^M\right)$. This means that the distribution of $\boldsymbol{A}$ does not depend on $\boldsymbol{\Sigma}_0$ but of course depends on $g(.)$, similarly to the over-sampled case. It results that the p.d.f. of $LR_{CES}^{u}\left(\boldsymbol{\Sigma}_0|\boldsymbol{X}_T, g\right)$ is independent of $\boldsymbol{\Sigma}_0$.

References

[1] 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.

[2] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, pp. 115–127, Mar. 1986.

[3] E. J. Kelly, "Adaptive detection in non-stationary interference, Part III," Massachusetts Inst. of Technol., Lincoln Lab., Lexington, MA, USA, Tech. Rep. 761, 1987.

[4] E. J. Kelly and K. M. Forsythe, "Adaptive detection and parameter estimation for multidimensional signal models," Massachusetts Inst. of Technol., Lincoln Lab., Lexington, MA, USA, Tech. Rep. 848, 1989.

[5] S. Kraut and L. L. Scharf, "The cfar adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2538–2541, Sep. 1999.

[6] S. Kraut, L. L. Scharf, and T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.

[7] S. Kraut, L. L. Scharf, and R. W. Butler, "The adaptive coherence estimator: A uniformly most powerful invariant adaptive detection statistic," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 427–438, Feb. 2005.

[8] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, Mar. 1963.

[9] C. G. Khatri, "Classical statistical analysis based on a certain multivariate complex Gaussian distribution," *Ann. Math. Statist.*, vol. 36, no. 1, pp. 98–114, Feb. 1965.

[10] R. J. Muirhead, *Aspects of Multivariate Statist. Theory*. New York, NY, USA: Wiley, 1982.

[11] L. L. Scharf, *Statist. Signal Processing: Detection, Estimation and Time Series Anal. Reading*. Reading, MA, USA: Addison-Wesley, 1991.

[12] S. M. Kay, *Fund. of Statist. Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[13] Y. I. Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum SNR," *Radio Eng. Elect. 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] 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.

[16] 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.

[17] 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.

[18] 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.

[19] E. Conte and M. Longo, "Characterisation of radar clutter as a spherically invariant process," *Proc. Inst. Electr. Eng.—Radar, Sonar, Navig.*, vol. 134, no. 2, pp. 191–197, Apr. 1987.

[20] 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.

[21] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847–1859, Dec. 2002.

[22] E. Conte, A. D. Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound-Gaussian process and its application to adaptive CFAR detection," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1908–1915, Aug. 2002.

[23] 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.

[24] 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.

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

[26] T. W. Anderson and K.-T Fang, "Theory and applications of elliptically contoured and related distributions," Statist. Dept., Stanford Univ., Stanford, CA, USA, Tech. Rep. 24, 1990.

[27] 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.

[28] 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.

[29] Y. I. Abramovich and O. Besson, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach—Part 1: The oversampled case," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5807–5818, 2013.

[30] Y. I. Abramovich and B. A. Johnson, "GLRT-based detection-estimation for undersampled training conditions," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3600–3612, Aug. 2008.

[31] B. Johnson and Y. Abramovich, “GlRT-based outlier prediction and cure in under-sampled training conditions using a singular likelihood ratio,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 15–20, 2007, pp. 1129–1132.
[32] Y. Abramovich and B. Johnson, “Use of an under-sampled likelihood ratio for GLRT and AMF detection,” in *Proc. IEEE Conf. Radar*, Apr. 24–27, 2006, pp. 539–545.
[33] Y. Abramovich and B. Johnson, “A modified likelihood ratio test for detection-estimation in under-sampled training conditions,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toulouse, France, May 14–19, 2006, pp. 1105–1108.
[34] D. E. Tyler, “Statistical analysis for the angular central Gaussian distribution on the sphere,” *Biometrika*, vol. 74, no. 3, pp. 579–589, Sep. 1987.
[35] Y. Chikuse, *Statistics on Special Manifolds*. New York, NY, USA: Springer-Verlag, 2003.
[36] D. E. Tyler, “A distribution-free M-estimator of multivariate scatter,” *Ann. Statist.*, vol. 15, no. 1, pp. 234–251, Mar. 1987.
[37] M. Siotani, T. Hayakawa, and Y. Fujikoto, *Modern Multivariate Statistical Analysis*. Cleveland, OH, USA: Amer. Science Press, 1985.
[38] Y. I. Abramovich and N. K. Spencer, “Diagonally loaded normalised sample matrix inversion (LNSMI) for outlier-resistant adaptive filtering,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, pp. 1105–1108.
[39] 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.
[40] 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.

Olivier Besson (SM’04) received the M. S. and PhD degrees in signal processing in 1988 and 1992 respectively, both from Institut National Polytechnique, Toulouse. He is currently a Professor with the Department of Electronics, Optronics and Signal of ISAE (Institut Supérieur de l’Aéronautique et de l’Espace), 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.

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, Ukraine, 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. From 2000, he was with the Australian Defence Science and Technology Organisation (DSTO), Adelaide, as principal research scientist, seconded to CSSIP until its closure. As of January 2012, Dr. Abramovich is with W.R. Systems Ltd., Fairfax, Virginia USA.

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.