Adaptive detection in elliptically distributed noise and under-sampled scenario

Olivier Besson, Yuri Abramovich

To link to this article: DOI: 10.1109/LSP.2014.2344695
URL: http://dx.doi.org/10.1109/LSP.2014.2344695

To cite this version: Besson, Olivier and Abramovich, Yuri *Adaptive detection in elliptically distributed noise and under-sampled scenario.* (2014) IEEE Signal Processing Letters, vol. 21 (n° 12). pp. 1531-1535. ISSN 1070-9908

Adaptive Detection in Elliptically Distributed Noise and Under-Sampled Scenario

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

***Abstract*—The problem of adaptive detection of a signal of interest embedded in elliptically distributed noise with unknown scatter matrix $\boldsymbol{R}$ is addressed, in the specific case where the number of training samples T is less than the dimension M of the observations. In this under-sampled scenario, whenever $\boldsymbol{R}$ is treated as an arbitrary positive definite Hermitian matrix, one cannot resort directly to the generalized likelihood ratio test (GLRT) since the maximum likelihood estimate (MLE) of $\boldsymbol{R}$ is not well-defined, the likelihood function being unbounded. Indeed, inference of $\boldsymbol{R}$ can only be made in the subspace spanned by the observations. In this letter, we present a modification of the GLRT which takes into account the specific features of under-sampled scenarios. We come up with a test statistic that, surprisingly enough, coincides with a subspace detector of Scharf and Friedlander: the detector proceeds in the subspace orthogonal to the training samples and then compares the energy along the signal of interest to the total energy. Moreover, this detector does not depend on the density generator of the noise elliptical distribution. Numerical simulations illustrate the performance of the test and compare it with schemes based on regularized estimates of $\boldsymbol{R}$.**

***Index Terms*—Adaptive detection, elliptically contoured distributions, generalized likelihood ratio test, under-sampled scenarios.**

I. Introduction and Problem Statement

We consider a conventional radar problem where presence of a target, with signature $\boldsymbol{v}$, is to be detected from the observation of the radar return $\boldsymbol{x} \in \mathbb{C}^M$ in a range cell under test (CUT) [1]. As usually done [1], we assume that a set of T training samples $\boldsymbol{x}_t, t = 1, \dots, T$, obtained from cells adjacent to the CUT, contain noise only, and share the same statistics as the noise in the CUT. The problem can be formulated as the following hypotheses testing problem [2]

$$\begin{aligned} H_0 &: \boldsymbol{x} = \boldsymbol{n}; \qquad \boldsymbol{x}_t = \boldsymbol{n}_t; \quad t = 1, \cdots, T \\ H_1 &: \boldsymbol{x} = \alpha \boldsymbol{v} + \boldsymbol{n}; \qquad \boldsymbol{x}_t = \boldsymbol{n}_t; \quad t = 1, \cdots, T \end{aligned} \tag{1}$$

where α denotes the unknown target complex amplitude, and $\boldsymbol{n}$ and $\boldsymbol{n}_t$ are independent and identically distributed (i.i.d.) random noise vectors. When the latter follow a complex Gaussian distribution and $T \geq M$, the generalized likelihood ratio test (GLRT) was derived and analyzed by Kelly [1]. An alternative 2-step GLRT solution was proposed in [3] which is referred to as the adaptive matched filter (AMF). However, when T is slightly larger but of the order of M, their performance has been observed to deteriorate. Moreover, when $T < M$, they cannot be implemented mainly because the maximum likelihood estimate (MLE) of the noise covariance matrix is not defined. To circumvent this problem, either a parametric model or regularization can be advocated. For instance, in [4], [5], a low-rank plus white noise structure is assumed for the noise covariance matrix and a fast ML estimator is derived, which in turn yields the GLRT under this assumption. Other parametric models can be used, e.g., Toeplitz matrices [6], autoregressive models [7], low-rank Toeplitz structures [8]. The alternative approach, namely regularization, is treated in detail in [9] where diagonally loaded versions of the AMF are proposed, i.e., the sample covariance matrix is regularized by adding a scaled identity matrix. The expected likelihood (EL) principle developed in [10] is then used to obtain a statistically sound value of the loading factor. Regularization is somehow a necessity when $T < M$ and the EL approach was extended to under-sampled Gaussian scenarios in [11] and used to provide efficient regularized estimates.

In this paper, we consider the problem in (1) for the specific case where $\boldsymbol{n}$ and $\boldsymbol{n}_t$ are i.i.d. random vectors drawn from a *complex multivariate elliptically contoured distribution* (ECD) [12], [13], see also the recent excellent survey in [14] for properties and applications of EC distributions. The literature about adaptive detection in ECD is much less abundant than that for the Gaussian case. In [15], Richmond showed that the GLRT in vector elliptical distributions is still Kelly's detector, and in [16, chapter 4] he investigated an AMF-type approach but faced the problem that the first step is not tractable unless the density generator of the ECD is known.

II. Modified GLRT

Assuming that $\boldsymbol{n}$ and $\boldsymbol{x}_t$ are i.i.d. random vectors drawn from a complex multivariate EC distribution with zero mean and positive definite scatter matrix $\boldsymbol{R}$, the joint distribution of $\boldsymbol{x}$ and $\boldsymbol{X} = [\,\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \dots \quad \boldsymbol{x}_T\,]$, under hypotheses $H_\ell, \ell \in \{0, 1\}$, is given by [13], [14]

$$\begin{aligned} p_\ell(\boldsymbol{x}, \boldsymbol{X} | \alpha, \boldsymbol{R}, g) \propto |\boldsymbol{R}|^{-(T+1)} \prod_{t=1}^{T} g(\boldsymbol{x}_t^H \boldsymbol{R}^{-1} \boldsymbol{x}_t) \\ \times g([\boldsymbol{x} - \ell \alpha \boldsymbol{v}]^H \boldsymbol{R}^{-1} [\boldsymbol{x} - \ell \alpha \boldsymbol{v}]) \end{aligned} \tag{2}$$

where $\propto$ means proportional to. In (2), $g : \mathbb{R}^+ \to \mathbb{R}^+$ is called the density generator and satisfies finite moment condition $\delta_{M,g} = \int_0^\infty t^{M-1} g(t) dt < \infty$ to ensure integrability of $p(\boldsymbol{x}_t | \boldsymbol{R}, g)$. Similarly to [14], we make the technical assumption that the function $t\varphi(t)$, where $\varphi(t) = -g'(t)/g(t)$, is non-decreasing. Moreover, we consider here the **under-sampled scenario** for which $T < M$: this scenario is very relevant in most applications, especially with large M.

Manuscript received June 13, 2014; revised July 21, 2014; accepted July 26, 2014. Date of publication August 05, 2014; date of current version August 07, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joseph Tabrikian.

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

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

As said previously, when no assumption is made on $\boldsymbol{R}$, i.e., it is a completely arbitrary positive definite Hermitian (PDH) matrix, then the maximum likelihood estimate (MLE) of $\boldsymbol{R}$ is not defined, since the likelihood function is unbounded above, see below for details. Usually, $\boldsymbol{R}$ is found under a set of constraints, typically assuming a parametric or regularized model $\boldsymbol{R}(\boldsymbol{\theta})$, which ensures that $\boldsymbol{\theta}$ is uniquely identifiable from T snapshots. However, whatever the model, arbitrary PDH $\boldsymbol{R}$ or $\boldsymbol{R}(\boldsymbol{\theta})$, inference about the scatter matrix can only be made in the low-dimensional subspace spanned by the observations, and this is inherent to the fact that $T < M$. While $\boldsymbol{R}$ cannot be recovered from its projection on this subspace, it may be possible that projection of $\boldsymbol{R}(\boldsymbol{\theta})$ enables to identify the latter. Herein, rather than enforcing a constraint on $\boldsymbol{R}$, we do not make any assumption and follow the lines of [11], viz. consider (a modification of) the generalized likelihood ratio test in the under-sampled case for arbitrary $\boldsymbol{R}$, knowing that identifiability issues exist.

Let us start with hypothesis H_0. Let $\tilde{\boldsymbol{X}} = [\boldsymbol{x} \quad \boldsymbol{X}] = \boldsymbol{U}_1\boldsymbol{\Lambda}_1^{1/2}\boldsymbol{V}^H$ denote the thin singular value decomposition of the whole data matrix, where $\boldsymbol{U}_1^H\boldsymbol{U}_1 = \boldsymbol{I}_{T+1}$ (the rank of $\tilde{\boldsymbol{X}}$ is $T+1$ with probability 1) and $\boldsymbol{V}$ is a unitary matrix. Let $\boldsymbol{U}_2$ be the $M \times (M-T-1)$ matrix whose columns form an orthonormal basis for the subspace orthogonal to $\boldsymbol{U}_1$ and let $\boldsymbol{U} = [\boldsymbol{U}_1 \quad \boldsymbol{U}_2]$. Let us rewrite $\boldsymbol{R}$ as

$$\boldsymbol{R} = [\boldsymbol{U}_1 \quad \boldsymbol{U}_2] \begin{bmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_1^H \\ \boldsymbol{U}_2^H \end{bmatrix} \tag{3}$$

and let us define $\boldsymbol{Q}_{1.2} = \boldsymbol{Q}_{11} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21}$ so that $\boldsymbol{R}$ can be parameterized by $\boldsymbol{Q}_{1.2}$, $\boldsymbol{Q}_{12}$ and $\boldsymbol{Q}_{22}$. Under H_0, (2) can be rewritten as

$$\begin{aligned} p_0(\boldsymbol{x}, \boldsymbol{X}|\boldsymbol{R}, g) &\propto |\boldsymbol{Q}_{22}|^{-(T+1)}|\boldsymbol{Q}_{1.2}|^{-(T+1)} \\ &\times \prod_{t=0}^{T} g(\boldsymbol{x}_t^H\boldsymbol{U}_1\boldsymbol{Q}_{1.2}^{-1}\boldsymbol{U}_1^H\boldsymbol{x}_t) \end{aligned} \tag{4}$$

where, with some slight abuse of notation, we set $\boldsymbol{x}_0 = \boldsymbol{x}$. Consider first maximization with respect to $\boldsymbol{Q}_{1.2}$. It is well known that the maximum likelihood estimator of $\boldsymbol{Q}_{1.2}$ is the solution to the following implicit equation [14]

$$\boldsymbol{Q}_{1.2} = \frac{1}{T+1}\sum_{t=0}^{T}\varphi(\boldsymbol{x}_t^H\boldsymbol{U}_1\boldsymbol{Q}_{1.2}^{-1}\boldsymbol{U}_1^H\boldsymbol{x}_t)\boldsymbol{U}_1^H\boldsymbol{x}_t\boldsymbol{x}_t^H\boldsymbol{U}_1. \tag{5}$$

Let us now show that the solution is given by $\boldsymbol{Q}_{1.2}^{\mathrm{ML}} = \gamma_\star^{-1}\boldsymbol{\Lambda}_1$ where $\gamma_\star$ is defined next. Indeed, let us investigate a solution of the form $\boldsymbol{Q}_{1.2} = \gamma^{-1}\boldsymbol{\Lambda}_1$. Since $\tilde{\boldsymbol{X}}^H\boldsymbol{U}_1\boldsymbol{\Lambda}_1^{-1}\boldsymbol{U}_1^H\tilde{\boldsymbol{X}} = \boldsymbol{V}\boldsymbol{V}^H = \boldsymbol{I}_{T+1}$ it follows that $\boldsymbol{x}_t^H\boldsymbol{U}_1\boldsymbol{\Lambda}_1^{-1}\boldsymbol{U}_1^H\boldsymbol{x}_t = 1$ and

$$\sum_{t=0}^{T}\varphi(\gamma\boldsymbol{x}_t^H\boldsymbol{U}_1\boldsymbol{\Lambda}_1^{-1}\boldsymbol{U}_1^H\boldsymbol{x}_t)\boldsymbol{U}_1^H\boldsymbol{x}_t\boldsymbol{x}_t^H\boldsymbol{U}_1 = \varphi(\gamma)\boldsymbol{\Lambda}_1. \tag{6}$$

It ensues that $\boldsymbol{Q}_{1.2}^{\mathrm{ML}} = \gamma_\star^{-1}\boldsymbol{\Lambda}_1$ where $\gamma_\star$ is the unique solution to $\gamma\varphi(\gamma) = T+1$. We would like to emphasize that $\gamma_\star$ *does only depend on* $g(.)$. Consequently, we get

$$\begin{aligned} \max_{\boldsymbol{Q}_{1.2}} p_0(\boldsymbol{x}, \boldsymbol{X}|\boldsymbol{R}, g) &\propto |\boldsymbol{Q}_{22}|^{-(T+1)}|\boldsymbol{\Lambda}_1|^{-(T+1)} \\ &\propto |\boldsymbol{Q}_{22}|^{-(T+1)}|\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^H|_+^{-(T+1)} \end{aligned} \tag{7}$$

where $|.|_+$ stands for the product of the positive eigenvalues of a matrix. Note that the maximum is achieved for any $\boldsymbol{Q}_{11}$, $\boldsymbol{Q}_{12}$ such that $\boldsymbol{Q}_{11} - \boldsymbol{Q}_{12}\boldsymbol{Q}_{22}^{-1}\boldsymbol{Q}_{21} = \gamma_\star^{-1}\boldsymbol{\Lambda}_1$ and hence with no loss of generality we can set $\boldsymbol{Q}_{12} = \boldsymbol{0}$ so that $\gamma_\star^{-1}\boldsymbol{\Lambda}_1$ is the MLE of $\boldsymbol{Q}_{11} = \boldsymbol{U}_1^H\boldsymbol{R}\boldsymbol{U}_1$. Observe also that (7) is unbounded above, implying that $\boldsymbol{Q}_{22} = \boldsymbol{U}_2^H\boldsymbol{R}\boldsymbol{U}_2$ is not identifiable.

Let us consider the equivalent problem under H_1. Using the same reasoning, it comes that

$$\begin{aligned} \max_{\boldsymbol{Q}_{1.2}} p_1(\boldsymbol{x}, \boldsymbol{X}|\alpha, \boldsymbol{R}, g) &\propto |\boldsymbol{Q}_{22}|^{-(T+1)} \\ &\times |(\boldsymbol{x} - \alpha\boldsymbol{v})(\boldsymbol{x} - \alpha\boldsymbol{v})^H + \boldsymbol{X}\boldsymbol{X}^H|_+^{-(T+1)} \end{aligned} \tag{8}$$

provided that the rank of the matrix $[\boldsymbol{x} - \alpha\boldsymbol{v} \quad \boldsymbol{X}]$ is $T+1$. For this latter condition not to be fulfilled, $\boldsymbol{x} - \alpha\boldsymbol{v}$ should belong to the range space of $\boldsymbol{X}$, or equivalently $\boldsymbol{v}$ would reside in the range space of $\tilde{\boldsymbol{X}}$, and the latter occurs with probability zero, unless $T+1 = M$. Consequently, the $(T+1)$-th root of the *modified generalized likelihood ratio* (MGLR) is defined as

$$MGLR^{1/(T+1)} = \frac{|\boldsymbol{x}\boldsymbol{x}^H + \boldsymbol{S}|_+}{\min_\alpha |(\boldsymbol{x} - \alpha\boldsymbol{v})(\boldsymbol{x} - \alpha\boldsymbol{v})^H + \boldsymbol{S}|_+} \tag{9}$$

where $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^H$. As said previously, the exact GLR is not well-defined since the likelihood is unbounded under both hypotheses due to the term $|\boldsymbol{Q}_{22}|^{-(T+1)}$. However, the latter is *the same* under both hypotheses and therefore, one may argue that $\boldsymbol{Q}_{22}$ should be selected the same under H_0 and H_1: anyhow nothing can be inferred about the projection of $\boldsymbol{R}$ onto the subspace orthogonal to the observations. With this constraint, (9) can be viewed as a likelihood ratio.

Before pursuing our derivation, it is noteworthy that *the above detector does not depend on* $g(.)$: hence, in under-sampled scenarios, the detector does not depend on the specific form of the density generator. In order to come up with the MGLRT, it now remains to solve the minimization problem at the denominator of (9). For any PDH matrix $\boldsymbol{A}$ whose rank is K and whose eigenvalues are $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K > \lambda_{K+1} = \ldots = \lambda_M = 0$, we have $|\boldsymbol{A}|_+ = \prod_{k=1}^{K}\lambda_k$ and $|\boldsymbol{A} + \epsilon\boldsymbol{I}| = \epsilon^{M-K}\prod_{k=1}^{K}(\lambda_k + \epsilon)$. Therefore, it follows that

$$|\boldsymbol{A}|_+ = \lim_{\epsilon\to 0}\epsilon^{-(M-K)}|\boldsymbol{A} + \epsilon\boldsymbol{I}|. \tag{10}$$

The previous equation is now used to obtain a more convenient expression for (9). Let $\boldsymbol{y}$ be an arbitrary vector in $\mathbb{C}^M$ and let

$$\boldsymbol{S} = \boldsymbol{V}\boldsymbol{L}\boldsymbol{V}^H = [\boldsymbol{V}_1 \quad \boldsymbol{V}_2] \begin{bmatrix} \boldsymbol{L}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1^H \\ \boldsymbol{V}_2^H \end{bmatrix} \tag{11}$$

denote the eigenvalue decomposition of $\boldsymbol{S}$ with $\boldsymbol{L}_1$ the $T \times T$ diagonal matrix of its eigenvalues. One can write

$$\begin{aligned} d(\epsilon) &= |\boldsymbol{S} + \boldsymbol{y}\boldsymbol{y}^H + \epsilon\boldsymbol{I}| \\ &= |\boldsymbol{L} + \epsilon\boldsymbol{I} + \boldsymbol{V}^H\boldsymbol{y}\boldsymbol{y}^H\boldsymbol{V}| \\ &= |\boldsymbol{L} + \epsilon\boldsymbol{I}||\boldsymbol{I} + (\boldsymbol{L} + \epsilon\boldsymbol{I})^{-1/2}\boldsymbol{V}^H\boldsymbol{y}\boldsymbol{y}^H\boldsymbol{V}(\boldsymbol{L} + \epsilon\boldsymbol{I})^{-1/2}| \\ &= \epsilon^{M-T}|\boldsymbol{L}_1 + \epsilon\boldsymbol{I}|[1 + \boldsymbol{y}^H\boldsymbol{V}(\boldsymbol{L} + \epsilon\boldsymbol{I})^{-1}\boldsymbol{V}^H\boldsymbol{y}] \\ &= \epsilon^{M-(T+1)}|\boldsymbol{L}_1 + \epsilon\boldsymbol{I}| \\ &\quad \times [\epsilon + \epsilon\boldsymbol{y}^H\boldsymbol{V}_1(\boldsymbol{L}_1 + \epsilon\boldsymbol{I})^{-1}\boldsymbol{V}_1^H\boldsymbol{y} + \boldsymbol{y}^H\boldsymbol{V}_2\boldsymbol{V}_2^H\boldsymbol{y}] \end{aligned} \tag{12}$$

where we used Sylvester's determinant theorem. It ensues that

$$\lim_{\epsilon \to 0} \epsilon^{-M+(T+1)} d(\epsilon) = |\boldsymbol{L}_1| (\boldsymbol{y}^H \boldsymbol{P}_{\boldsymbol{X}}^{\perp} \boldsymbol{y}) \tag{13}$$

where $\boldsymbol{P}_{\boldsymbol{X}}^{\perp} = \boldsymbol{V}_2 \boldsymbol{V}_2^H$ denote the projection onto the space orthogonal to $\boldsymbol{X}$. Equation (9) can thus be rewritten as

$$\begin{aligned} MGLR^{1/(T+1)} &= \frac{\boldsymbol{x}^H \boldsymbol{P}_{\boldsymbol{X}}^{\perp} \boldsymbol{x}}{\min_\alpha (\boldsymbol{x} - \alpha \boldsymbol{v})^H \boldsymbol{P}_{\boldsymbol{X}}^{\perp} (\boldsymbol{x} - \alpha \boldsymbol{v})} \\ &= \frac{\boldsymbol{x}^H \boldsymbol{P}_{\boldsymbol{X}}^{\perp} \boldsymbol{x}}{\boldsymbol{x}^H \boldsymbol{P}_{\boldsymbol{X}}^{\perp} \boldsymbol{x} - \boldsymbol{x}^H \boldsymbol{P}_{\boldsymbol{g}} \boldsymbol{x}} \equiv \frac{\boldsymbol{x}^H \boldsymbol{P}_{\boldsymbol{g}} \boldsymbol{x}}{\boldsymbol{x}^H \boldsymbol{P}_{\boldsymbol{X}}^{\perp} \boldsymbol{x}} \end{aligned} \tag{14}$$

where $\boldsymbol{g} = \boldsymbol{P}_{\boldsymbol{X}}^{\perp} \boldsymbol{v}$ and $\boldsymbol{P}_{\boldsymbol{g}} = \boldsymbol{P}_{\boldsymbol{P}_{\boldsymbol{X}}^{\perp} \boldsymbol{v}}$ stands for the orthogonal projector onto $\boldsymbol{g}$. Note that $\boldsymbol{g}$ is the part of $\boldsymbol{v}$ in the subspace orthogonal to $\boldsymbol{X}$. Some observations are noteworthy regarding (14). First, observe that the MGLR is a *ratio of energies*: the denominator is the energy of $\boldsymbol{x}$ in $\mathcal{R}(\boldsymbol{X})^{\perp}$ while the numerator measures the energy of $\boldsymbol{x}$ along the component of $\boldsymbol{v}$ within $\mathcal{R}(\boldsymbol{X})^{\perp}$. Note that, since the columns of $\boldsymbol{X}$ contain noise only, the detector operates in the subspace orthogonal to them and look there for energy along $\boldsymbol{v}$. Surprisingly enough, this detector *coincides with Scharf and Friedlander GLRT* for detecting a known signal ($\boldsymbol{v}$) in subspace interference (spanned by the columns of $\boldsymbol{X}$) and white noise of unknown level, see [17, eqn. (6.2)], i.e., it is the GLRT for the detection problem

$$\begin{aligned} H_0 &: \boldsymbol{x} \sim \mathbb{C}\mathcal{N}(\boldsymbol{X}\phi, \sigma^2 \boldsymbol{I}) \\ H_1 &: \boldsymbol{x} \sim \mathbb{C}\mathcal{N}(\alpha \boldsymbol{v} + \boldsymbol{X}\phi, \sigma^2 \boldsymbol{I}). \end{aligned} \tag{15}$$

This is quite a remarkable result as the initial detection problem is very different from (15). However, it is not illogical in view of the above discussion.

The *MGLR does not depend on the density generator* $g(.)$ of the CES distribution. This means, that in the under-sampled scenario, the underlying distribution of the data does not account as much as the subspace where the data lies. Additionally, the *probability density function (p.d.f.) of MGLR under* H_0 *does not depend on* $g(.)$. Indeed, under the stated assumptions about $\boldsymbol{x}$ and $\boldsymbol{x}_t$, they admit the following stochastic representation [13], [14] $\boldsymbol{x} \overset{d}{=} \mathcal{R}\boldsymbol{R}^{1/2}\boldsymbol{u}$ and $\boldsymbol{x}_t \overset{d}{=} \mathcal{R}_t \boldsymbol{R}^{1/2}\boldsymbol{u}_t$ where $\overset{d}{=}$ means "has the same distribution as". The non-negative real random variable $\mathcal{R} = \sqrt{\mathcal{Q}}$ is referred to as the modular variate and is independent of the complex random vector $\boldsymbol{u}$, which is uniformly distributed on the complex M-sphere. Using the fact that $\boldsymbol{u} \overset{d}{=} \frac{\boldsymbol{n}}{\sqrt{\mathbb{C}\chi^2_M}}$ where $\boldsymbol{n} \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, we can equivalently write $\boldsymbol{x} \overset{d}{=} \tilde{\mathcal{R}}\boldsymbol{R}^{1/2}\boldsymbol{n}$ and $\boldsymbol{x}_t \overset{d}{=} \tilde{\mathcal{R}}_t \boldsymbol{R}^{1/2}\boldsymbol{n}_t$. This implies that $\boldsymbol{X} \overset{d}{=} \boldsymbol{R}^{1/2}\boldsymbol{N}\boldsymbol{D}$ where $\boldsymbol{D} = \operatorname{diag}\{\tilde{\mathcal{R}}_t\}$. Consequently, the range space $\mathcal{R}(\boldsymbol{X})$ of $\boldsymbol{X}$ is the same as that of $\boldsymbol{R}^{1/2}\boldsymbol{N}$, which implies that $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{P}_{\boldsymbol{R}^{1/2}\boldsymbol{N}}$: hence, the distribution of $\boldsymbol{P}_{\boldsymbol{X}}^{\perp}$ does not depend on $g(.)$ and is the same as that when $\boldsymbol{X}$ is Gaussian distributed. The same reasoning holds for $\boldsymbol{P}_{\boldsymbol{g}}$ since $\boldsymbol{g} = \boldsymbol{P}_{\boldsymbol{X}}^{\perp}\boldsymbol{v}$. Finally, the ratio

$$\frac{\boldsymbol{x}^H \boldsymbol{P}_{\boldsymbol{g}} \boldsymbol{x}}{\boldsymbol{x}^H \boldsymbol{P}_{\boldsymbol{X}}^{\perp} \boldsymbol{x}} \overset{d}{=} \frac{\boldsymbol{n}^H \boldsymbol{R}^{1/2} \boldsymbol{P}_{\boldsymbol{g}} \boldsymbol{R}^{1/2} \boldsymbol{n}}{\boldsymbol{n}^H \boldsymbol{R}^{1/2} \boldsymbol{P}_{\boldsymbol{X}}^{\perp} \boldsymbol{R}^{1/2} \boldsymbol{n}}$$

does not depend on $\mathcal{R}$ and is independent of the density generator. Consequently, the distribution of $MGLR$ under H_0 is independent of $g(.)$.

Finally, note that the specific form in (14) is inherently due to the fact that we consider an arbitrary PDH matrix $\boldsymbol{R}$ rather than investigating directly a regularized or parameterized form for $\boldsymbol{R}$, e.g., as in [9]. It remains to evaluate the respective merit of each method, which is what we do in the next section through simulations.

III. Numerical Simulations

The modified GLR as given by (14) should be compared to possible competitors as well as a reference. The latter is provided by the matched filter (MF) which is given by [16]

$$\mathrm{MF}(\boldsymbol{x}|\boldsymbol{R}, g) = \frac{g\left(\boldsymbol{x}^H \boldsymbol{R}^{-1} \boldsymbol{x} - \frac{|\boldsymbol{x}^H \boldsymbol{R}^{-1} \boldsymbol{v}|^2}{\boldsymbol{v}^H \boldsymbol{R}^{-1} \boldsymbol{v}}\right)}{g(\boldsymbol{x}^H \boldsymbol{R}^{-1} \boldsymbol{x})}. \tag{16}$$

For a Student distribution with d degrees of freedom that will be considered in the sequel, the MF amounts to

$$\mathrm{MFS}(\boldsymbol{x}|\boldsymbol{R}) = \frac{|\boldsymbol{x}^H \boldsymbol{R}^{-1} \boldsymbol{v}|^2}{(\boldsymbol{v}^H \boldsymbol{R}^{-1} \boldsymbol{v})(d + \boldsymbol{x}^H \boldsymbol{R}^{-1} \boldsymbol{x})}. \tag{17}$$

The matched filter assumes that both $\boldsymbol{R}$ and the density generator $g(.)$ are known, which is unrealistic but can serve as a reference. For a fair comparison with MGLRT, some adaptive detection schemes ought to be investigated. A natural method is the AMF which consists in replacing $\boldsymbol{R}$ in (17) by an estimate based on $\boldsymbol{X}$. Since $T < M$, the sample covariance matrix obtained from $\boldsymbol{X}$ is singular and, hence, regularization is somehow needed. An effective approach, often advocated in Gaussian settings, consists in using shrinkage of the sample covariance matrix [9] as an estimate of $\boldsymbol{R}$. More precisely, one can consider

$$\mathrm{RAMFS_{DL}} = \frac{|\boldsymbol{x}^H \boldsymbol{R}_{\mathrm{DL}}^{-1}(\boldsymbol{X}, \beta_{\mathrm{DL}}) \boldsymbol{v}|^2}{(\boldsymbol{v}^H \boldsymbol{R}_{\mathrm{DL}}^{-1}(\boldsymbol{X}, \beta_{\mathrm{DL}}) \boldsymbol{v})[d + \boldsymbol{x}^H \boldsymbol{R}_{\mathrm{DL}}^{-1}(\boldsymbol{X}, \beta_{\mathrm{DL}}) \boldsymbol{x}]} \tag{18}$$

where $\boldsymbol{R}_{\mathrm{DL}}(\boldsymbol{X}, \beta) = (1-\beta)\frac{\boldsymbol{X}\boldsymbol{X}^H}{T} + \beta \boldsymbol{I}$. The parameter β_{DL} is selected according to the expected likelihood (EL) principle [11], i.e., β_{DL} is such that

$$[LR_G^u(\boldsymbol{R}_{\mathrm{DL}}(\beta_{\mathrm{DL}})|\boldsymbol{X})]^{1/T} \simeq \mathrm{med}[p_G^u(LR_G|M, T)] \tag{19}$$

where $LR_G^u(\boldsymbol{R}_{\mathrm{DL}}(\beta)|\boldsymbol{X})$ is the Gaussian likelihood ratio (LR) for testing $\mathcal{E}\{\boldsymbol{X}\boldsymbol{X}^H\} = \boldsymbol{R}_{\mathrm{DL}}(\beta)$ [11], $p_G^u(LR_G|M, T)$ is the scenario-invariant p.d.f. of the T-th root of $LR_G^u(\boldsymbol{R}|\boldsymbol{X})$ evaluated at the true covariance matrix $\boldsymbol{R}$, and $\mathrm{med}[.]$ stands for the median value. Another possibility is to consider the adaptive coherence estimator (ACE) [18], [19] (also called normalized AMF), or more precisely a regularized version of it

$$\mathrm{RACE_{DL}} = \frac{|\boldsymbol{x}^H \boldsymbol{R}_{\mathrm{DL}}^{-1}(\boldsymbol{X}, \beta_{\mathrm{DL}}) \boldsymbol{v}|^2}{(\boldsymbol{v}^H \boldsymbol{R}_{\mathrm{DL}}^{-1}(\boldsymbol{X}, \beta_{\mathrm{DL}}) \boldsymbol{v})(\boldsymbol{x}^H \boldsymbol{R}_{\mathrm{DL}}^{-1}(\boldsymbol{X}, \beta_{\mathrm{DL}}) \boldsymbol{x})}. \tag{20}$$

The use of $\boldsymbol{R}_{\mathrm{DL}}(\beta_{\mathrm{DL}})$ in (18) and (20) is suitable when $\boldsymbol{X}$ follows a Gaussian distribution, and indeed relies on this assumption. In order to account for a possibly complex elliptical distribution, yet without needing $g(.)$ to estimate $\boldsymbol{R}$, a pertinent choice is to normalize the training samples and proceed with $\boldsymbol{z}_t = \boldsymbol{x}_t / \|\boldsymbol{x}_t\|$ [20]. The vector $\boldsymbol{z}_t$ follows a complex angular central Gaussian (ACG) distribution [14], [21] which does no longer depend on $g(.)$. An efficient regularization scheme for estimating $\boldsymbol{R}$ from $\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}_1 & \ldots & \boldsymbol{z}_T \end{bmatrix}$ has been proposed in [22], [23], and referred to as the fixed point diagonally loaded estimator (FP-DL) in [20]. Let us denote it by $\boldsymbol{R}_{\text{FP-DL}}(\boldsymbol{Z}, \beta)$. The value of β can again be chosen according to the EL principle

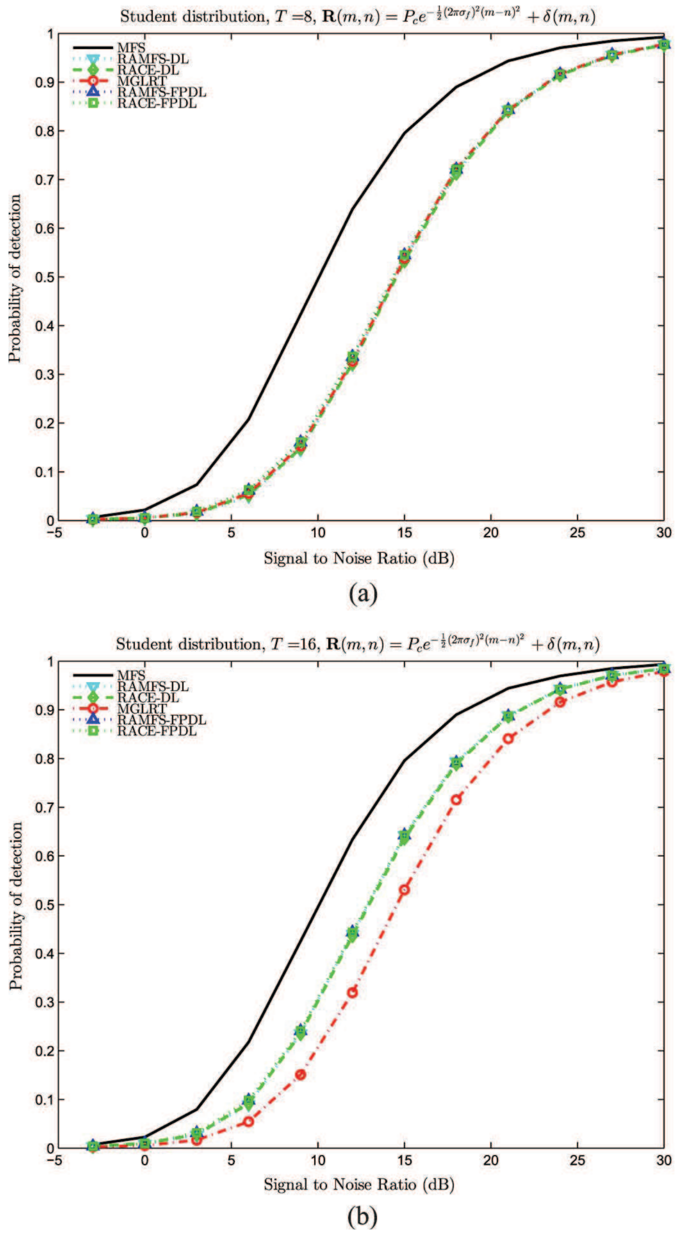

Fig. 1. Probability of detection of the various detectors versus SNR. (a) $T = 8$, (b) $T = 16$.

[20]. The matrix $\boldsymbol{R}_{\text{FP-DL}}(\boldsymbol{Z}, \beta_{\text{FP}-\text{DL}})$ can be used in place of $\boldsymbol{R}_{\text{DL}}(\boldsymbol{X}, \beta_{\text{DL}})$ in (18) and (20) to yield

$$\begin{aligned}&\text{RAMFS}_{\text{FP-DL}}\\ &= \text{MFS}(\boldsymbol{x}|\boldsymbol{R}_{\text{FP-DL}}(\boldsymbol{Z}, \beta_{\text{FP-DL}})) \qquad (21)\\ &\text{RACE}_{\text{FP-DL}}\\ &= \frac{|\boldsymbol{x}^H \boldsymbol{R}_{\text{FP-DL}}^{-1}(\boldsymbol{Z}, \beta_{\text{FP-DL}})\boldsymbol{v}|^2}{(\boldsymbol{v}^H \boldsymbol{R}_{\text{FP-DL}}^{-1}(\boldsymbol{Z}, \beta_{\text{FP-DL}})\boldsymbol{v})(\boldsymbol{x}^H \boldsymbol{R}_{\text{FP-DL}}^{-1}(\boldsymbol{Z}, \beta_{\text{FP-DL}})\boldsymbol{x})}. \qquad (22)\end{aligned}$$

The two adaptive detectors do not require knowledge of $g(.)$ to estimate $\boldsymbol{R}$ and (20) does not need it at all. Comparison of (18) and (21) will allow one to quantify the gain/loss achieved by using the normalized samples $\boldsymbol{z}_t$ instead of the original samples $\boldsymbol{x}_t$.

The above detectors are now compared under the following scenario. We consider an array of $M = 32$ elements separated a half-wavelength apart. The signal of interest impinges with a direction of arrival of $5°$ so that $\boldsymbol{v} = \boldsymbol{a}(5°)$ where $\boldsymbol{a}(\theta) = [1 \quad e^{i\pi \sin\theta} \quad \ldots e^{i\pi(M-1)\sin\theta}]^T$. We consider a covariance matrix of the type $\boldsymbol{R}(m,n) = P_c \exp\{-\frac{1}{2}(2\pi\sigma_f)^2(m-n)^2\} + \delta(m,n)$. The parameter σ_f is chosen as $\sigma_f = 0.01$. As for P_c, we set it to $P_c = 100$. The signal to noise ratio is defined as $SNR = |\alpha|^2 \boldsymbol{v}^H \boldsymbol{R}^{-1}\boldsymbol{v}$. We consider that the data follow Student distribution, with $d = 1$ degree of freedom. Finally, the probability of false alarm is set to $P_{fa} = 10^{-3}$. The results are shown in Fig. 1. The following main observations can be made.

When $T = 8$, MGLRT is an interesting solution as it performs as well as the best regularized schemes, yet with a lower computational complexity. When T increases, regularization appears to be more effective. In fact, MGLRT does not really improve when T increases while the other detectors tend to perform better.

Regularized estimation of $\boldsymbol{R}$ from either $\boldsymbol{x}_t$ or $\boldsymbol{z}_t = \boldsymbol{x}_t/||\boldsymbol{x}_t||$ results in the same probability of detection for RAMF and RACE, whatever the underlying distribution of $\boldsymbol{x}_t$. From EL point of view, any solution with the targeted LR value is statistically as "likely as the truth". Yet some of them could be better than others, if regularization matches the actual structure of the covariance matrix. For example, for AR covariance matrices, TVAR regularization can provide better detection performance than diagonal loading with the same EL threshold. Herein, the same type of regularization was used (diagonal loading) which results in similar performance for the direct and fixed point regularization. For under-sampled scenarios when the subspace spanned by the training data does not depend on the relative powers of the samples, this is not so surprising. In fact, it is reminiscent of the fact that made the MGLR detector does not dependent on $g(.)$. The conclusion might be different for over-sampled scenarios.

IV. Concluding Remarks

We investigated the adaptive detection of a signal of interest buried in noise with an elliptically contoured distribution, when the number of training samples T is much less than the size M of the observation space. In this under-sampled scenario, inference about the scatter matrix $\boldsymbol{R}$ can be made only in the T-dimensional subspace spanned by the training samples, and therefore considering $\boldsymbol{R}$ as arbitrary leads to identifiability issues. Since the latter are the same under both hypotheses, we proposed a modified GLR test. This test does not depend on the density generator of the noise EC distribution but is based on projection onto the subspace orthogonal to the training samples, followed by energy search along a direction, similarly to a test derived by Scharf and Friedlander in the context of matched subspace detection. The new detector is very simple from a computational point of view and it is seen to offer a valuable solution when T is very low. When T increases however, it appears that regularization offers a better solution.

References

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

[2] L. L. Scharf, *Statistical Signal Proc.: Detection, Estimation and Time Series Analysis*. Reading, MA, USA: Addison Wesley, 1991.

[3] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[4] M. Steiner and K. Gerlach, "Fast converging adaptive processor of a structured covariance matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1115–1126, Oct. 2000.

[5] K. Gerlach and M. J. Steiner, "Fast converging adaptive detection of Doppler-shifted range-distributed targets," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2686–2690, Sep. 2000.

[6] D. R. Fuhrmann, "Application of Toeplitz covariance estimation to adaptive beamforming and detection," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 39, no. 10, pp. 2194–2198, Oct. 1991.

[7] 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.

[8] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 501–515, Jan. 2014.

[9] 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.

[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 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.

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

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

[14] 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.

[15] C. D. Richmond, "A note on non-Gaussian adaptive array detection and signal parameter estimation," *IEEE Signal Process. Lett.*, vol. 3, no. 8, pp. 251–252, Aug. 1996.

[16] C. D. Richmond, "Adaptive array signal processing and performance analysis in non-Gaussian environments," Ph.D. dissertation, Mass. Inst. of Technol., Cambridge, MA, USA, 1996.

[17] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[18] L. L. Scharf and T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," in *Proc. 30th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 3–6, 1996, pp. 1114–1117.

[19] 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.

[20] O. Besson and Y. I. Abramovich, "Regularized covariance matrix estimation in complex elliptically symmetric distributions using the expected likelihood approach - Part 2: The under-sampled case," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5819–5829, Dec. 2013.

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

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

[23] 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.