DOI: 10.1109/TSP.2016.2633241
To cite this article: O. Besson, A. Coluccia, E. Chaumette, G. Ricci and F. Vincent, “Generalized Likelihood Ratio Test for Detection of Gaussian Rank-One Signals in Gaussian Noise With Unknown Statistics,” in IEEE Transactions on Signal Processing, vol. 65, no. 4, pp. 1082-1092, Feb. 15, 2017, doi: 10.1109/TSP.2016.2633241.

Generalized Likelihood Ratio Test for Detection of Gaussian Rank-One Signals in Gaussian Noise With Unknown Statistics

Olivier Besson, Angelo Coluccia, Eric Chaumette, Giuseppe Ricci, and François Vincent

*Abstract***—We consider the classical radar problem of detecting a target in Gaussian noise with unknown covariance matrix. In contrast to the usual assumption of deterministic target amplitudes, we assume here that the latter are drawn from a Gaussian distribution. The generalized likelihood ratio test (GLRT) is derived based on multiple primary data and a set of secondary data containing noise only. The new GLRT is shown to be the product of Kelly's GLRT and a corrective, data dependent term. We also investigate two-step approaches where the GLRT for a known disturbance covariance matrix is first derived. In order to come up with detectors that provide a good tradeoff between detection of matched signals and rejection of mismatched signals, we also investigate the two-step GLRT when a fictitious signal is included in the null hypothesis. The constant false alarm rate properties of the detectors are analyzed. Numerical simulations are presented, which show that for small sample sizes the newly-proposed GLRT can outperform Kelly's GLRT and, in addition, that detectors including a fictitious signal are very powerful, at least for low-to-intermediate clutter to noise ratio values.**

*Index Terms***—Adaptive detection, Gaussian rank-one signals, generalized likelihood ratio test.**

I. Problem Statement

RADAR systems are meant at detecting and tracking targets of interest in a possibly complicated noise environment, which is often unknown and must be learned from the data itself. A main objective is thus to produce detectors which offer good detection probability of matched signals with a constant false alarm rate (CFAR) so that their threshold can be set irrespective of the noise statistics, namely its covariance matrix in the case of Gaussian noise. In its more general form, the problem can be formulated as the following composite binary hypothesis test

$$H_0 : \begin{cases} \boldsymbol{x}_{t_p} = \boldsymbol{n}_{t_p};\ t_p = 1, \ldots, T_p \\ \boldsymbol{y}_{t_s} = \boldsymbol{n}_{T_p + t_s};\ t_s = 1, \ldots, T_s \end{cases}$$

$$H_1 : \begin{cases} \boldsymbol{x}_{t_p} = \alpha_{t_p} \boldsymbol{v} + \boldsymbol{n}_{t_p};\ t_p = 1, \ldots, T_p \\ \boldsymbol{y}_{t_s} = \boldsymbol{n}_{T_p + t_s};\ t_s = 1, \ldots, T_s \end{cases} \tag{1}$$

where $\boldsymbol{X} = \left[\boldsymbol{x}_1 \cdots \boldsymbol{x}_{T_p} \right] \in \mathbb{C}^{M \times T_p}$ stands for the observation matrix and corresponds to the radar returns at different range cells under test (case of range spread target [1]) or at a given range cell over multiple coherent processing intervals (CPI). The presence of a target is sought among $\boldsymbol{X}$ and $\boldsymbol{v}$ denotes its space or time or space-time signature. For instance, in the case of M pulses within the CPI, $\boldsymbol{v} = \left[1\ e^{i2\pi f_d}\ \ldots\ e^{i2\pi(M-1)f_d} \right]^T$ with $f_d = 2\lambda^{-1} v T_R$ the Doppler frequency and where v stands for the radial velocity of the target, λ is the wavelength of the radar, and T_R the pulse repetition period. α_{t_p} stands for the complex amplitude of the target and is assumed to be constant over the CPI but varies from one series of observations to the other. $\boldsymbol{n}_{t_p}$ corresponds to the additive noise, which is assumed to be zero-mean, complex Gaussian distributed with unknown positive definite covariance matrix $\boldsymbol{R} \in \mathbb{C}^{M \times M}$, which we denote as $\boldsymbol{n}_{t_p} \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$. Additionally, it is assumed that T_s snapshots $\boldsymbol{y}_{t_s}$ are available, which contain noise only, i.e., $\boldsymbol{y}_{t_s}$ are independent, zero-mean complex Gaussian vectors drawn from $\boldsymbol{y}_{t_s} \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$. These radar data can be collected from range cells in the vicinity of the cells under test (CUT) [2].

This problem was thoroughly investigated by Kelly in a series of technical reports and papers now become classic references [2]–[6]. Kelly's generalized likelihood ratio test (GLRT) set the pace and every newly developed detector since then has been quasi systematically compared to it. The GLR statistic in [2], [5] was obtained under the assumption that α_{t_p} are unknown deterministic quantities. For multiple primary data, it takes the following form

$$GLR_{\text{Kelly}}^{1/T_t} = \frac{|\boldsymbol{I}_{T_p} + \boldsymbol{X}^H \boldsymbol{S}_y^{-1} \boldsymbol{X}|}{|\boldsymbol{I}_{T_p} + \boldsymbol{X}^H \boldsymbol{S}_y^{-1/2} \boldsymbol{P}^{\perp}_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}} \boldsymbol{S}_y^{-1/2} \boldsymbol{X}|} \tag{2}$$

where $T_t = T_p + T_s$, $\boldsymbol{S}_y = \boldsymbol{Y}\boldsymbol{Y}^H$ is T_s times the sample covariance matrix of the secondary data $\boldsymbol{Y} = \left[\boldsymbol{y}_1 \ldots \boldsymbol{y}_{T_s} \right]$, $\boldsymbol{P}^{\perp}_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}}$ denotes the orthogonal projector onto the subspace orthogonal to $\boldsymbol{S}_y^{-1/2}\boldsymbol{v}$. Through the paper, $\boldsymbol{I}_{T_p}$ denotes the identity matrix of size T_p, $|.|$ stands for the determinant of a matrix and $.^H$ is the Hermitian (conjugate transpose) operator. Kelly provided a detailed statistical analysis of this detector both in

The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Amir Asif. The work of O. Besson and E. Chaumette was supported in part by the DGA/MRIS under Grant 2015.60.0090.00.470.75.01. The work of A. Coluccia was supported in part by Fondation ISAE-Supaéro. This paper was presented in part at the 6th IEEE Computational Advances in Multi-Sensor Adaptive Processing workshop, Cancun, Mexico, December 2015.

O. Besson, E. Chaumette, and F. Vincent are with Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-Supaéro), Toulouse 31055, France (e-mail: olivier.besson@isae-supaero.fr; eric.chaumette@isae.fr; francois.vincent@isae.fr).

A. Coluccia and G. Ricci are with the Dipartimento di Ingegneria dell'Innovazione, Università del Salento, Via Monteroni, Lecce 73100, Italy (e-mail: angelo.coluccia@unisalento.it; giuseppe.ricci@unisalento.it).

the case of matched and mismatched signature [4], [6]. Under the same assumption and in the case $T_p = 1$, Robey *et al.* derived the adaptive matched filter in [7]. This is indeed a two-step GLRT where at the first step $\boldsymbol{R}$ is assumed to be known (and the GLR is derived from $\boldsymbol{X}$ only), and at the second step $T_s^{-1}\boldsymbol{S}_y$ is substituted for $\boldsymbol{R}$.

Surprisingly enough, considering α_{t_p} as a random variable has received little attention, and the quasi totality of recent studies followed the lead of [2] and considered α_{t_p} as deterministic parameters. To the best of the authors knowledge, no references have addressed detection of a Gaussian signal in colored noise with unknown covariance matrix (while the case of white noise has been examined thoroughly). In [8], detection of an arbitrary Gaussian signal is addressed but this signal is not aligned on a known signature. However, a stochastic assumption for α_{t_p} makes sense to take into account the unpredictable fluctuation of the radar cross-section. Indeed, the widely accepted Swerling I-II target model [9], [10] corresponds to assuming that α_{t_p} are independent and drawn from a complex Gaussian distribution with zero mean and unknown variance P, i.e., $\alpha_{t_p} \sim \mathbb{C}\mathcal{N}(0, P)$. This is the approach we take in this paper. In fact, when $T_p = 1$, we have a constant but random amplitude along the CPI which corresponds to a Swerling I target, while for $T_p > 1$, the target amplitude varies randomly from CPI to CPI with no intra-pulse fluctuation, which leads to a Swerling II target. Compared to the classical "conditional" model where α_{t_p} are treated as deterministic unknowns, it may be felt that the "unconditional" model suffers some drawbacks. Firstly, a statistical assumption is made about α_{t_p} while this is not the case in the conditional model. Secondly, derivations in the conditional model involve a simple linear least-squares problem with respect to α_{t_p} while the derivations in the unconditional model are more complicated, see below. On the other hand, a drawback of the conditional model is that the number of unknowns grows with T_p while it is constant in the unconditional model. Therefore, an unconditional model is worthy of investigation and thus we address the equivalent of (1) in a stochastic framework, i.e., we consider the problem

$$H_0 : \begin{cases} \boldsymbol{X} \sim \mathbb{C}\mathcal{N}\left(\boldsymbol{0}, \boldsymbol{R}, \boldsymbol{I}_{T_p}\right) \\ \boldsymbol{Y} \sim \mathbb{C}\mathcal{N}\left(\boldsymbol{0}, \boldsymbol{R}, \boldsymbol{I}_{T_s}\right) \end{cases}$$
$$H_1 : \begin{cases} \boldsymbol{X} \sim \mathbb{C}\mathcal{N}\left(\boldsymbol{0}, \boldsymbol{R} + P\boldsymbol{v}\boldsymbol{v}^H, \boldsymbol{I}_{T_p}\right) \\ \boldsymbol{Y} \sim \mathbb{C}\mathcal{N}\left(\boldsymbol{0}, \boldsymbol{R}, \boldsymbol{I}_{T_s}\right) \end{cases} . \tag{3}$$

In (3), $\boldsymbol{Z} \in \mathbb{C}^{M\times T} \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}, \boldsymbol{\Omega})$ stands for the complex matrix-variate Gaussian distribution, given by $p(\boldsymbol{Z}) = \pi^{-MT}|\boldsymbol{\Sigma}|^{-T}|\boldsymbol{\Omega}|^{-M}\,\mathrm{etr}\left\{-\boldsymbol{Z}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}\boldsymbol{\Omega}^{-1}\right\}$, where $\mathrm{etr}\{.\}$ stands the exponential of the trace of a matrix. The main difference with the deterministic approach is that, under H_1, the signal of interest is embedded in the covariance matrix of $\boldsymbol{X}$ instead of in its mean value. In Section II, the exact GLRT for problem (3) is derived, see (4) below. Additionally, we investigate in Section III two-step approaches where $\boldsymbol{R}$ is first assumed to be known. Technical derivations are deferred to the appendices and we only state the main results, namely the expressions of the GLRT, see Propositions 1 and 2, and the analysis of the CFARness of all detectors. Numerical simulations are then reported in Section IV to illustrate the performance of the new detection schemes compared to Kelly's GLRT.

II. Generalized Likelihood Ratio Test

In this section, we derive the GLRT for the problem described in (3) and relate it to Kelly's GLRT in the deterministic case. The main result is stated in the following proposition.

Proposition 1: Suppose $T_s \geq M$. The GLR for the composite hypothesis testing problem in (3) is given by

$$\begin{aligned} GLR_{Gaussian}^{1/T_t} &= \frac{|\boldsymbol{I}_{T_p} + \boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}|}{|\boldsymbol{I}_{T_p} + \boldsymbol{X}^H\boldsymbol{S}_y^{-1/2}\boldsymbol{P}^{\perp}_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}}\boldsymbol{S}_y^{-1/2}\boldsymbol{X}|} \\ &\times \max_b \frac{|\boldsymbol{I}_{T_p} + (1+b)^{-1}\boldsymbol{X}^H\boldsymbol{S}_y^{-1/2}\boldsymbol{P}^{\perp}_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}}\boldsymbol{S}_y^{-1/2}\boldsymbol{X}|}{(1+b)^{T_p/T_t}|\boldsymbol{I}_{T_p} + (1+b)^{-1}\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}|} \\ &= \frac{\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}}{\boldsymbol{v}^H\left(\boldsymbol{S}_y + \boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}} \\ &\times \max_b \frac{\boldsymbol{v}^H\left(\boldsymbol{S}_y + (1+b)^{-1}\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}}{(1+b)^{T_p/T_t}\left(\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}\right)}. \end{aligned} \tag{4}$$

Proof: See Appendix A. ■

A few remarks are in order from the expression in (4):

- The GLR is a product of two terms, the first of which is recognized as Kelly's test statistic, i.e., the GLR for deterministic amplitudes α_{t_p}. The second term (which is always lower than one) is a corrective term due to the fact that now α_{t_p} are considered as Gaussian distributed random variables. Note that b corresponds to the signal-to-noise ratio and is proportional to P, see Appendix A. The numerator of the second term measures the gain of the filter $\left(\boldsymbol{S}_y + (1+b)^{-1}\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}$ towards the target. When the latter is absent or very weak, $\boldsymbol{X}$ contains noise only, and the filter is more efficient when $(1+b)^{-1}$ grows close to 1, or equivalently when b goes to 0: in this case, the estimated target power will be small. In contrast, if $\boldsymbol{X}$ contains a strong target, $\left(\boldsymbol{S}_y + (1+b)^{-1}\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}$ will tend to eliminate it if $(1+b)^{-1}$ is too large: in such a case, the estimated b should be large.
- Since the above GLR involves the same quantities as Kelly's GLR, it follows that is has a constant false alarm rate with respect to $\boldsymbol{R}$, i.e., its distribution under H_0 is independent of $\boldsymbol{R}$.
- The new detector requires solving the optimization problem in (4). In order to solve it efficiently, let us define $\eta = (1+b)^{-1} \in [0,1]$ and $\boldsymbol{S}_{xy} = \boldsymbol{S}_y + \boldsymbol{X}\boldsymbol{X}^H$. Then, if the determinant form is employed, one can make use of the fact that $|\boldsymbol{I} + \eta\boldsymbol{M}| = \prod_j\left[1 + \eta\lambda_j(\boldsymbol{M})\right]$ where $\lambda_j(\boldsymbol{M})$ are the eigenvalues of $\boldsymbol{M}$, to efficiently compute the function to be maximized with respect to η. Likewise, if the

TABLE I
VALUE OF λ_1 FOR THE CONSIDERED ASSUMPTIONS ON $\boldsymbol{u}$. $\lambda_{\max}(\boldsymbol{A}, \boldsymbol{B})$ STANDS FOR THE LARGEST GENERALIZED EIGENVALUE OF THE MATRIX PENCIL $(\boldsymbol{A}, \boldsymbol{B})$

$\boldsymbol{u} = \boldsymbol{0}$	$\boldsymbol{u} \perp \boldsymbol{v}$	$\boldsymbol{R}^{-1/2}\boldsymbol{u} \perp \boldsymbol{R}^{-1/2}\boldsymbol{v}$
1	$\lambda_{\max}\left(\boldsymbol{V}_\perp^H \boldsymbol{R}^{-1}\boldsymbol{S}_x \boldsymbol{R}^{-1}\boldsymbol{V}_\perp, \boldsymbol{V}_\perp^H \boldsymbol{R}^{-1}\boldsymbol{V}_\perp\right)$	$\lambda_{\max}\left(\boldsymbol{V}_\perp^H \boldsymbol{S}_x \boldsymbol{V}_\perp, \boldsymbol{V}_\perp^H \boldsymbol{R}\boldsymbol{V}_\perp\right)$

second form of the detector is used, one can notice that

$$\begin{aligned} f(\eta) =& \ \boldsymbol{v}^H \left(\boldsymbol{S}_y + \eta \boldsymbol{X}\boldsymbol{X}^H\right)^{-1} \boldsymbol{v} \\ =& \ \boldsymbol{v}^H \left(\boldsymbol{S}_{xy} + (\eta-1) \boldsymbol{X}\boldsymbol{X}^H\right)^{-1} \boldsymbol{v} \\ =& \ \boldsymbol{v}^H \boldsymbol{S}_{xy}^{-1}\boldsymbol{v} - (\eta-1)\boldsymbol{v}^H \boldsymbol{S}_{xy}^{-1}\boldsymbol{X} \\ & \left[\boldsymbol{I}_{T_p} + (\eta-1)\boldsymbol{X}^H \boldsymbol{S}_{xy}^{-1}\boldsymbol{X}\right]^{-1} \boldsymbol{X}^H \boldsymbol{S}_{xy}^{-1}\boldsymbol{v} \quad (5) \end{aligned}$$

which can be used, e.g., to compute efficiently $f(\eta)$ over a grid of values of η and solve the optimization problem.

- As a final remark, note that the GLRT exists under the weaker condition $T_p + T_s \geq M$, see equations (21) and (34) of Appendix A.

III. TWO-STEP APPROACHES

The previous GLR is usually referred to as a one-step GLRT as it is computed from both $\boldsymbol{X}$ and $\boldsymbol{Y}$ and maximization of the likelihood function is carried out with respect to all unknown parameters, namely P and $\boldsymbol{R}$ here. Similarly to what was proposed in [7], we now investigate two-step approaches where $\boldsymbol{R}$ is first assumed to be known, and the GLRT is derived based on $\boldsymbol{X}$ only. Then, an estimate of $\boldsymbol{R}$, based on $\boldsymbol{Y}$ only, is substituted for $\boldsymbol{R}$. This is the principle of the adaptive matched filter (AMF) of [7]. As was illustrated in the literature, AMF yields some loss compared to Kelly but it is not that significant, at least for T_s large enough and in the matched case where the assumed target signature coincides with the actual one. However, in the mismatched case, i.e., when $\boldsymbol{v}$ differs from the actual target signature, AMF lacks sensitivity, i.e., it still provides a good probability of detection even for non negligible mismatches. In order to overcome this drawback, a common approach consists in injecting a fictitious signal under H_0, "orthogonal" to $\boldsymbol{v}$, so that the detector is less inclined to decide in favor of H_1 in the case of signature mismatch, see e.g., [11]–[13]. We adopt the same philosophy here and consider the following detection problem

$$\begin{aligned} H_0 &: \ \boldsymbol{X} \sim \mathbb{C}\mathcal{N}\left(\boldsymbol{0}, \boldsymbol{R} + \boldsymbol{u}\boldsymbol{u}^H, \boldsymbol{I}_{T_p}\right) \\ H_1 &: \ \boldsymbol{X} \sim \mathbb{C}\mathcal{N}\left(\boldsymbol{0}, \boldsymbol{R} + P\boldsymbol{v}\boldsymbol{v}^H, \boldsymbol{I}_{T_p}\right). \end{aligned} \quad (6)$$

where the vector $\boldsymbol{u}$ can be either zero (similarly to AMF), orthogonal to $\boldsymbol{v}$, i.e., $\boldsymbol{u} \perp \boldsymbol{v}$ or orthogonal to $\boldsymbol{v}$ in the whitened space, i.e., $\boldsymbol{R}^{-1/2}\boldsymbol{u} \perp \boldsymbol{R}^{-1/2}\boldsymbol{v}$.

A. Expressions of the Two-Step Detectors

In Appendix B, we successively derive the three corresponding detectors and summarize the result below.

Proposition 2: Assuming that $\boldsymbol{R}$ is known, the GLR corresponding to the composite hypothesis testing problem (6) is given by

$$GLR_{|\boldsymbol{R}}(\boldsymbol{X}) = \frac{g\left(\max\left[\frac{\boldsymbol{v}^H \boldsymbol{R}^{-1}\boldsymbol{S}_x \boldsymbol{R}^{-1}\boldsymbol{v}}{(\boldsymbol{v}^H \boldsymbol{R}^{-1}\boldsymbol{v})T_p}, 1\right]\right)}{g\left(\max\left[\frac{\lambda_1}{T_p}, 1\right]\right)} \quad (7)$$

where $\boldsymbol{S}_x = \boldsymbol{X}\boldsymbol{X}^H$, $g(x) = x^{-T_p}\exp\{(x-1)T_p\}$, and the value of λ_1 is given in Table I.

In order to make the detectors adaptive (and assuming $T_s \geq M$), $T_s^{-1}\boldsymbol{S}_y$ should be substituted for $\boldsymbol{R}$ in (7). We will refer to the three detectors above as two-step GLRT, two-step ABORT and two-step WABORT, respectively.

B. Analysis

We now provide stochastic representations of the test statistics upon which the various detectors depend. Although they do not allow to provide expressions for the probability of false alarm P_{fa} and probability of detection P_d, they provide qualitative insights into the detectors properties. In particular, we show that the two-step GLRT and two-step WABORT are CFAR with respect to the covariance matrix $\boldsymbol{R}$, and that their probability of detection depends only on signal-to-noise ratio. In contrast, the two-step ABORT is not CFAR.

Let us first observe that the detectors are function of the following statistics

$$t_1(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\boldsymbol{v}^H \boldsymbol{S}_y^{-1}\boldsymbol{S}_x \boldsymbol{S}_y^{-1}\boldsymbol{v}}{\boldsymbol{v}^H \boldsymbol{S}_y^{-1}\boldsymbol{v}} \quad (8a)$$

$$t_2(\boldsymbol{X}, \boldsymbol{Y}) = \lambda_{\max}\left(\boldsymbol{V}_\perp^H \boldsymbol{S}_y^{-1}\boldsymbol{S}_x \boldsymbol{S}_y^{-1}\boldsymbol{V}_\perp, \boldsymbol{V}_\perp^H \boldsymbol{S}_y^{-1}\boldsymbol{V}_\perp\right) \quad (8b)$$

$$t_3(\boldsymbol{X}, \boldsymbol{Y}) = \lambda_{\max}\left(\boldsymbol{V}_\perp^H \boldsymbol{S}_x \boldsymbol{V}_\perp, \boldsymbol{V}_\perp^H \boldsymbol{S}_y \boldsymbol{V}_\perp\right). \quad (8c)$$

More precisely, the two-step GLRT depends on $t_1(\boldsymbol{X}, \boldsymbol{Y})$ only, while the two-step ABORT and two-step WABORT depend on the couples $(t_1(\boldsymbol{X}, \boldsymbol{Y}), t_2(\boldsymbol{X}, \boldsymbol{Y}))$ and $(t_1(\boldsymbol{X}, \boldsymbol{Y}), t_3(\boldsymbol{X}, \boldsymbol{Y}))$, respectively. Let $\boldsymbol{R} = \boldsymbol{G}\boldsymbol{G}^H$ denote a square-root factorization of $\boldsymbol{R}$. It follows that under both H_0 and H_1, $\boldsymbol{Y} \overset{d}{=} \boldsymbol{G}\boldsymbol{N}_y$ where $\overset{d}{=}$ means "distributed as" and $\boldsymbol{N}_y \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_M, \boldsymbol{I}_{T_s})$. Moreover, $\boldsymbol{X} \overset{d}{=} \boldsymbol{G}\boldsymbol{F}\boldsymbol{N}_x$ where $\boldsymbol{N}_x \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_M, \boldsymbol{I}_{T_p})$ and $\boldsymbol{F} = \boldsymbol{I}_M$ under H_0, $\boldsymbol{F}\boldsymbol{F}^H = \boldsymbol{I}_M + P\boldsymbol{G}^{-1}\boldsymbol{v}\boldsymbol{v}^H\boldsymbol{G}^{-H}$ under H_1, where $\boldsymbol{G}^{-H}$ is a short-hand notation for $(\boldsymbol{G}^H)^{-1}$. For any unitary matrix $\boldsymbol{Q}$, one can write that

$$\begin{aligned} \boldsymbol{S}_y^{-1}\boldsymbol{S}_x\boldsymbol{S}_y^{-1} \overset{d}{=}& \left(\boldsymbol{G}\boldsymbol{N}_y\boldsymbol{N}_y^H\boldsymbol{G}^H\right)^{-1}\left(\boldsymbol{G}\boldsymbol{F}\boldsymbol{N}_x\boldsymbol{N}_x^H\boldsymbol{F}^H\boldsymbol{G}^H\right) \\ & \left(\boldsymbol{G}\boldsymbol{N}_y\boldsymbol{N}_y^H\boldsymbol{G}^H\right)^{-1} \\ \overset{d}{=}& \ \boldsymbol{G}^{-H}\left(\boldsymbol{N}_y\boldsymbol{N}_y^H\right)^{-1} \\ & \left(\boldsymbol{F}\boldsymbol{N}_x\boldsymbol{N}_x^H\boldsymbol{F}^H\right)\left(\boldsymbol{N}_y\boldsymbol{N}_y^H\right)^{-1}\boldsymbol{G}^{-1} \\ \overset{d}{=}& \ \boldsymbol{G}^{-H}\boldsymbol{Q}\left[\boldsymbol{W}_y^{-1}\boldsymbol{W}_x\boldsymbol{W}_y^{-1}\right]\boldsymbol{Q}^H\boldsymbol{G}^{-1} \end{aligned} \quad (9)$$

where $\boldsymbol{W}_y = \boldsymbol{Q}^H \boldsymbol{N}_y \boldsymbol{N}_y^H \boldsymbol{Q}$ and $\boldsymbol{W}_x = \boldsymbol{Q}^H \boldsymbol{F} \boldsymbol{N}_x \boldsymbol{N}_x^H \boldsymbol{F}^H \boldsymbol{Q}$. Similarly, one has

$$\boldsymbol{S}_y^{-1} \overset{d}{=} \boldsymbol{G}^{-H} \boldsymbol{Q} \boldsymbol{W}_y^{-1} \boldsymbol{Q}^H \boldsymbol{G}^{-1}. \tag{10}$$

Let us consider the following unitary matrix

$$\boldsymbol{Q} = \begin{bmatrix} \frac{\boldsymbol{G}^{-1}\boldsymbol{v}}{\|\boldsymbol{G}^{-1}\boldsymbol{v}\|} & \boldsymbol{G}^H \boldsymbol{V}_\perp \boldsymbol{T}^{-H} \end{bmatrix} \tag{11}$$

where $\boldsymbol{V}_\perp^H \boldsymbol{R} \boldsymbol{V}_\perp = \boldsymbol{T}\boldsymbol{T}^H$. It is readily verified that $\boldsymbol{Q}^H \boldsymbol{G}^{-1} \boldsymbol{v} = (\boldsymbol{v}^H \boldsymbol{R}^{-1} \boldsymbol{v})^{1/2} \boldsymbol{e}_1$ where $\boldsymbol{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$. Consequently,

$$\begin{aligned} t_1(\boldsymbol{X}, \boldsymbol{Y}) &= \frac{\boldsymbol{v}^H \boldsymbol{S}_y^{-1} \boldsymbol{S}_x \boldsymbol{S}_y^{-1} \boldsymbol{v}}{\boldsymbol{v}^H \boldsymbol{S}_y^{-1} \boldsymbol{v}} \\ &\overset{d}{=} \frac{\boldsymbol{e}_1^T \left[\boldsymbol{W}_y^{-1} \boldsymbol{W}_x \boldsymbol{W}_y^{-1}\right] \boldsymbol{e}_1}{\boldsymbol{e}_1^T \boldsymbol{W}_y^{-1} \boldsymbol{e}_1}. \end{aligned} \tag{12}$$

Since $\boldsymbol{Q}^H \boldsymbol{N}_y \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_M, \boldsymbol{I}_{T_s})$ and $\boldsymbol{Q}^H \boldsymbol{F} \boldsymbol{N}_x \sim \mathbb{C}\mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^H \boldsymbol{F}\boldsymbol{F}^H \boldsymbol{Q}, \boldsymbol{I}_{T_p})$ with

$$\begin{aligned} \boldsymbol{Q}^H \boldsymbol{F}\boldsymbol{F}^H \boldsymbol{Q} &= \boldsymbol{Q}^H \left[\boldsymbol{I}_M + P\boldsymbol{G}^{-1}\boldsymbol{v}\boldsymbol{v}^H \boldsymbol{G}^{-H}\right] \boldsymbol{Q} \\ &= \boldsymbol{I}_M + P(\boldsymbol{v}^H \boldsymbol{R}^{-1} \boldsymbol{v}) \boldsymbol{e}_1 \boldsymbol{e}_1^T \end{aligned} \tag{13}$$

it follows that, under H_0 (corresponding to $P = 0$), the distribution of $t_1(\boldsymbol{X}, \boldsymbol{Y})$ does not depend on $\boldsymbol{R}$, while, under H_1, it depends only on $P(\boldsymbol{v}^H \boldsymbol{R}^{-1} \boldsymbol{v})$. As a consequence, the two-step GLRT is CFAR with respect to $\boldsymbol{R}$ and its probability of detection depends only on the signal-to-noise ratio $\mathrm{SNR} = P(\boldsymbol{v}^H \boldsymbol{R}^{-1} \boldsymbol{v})$.

Let us consider now the two-step WABORT, whose distribution depends on the joint distribution of $(t_1(\boldsymbol{X}, \boldsymbol{Y}), t_3(\boldsymbol{X}, \boldsymbol{Y}))$. Considering the same transformation $\boldsymbol{Q}$ as for $t_1(\boldsymbol{X}, \boldsymbol{Y})$ and observing that

$$\boldsymbol{Q}^H \boldsymbol{G}^H \boldsymbol{V}_\perp = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{T}^H \end{bmatrix} = \boldsymbol{E} \tag{14}$$

it ensues that

$$\begin{aligned} \boldsymbol{V}_\perp^H \boldsymbol{S}_x \boldsymbol{V}_\perp &\overset{d}{=} \boldsymbol{V}_\perp^H \left(\boldsymbol{G}\boldsymbol{F}\boldsymbol{N}_x \boldsymbol{N}_x^H \boldsymbol{F}^H \boldsymbol{G}^H\right) \boldsymbol{V}_\perp \\ &\overset{d}{=} \boldsymbol{V}_\perp^H \boldsymbol{G}\boldsymbol{Q}\boldsymbol{W}_x \boldsymbol{Q}^H \boldsymbol{G}^H \boldsymbol{V}_\perp \\ &\overset{d}{=} \boldsymbol{E}^H \boldsymbol{W}_x \boldsymbol{E} \\ &\overset{d}{=} \boldsymbol{T} \left[\boldsymbol{W}_x\right]_{22} \boldsymbol{T}^H \end{aligned} \tag{15a}$$

$$\begin{aligned} \boldsymbol{V}_\perp^H \boldsymbol{S}_y \boldsymbol{V}_\perp &\overset{d}{=} \boldsymbol{V}_\perp^H \left(\boldsymbol{G}\boldsymbol{N}_y \boldsymbol{N}_y^H \boldsymbol{G}^H\right) \boldsymbol{V}_\perp \\ &\overset{d}{=} \boldsymbol{V}_\perp^H \boldsymbol{G}\boldsymbol{Q}\boldsymbol{W}_y \boldsymbol{Q}^H \boldsymbol{G}^H \boldsymbol{V}_\perp \\ &\overset{d}{=} \boldsymbol{E}^H \boldsymbol{W}_y \boldsymbol{E} \\ &\overset{d}{=} \boldsymbol{T} \left[\boldsymbol{W}_y\right]_{22} \boldsymbol{T}^H \end{aligned} \tag{15b}$$

where $[.]_{22}$ stands for the lower-right $(M-1) \times (M-1)$ corner of the matrix between brackets. Moreover, the generalized eigenvalues of the matrix pencil $(\boldsymbol{T}\boldsymbol{A}_{22}\boldsymbol{T}^H, \boldsymbol{T}\boldsymbol{B}_{22}\boldsymbol{T}^H)$ are the same as those of the matrix pencil $(\boldsymbol{A}_{22}, \boldsymbol{B}_{22})$ and therefore

$$t_3(\boldsymbol{X}, \boldsymbol{Y}) \overset{d}{=} \lambda_{\max}\left(\left[\boldsymbol{W}_x\right]_{22}, \left[\boldsymbol{W}_y\right]_{22}\right). \tag{16}$$

Now, from (13), $\left[\boldsymbol{Q}^H \boldsymbol{F}\boldsymbol{F}^H \boldsymbol{Q}\right]_{22} = \boldsymbol{I}_{M-1}$ and, hence, the distribution of $t_3(\boldsymbol{X}, \boldsymbol{Y})$ is the same and is independent of $\boldsymbol{R}$ under both H_0 and H_1. Moreover, the stochastic representations of $t_1(\boldsymbol{X}, \boldsymbol{Y})$ in (12) and $t_3(\boldsymbol{X}, \boldsymbol{Y})$ in (16) imply that the joint distribution of $(t_1(\boldsymbol{X}, \boldsymbol{Y}), t_3(\boldsymbol{X}, \boldsymbol{Y}))$ does not depend on $\boldsymbol{R}$ under H_0 and depends on SNR only under H_1. Finally, this proves that the two-step WABORT is also CFAR with a probability of detection that depends on SNR only.

Let us finally consider the two-step ABORT and thus the joint distribution of $(t_1(\boldsymbol{X}, \boldsymbol{Y}), t_2(\boldsymbol{X}, \boldsymbol{Y}))$. One can write

$$\begin{aligned} \boldsymbol{V}_\perp^H \boldsymbol{S}_y^{-1} \boldsymbol{S}_x \boldsymbol{S}_y^{-1} \boldsymbol{V}_\perp \overset{d}{=} \boldsymbol{V}_\perp^H \boldsymbol{G}^{-H} \boldsymbol{Q} \left[\boldsymbol{W}_y^{-1} \boldsymbol{W}_x \boldsymbol{W}_y^{-1}\right] \\ \boldsymbol{Q}^H \boldsymbol{G}^{-1} \boldsymbol{V}_\perp \end{aligned} \tag{17a}$$

$$\boldsymbol{V}_\perp^H \boldsymbol{S}_y^{-1} \boldsymbol{V}_\perp \overset{d}{=} \boldsymbol{V}_\perp^H \boldsymbol{G}^{-H} \boldsymbol{Q} \boldsymbol{W}_y^{-1} \boldsymbol{Q}^H \boldsymbol{G}^{-1} \boldsymbol{V}_\perp. \tag{17b}$$

Now, one has

$$\boldsymbol{Q}^H \boldsymbol{G}^{-1} \boldsymbol{V}_\perp = \begin{bmatrix} \boldsymbol{v}^H \boldsymbol{R}^{-1} \boldsymbol{V}_\perp / \left\|\boldsymbol{G}^{-1}\boldsymbol{v}\right\| \\ \boldsymbol{T}^{-1} \end{bmatrix} \tag{18}$$

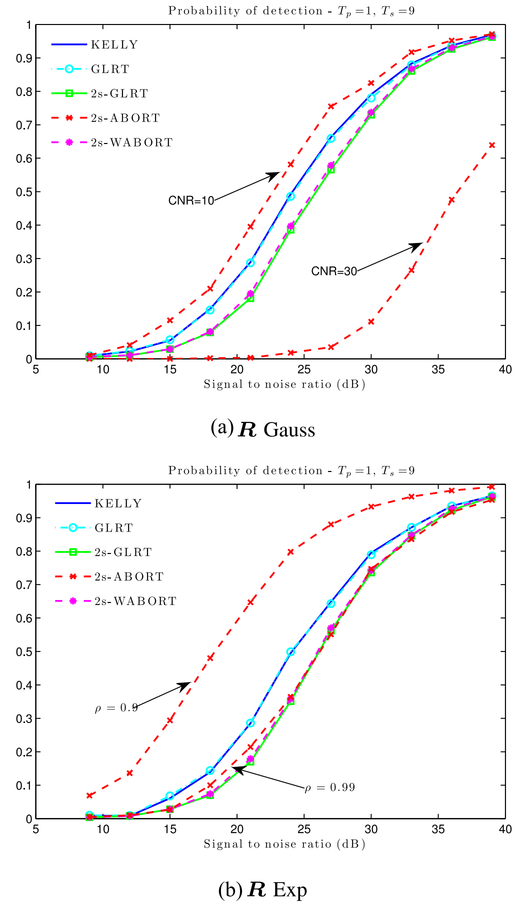
which shows that the joint distribution of $(t_1(\boldsymbol{X}, \boldsymbol{Y}), t_2(\boldsymbol{X}, \boldsymbol{Y}))$ depends on $\boldsymbol{R}$ and, consequently, the two-step ABORT is not CFAR.

Summarizing the results, we have that

1) the two-step GLRT and the two-step WABORT possess the CFAR property while the two-step ABORT does not;
2) the probability of detection of the two-step GLRT and the two-step WABORT depends on signal-to-noise ratio only.

IV. Numerical Simulations

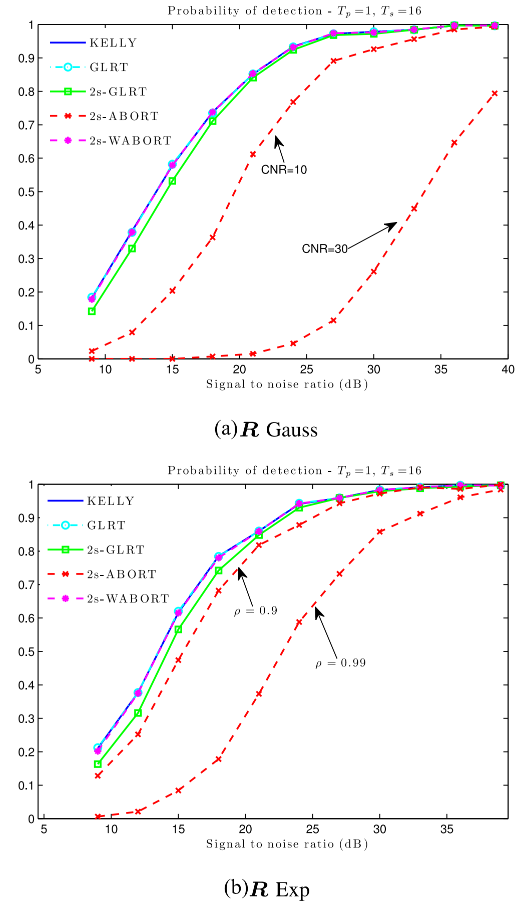
We now provide numerical illustrations of the performance of the new detectors and compare them with Kelly's GLRT. We consider a radar scenario with $M = 8$ pulses. The target signature is given by $\boldsymbol{v} = \begin{bmatrix} 1 & e^{i2\pi f_d} & \ldots & e^{i2\pi(M-1)f_d} \end{bmatrix}^T$ with $f_d = 0.09$, a value such that the target competes with noise. The amplitudes α_{t_p} are drawn from a complex Gaussian distribution with zero mean and power P. The signal-to-noise ratio is defined as $\mathrm{SNR} = P\boldsymbol{v}^H \boldsymbol{R}^{-1} \boldsymbol{v}$. In contrast to all other detectors the probability of false alarm and probability of detection of the two-step ABORT detector depend on $\boldsymbol{R}$ and, hence, we will consider two different noise covariance matrices $\boldsymbol{R}$ in order to measure the impact on two-step ABORT. In a first scenario, the noise vectors $\boldsymbol{n}_{t_p}$ and $\boldsymbol{n}_{T_p+t_s}$ include both thermal noise and clutter components, which are assumed to be uncorrelated so that $\boldsymbol{R} = \boldsymbol{R}_c + \sigma_n^2 \boldsymbol{I}_M$. The clutter covariance matrix is selected as $[\boldsymbol{R}_c]_{m_1,m_2} \propto \exp\{-2\pi^2\sigma_f^2(m_1 - m_2)^2\}$ with $\sigma_f^2 = 0.01$. The clutter to white noise ratio (CNR) is defined as $\mathrm{CNR} = \mathrm{Tr}\{\boldsymbol{R}_c\}/\mathrm{Tr}\{\sigma_n^2 \boldsymbol{I}_M\}$ and is set to $\mathrm{CNR} = 10$ dB or $\mathrm{CNR} = 30$ dB in the simulations. A second model for $\boldsymbol{R}$ will also be used, namely the standard exponentially-correlated matrix $[\boldsymbol{R}]_{m_1,m_2} = \rho^{|m_1 - m_2|}$ with different values of the one-lag correlation coefficient ρ. For the plain GLRT, the maximum



(a) $\boldsymbol{R}$ Gauss

(b) $\boldsymbol{R}$ Exp

Fig. 1. Probability of detection vs SNR for $T_p = 1, T_s = M + 1 = 9$.



(a) $\boldsymbol{R}$ Gauss

(b) $\boldsymbol{R}$ Exp

Fig. 2. Probability of detection vs SNR for $T_p = 1, T_s = 2M = 16$.

of the function $f(\eta)$ in (5) was searched over a grid with step 0.1 dB between -6 dB and 0 dB.

In all simulations, the probability of false alarm is set to $P_{fa} = 10^{-3}$. The thresholds of the detectors are obtained from $100/P_{fa}$ Monte-Carlo runs while 10^3 simulations are used to estimate the probability of detection P_d. We first consider the case of matched signals, then the case where the actual target signature differs from the assumed $\boldsymbol{v}$. To not burden too much the figures, only the curves that appreciably change with the varying parameters are shown.

A. Case of Matched Signals

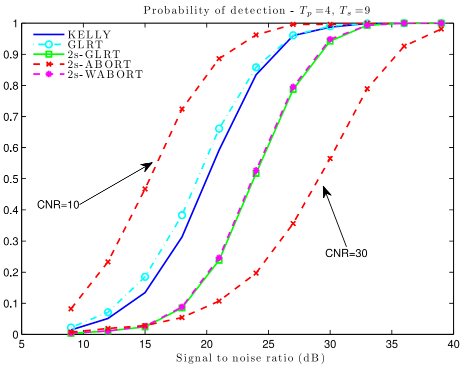
We first discuss the point-like target case $T_p = 1$. A different behavior can be observed based on both the number of training data T_s and CNR. Fig. 1 reports the P_d as function of SNR for $T_s = M + 1 = 9$ and different values of CNR: it is apparent that the two-step ABORT exhibits a peculiar robustness to ill-conditioning of the matrix $\boldsymbol{S}_y$ arising at small sample, which however weakens as CNR increases and practically vanishes at CNR $= 30$ dB. Analogous considerations hold true for the exponentially-correlated matrix.

This behavior is anyway absent when T_s is larger, and in that case one finds the typical loss of the two-step ABORT compared to plain GLRT, as reported in Fig. 2 for $T_s = 2M = 16$. Notice also that the loss of two-step GLRT and the two-step WABORT is limited compared to the corresponding one-step counterpart under matched conditions with sufficient number of training data.

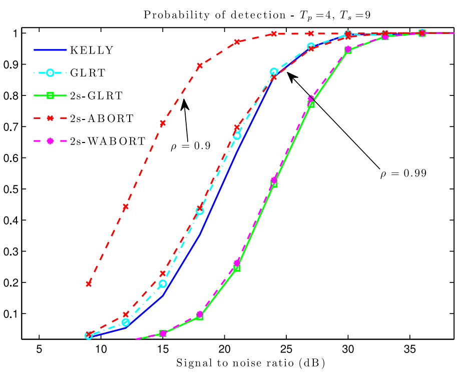
Finally, in the above results there is no significant difference between the stochastic GLRT detector and the corresponding deterministic one, namely Kelly's detector. An advantage is conversely observed for extended targets: Fig. 3 reports the case $T_p = 4$ for $T_s = M + 1 = 9$, which shows some gain with respect to Kelly's detector. Such a gain seems independent of the CNR; the two-step ABORT detector exhibits conversely a much larger gain that however turns into a loss for CNR $= 30$ dB. Analogous considerations hold true for the exponentially-correlated matrix. By increasing T_s to $2M = 16$, the superior performance of Kelly's detector show up, given that there is no mismatch compared to the assumed steering vector (Fig. 4); the stochastic version of the GLRT is comparable.

B. Mismatched Case

We evaluate now the performance of the detectors by simulating a mismatched scenario, where the actual steering vector has a Doppler frequency $f_s + \frac{\delta}{M}$ whereas the nominal one assumed by all detectors is f_s. We set $\delta = 0.25$, which for instance corresponds to a square cosine between the nominal steering vector and the true one equal to about 0.85 when $\boldsymbol{R}$ is exponentially-

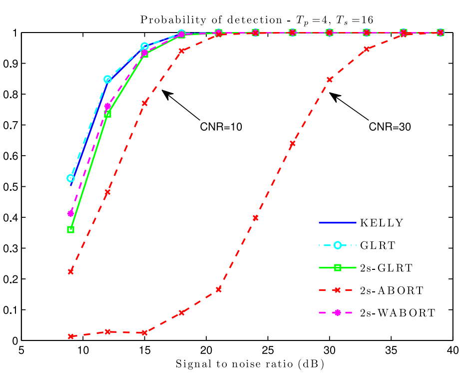


(a) $\boldsymbol{R}$ Gauss

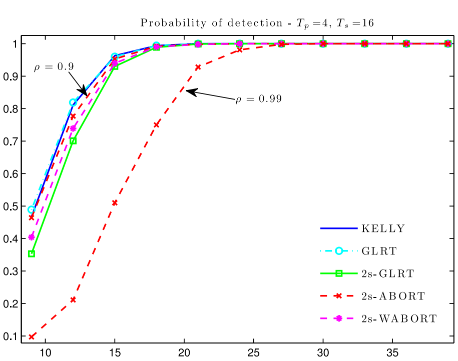


(b) $\boldsymbol{R}$ Exp

Fig. 3. Probability of detection vs SNR for $T_p = 4, T_s = M + 1 = 9$.

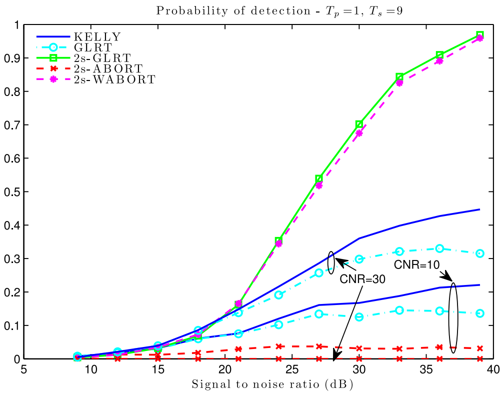


(a) $\boldsymbol{R}$ Gauss

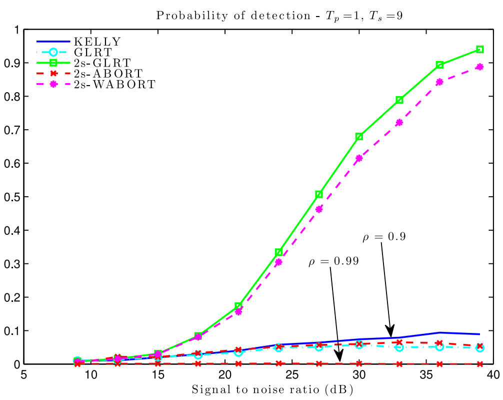


(b) $\boldsymbol{R}$ Exp

Fig. 4. Probability of detection vs SNR for $T_p = 4, T_s = 2M = 16$.



(a) $\boldsymbol{R}$ Gauss



(b) $\boldsymbol{R}$ Exp

Fig. 5. Probability of detection vs SNR for $T_p = 1$ in the mismatched case, $T_s = M + 1 = 9$.

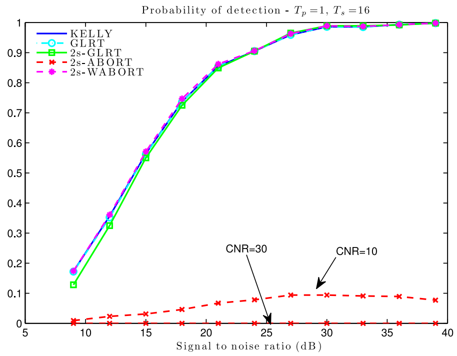
correlated with $\rho = 0.9$. We analyze the curves of P_d vs SNR in the same cases presented for the matched case.

Fig. 5 shows that, as consequence of the detection power loss in matched conditions for low T_s, all detectors have a low P_d except for the two-step GLRT and two-step WABORT. This is however a side-effect of the power detection loss in matched condition. In fact, as T_s increases, the GLRT lacks rejection capabilities (see Fig. 6). The two-step WABORT appears quite non-selective too, but this should be compared with analogous settings in the deterministic scenario, which in fact do not seem to provide better results (see curves in [12]).

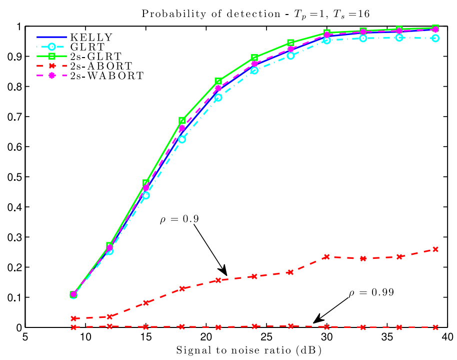
Finally, results for $T_p = 4$ are reported in Figs. 7-8 for $T_s = M + 1 = 9$ and $T_s = 2M = 16$, respectively. Interestingly, while in the former case there is no appreciable difference compared to $T_p = 1$ (see Fig. 5), for $T_s = 2M = 16$ Fig. 8 spotlights a dependency of the GLRT on the CNR which was not observed before: it seems that the GLRT eventually detects the mismatch at large SNR.

V. Conclusion

In this paper, the problem of detecting a target buried in Gaussian noise with unknown covariance matrix was addressed, using a stochastic model for the target's amplitude, namely a Gaussian assumption which complies with the Swerling I-II

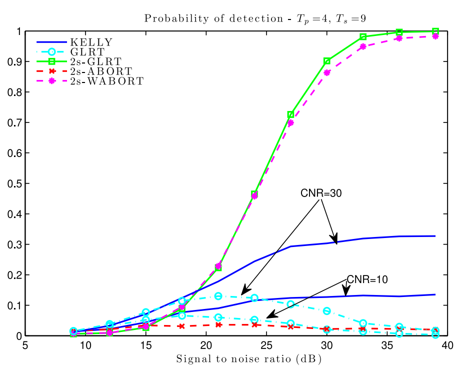


(a) $\boldsymbol{R}$ Gauss

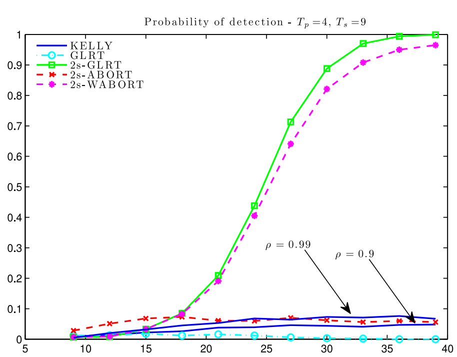


(b) $\boldsymbol{R}$ Exp

Fig. 6. Probability of detection vs SNR for $T_p = 1$ in the mismatched case, $T_s = 2M = 16$.

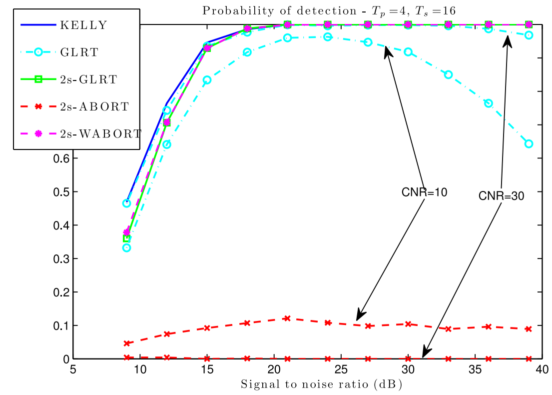


(a) $\boldsymbol{R}$ Gauss

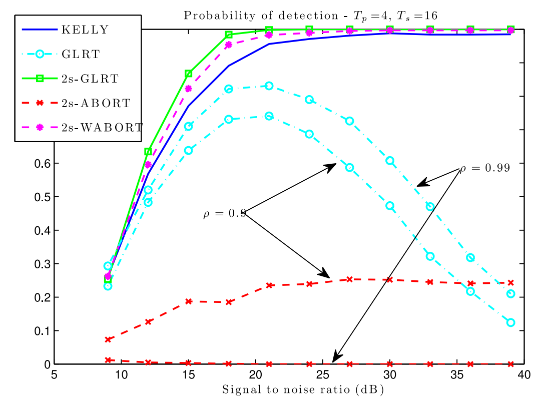


(b) $\boldsymbol{R}$ Exp

Fig. 7. Probability of detection vs SNR for $T_p = 4$ in the mismatched case, $T_s = M + 1 = 9$.



(a) $\boldsymbol{R}$ Gauss



(b) $\boldsymbol{R}$ Exp

Fig. 8. Probability of detection vs SNR for $T_p = 4$ in the mismatched case, $T_s = 2M = 16$.

target model. The exact generalized likelihood ratio test was derived and was shown to bear close resemblance to its deterministic counterpart, Kelly's GLRT. Two-step approaches were also presented where the noise covariance matrix is first assumed to be known. The GLRT was then derived for three different assumptions under the null hypothesis, either no signal or a fictitious signal orthogonal to the target signature. All but one detectors were proved to possess the CFAR property. Moreover, it was shown that the new detectors could significantly improve over Kelly's GLRT when the number of secondary (noise only) data is small. Additionally, some of the detectors achieve a good compromise between detection of matched signals and rejection of unwanted signals.

Appendix A
Derivation of Plain GLRT

This appendix is devoted to derivation of the plain GLRT for the composite hypothesis testing problem (3). Since both P and $\boldsymbol{R}$ are unknown, the GLR in this case writes

$$\frac{\max\limits_{P,\boldsymbol{R}} p_1\left(\boldsymbol{X},\boldsymbol{Y}\right)}{\max\limits_{\boldsymbol{R}} p_0\left(\boldsymbol{X},\boldsymbol{Y}\right)} \tag{19}$$

where $p_\ell(\boldsymbol{X},\boldsymbol{Y})$ is the probability density function (pdf) of the observations under hypothesis H_ℓ.

Under H_0 the pdf of the observations is given by

$$p_0(\boldsymbol{X},\boldsymbol{Y}) \propto |\boldsymbol{R}|^{-T_t} \operatorname{etr}\left\{-\boldsymbol{R}^{-1}\left(\boldsymbol{S}_y+\boldsymbol{X}\boldsymbol{X}^H\right)\right\} \tag{20}$$

where $\propto$ means proportional to. In this case, it is well known [14] that the maximum of $p_0(\boldsymbol{X},\boldsymbol{Y})$ is achieved for $\boldsymbol{R}=T_t^{-1}\left(\boldsymbol{S}_y+\boldsymbol{X}\boldsymbol{X}^H\right)$ and is thus given by

$$\max_{\boldsymbol{R}} p_0(\boldsymbol{X},\boldsymbol{Y}) \propto |\boldsymbol{S}_y+\boldsymbol{X}\boldsymbol{X}^H|^{-T_t}. \tag{21}$$

With no loss of generality, we assume that $\boldsymbol{v}$ is unit-norm and let $\boldsymbol{V}=\left[\boldsymbol{v}\ \boldsymbol{V}_\perp\right]$ be a unitary matrix, with $\boldsymbol{V}_\perp$ a basis for the subspace orthogonal to $\boldsymbol{v}$, i.e., $\boldsymbol{V}_\perp^H\boldsymbol{v}=\boldsymbol{0}$ and $\boldsymbol{V}_\perp^H\boldsymbol{V}_\perp=\boldsymbol{I}_{M-1}$. This transformation brings $\boldsymbol{v}$ to $\boldsymbol{V}^H\boldsymbol{v}=\boldsymbol{e}_1=\left[1\ 0\cdots 0\right]^T$. Let us define the transformed data $\tilde{\boldsymbol{X}}=\boldsymbol{V}^H\boldsymbol{X}=\begin{bmatrix}\tilde{\boldsymbol{X}}_1\\ \tilde{\boldsymbol{X}}_2\end{bmatrix}$ and $\tilde{\boldsymbol{Y}}=\boldsymbol{V}^H\boldsymbol{Y}=\begin{bmatrix}\tilde{\boldsymbol{Y}}_1\\ \tilde{\boldsymbol{Y}}_2\end{bmatrix}$, and transformed covariance matrix $\tilde{\boldsymbol{R}}=\boldsymbol{V}^H\boldsymbol{R}\boldsymbol{V}$. The joint pdf of $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be expressed as

$$\begin{aligned}p_1(\boldsymbol{X},\boldsymbol{Y}) \propto\ & |\tilde{\boldsymbol{R}}|^{-T_s}|\tilde{\boldsymbol{R}}+P\boldsymbol{e}_1\boldsymbol{e}_1^H|^{-T_p}\operatorname{etr}\left\{-\tilde{\boldsymbol{R}}^{-1}\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^H\right\}\\ &\times\operatorname{etr}\left\{-(\tilde{\boldsymbol{R}}+P\boldsymbol{e}_1\boldsymbol{e}_1^H)^{-1}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^H\right\}.\end{aligned} \tag{22}$$

Let us decompose $\tilde{\boldsymbol{R}}$ as

$$\tilde{\boldsymbol{R}}=\begin{pmatrix}\tilde{R}_{11} & \tilde{\boldsymbol{R}}_{12}\\ \tilde{\boldsymbol{R}}_{21} & \tilde{\boldsymbol{R}}_{22}\end{pmatrix} \tag{23}$$

and let $\tilde{R}_{1.2}=\tilde{R}_{11}-\tilde{\boldsymbol{R}}_{12}\tilde{\boldsymbol{R}}_{22}^{-1}\tilde{\boldsymbol{R}}_{21}$ and $\boldsymbol{\beta}=\tilde{\boldsymbol{R}}_{22}^{-1}\tilde{\boldsymbol{R}}_{21}$. Observe that $\tilde{\boldsymbol{R}}$ can be equivalently parametrized by $(\tilde{R}_{11},\ \tilde{\boldsymbol{R}}_{21},\ \tilde{\boldsymbol{R}}_{22})$ or $(\tilde{R}_{1.2},\ \boldsymbol{\beta},\ \tilde{\boldsymbol{R}}_{22})$. Using the facts that $|\tilde{\boldsymbol{R}}|=\tilde{R}_{1.2}|\tilde{\boldsymbol{R}}_{22}|$ and

$$\tilde{\boldsymbol{R}}^{-1}=\tilde{R}_{1.2}^{-1}\begin{pmatrix}1 & -\boldsymbol{\beta}^H\\ -\boldsymbol{\beta} & \boldsymbol{\beta}\boldsymbol{\beta}^H\end{pmatrix}+\begin{pmatrix}0 & \boldsymbol{0}\\ \boldsymbol{0} & \tilde{\boldsymbol{R}}_{22}^{-1}\end{pmatrix} \tag{24}$$

one can rewrite (22) as

$$\begin{aligned}p_1(\boldsymbol{X},\boldsymbol{Y}) \propto\ & |\tilde{\boldsymbol{R}}_{22}|^{-T_t}\tilde{R}_{1.2}^{-T_s}\left(P+\tilde{R}_{1.2}\right)^{-T_p}\\ &\times\operatorname{etr}\left\{-\tilde{\boldsymbol{R}}_{22}^{-1}\left(\tilde{\boldsymbol{Y}}_2\tilde{\boldsymbol{Y}}_2^H+\tilde{\boldsymbol{X}}_2\tilde{\boldsymbol{X}}_2^H\right)\right\}\\ &\times\exp\left\{-\left[1\ -\boldsymbol{\beta}^H\right]\tilde{\boldsymbol{A}}\begin{bmatrix}1\\ -\boldsymbol{\beta}\end{bmatrix}\right\}\end{aligned} \tag{25}$$

where we temporarily define

$$\tilde{\boldsymbol{A}}=\tilde{R}_{1.2}^{-1}\tilde{\boldsymbol{S}}_y+\left(P+\tilde{R}_{1.2}\right)^{-1}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^H \tag{26}$$

with $\tilde{\boldsymbol{S}}_y=\tilde{\boldsymbol{Y}}\tilde{\boldsymbol{Y}}^H$. Since

$$\begin{aligned}\left[1\ -\boldsymbol{\beta}^H\right]\tilde{\boldsymbol{A}}\begin{bmatrix}1\\ -\boldsymbol{\beta}\end{bmatrix}=&\left(\boldsymbol{\beta}-\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}\right)^H\tilde{\boldsymbol{A}}_{22}\\ &\times\left(\boldsymbol{\beta}-\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}\right)+\tilde{A}_{11}-\tilde{\boldsymbol{A}}_{12}\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}\end{aligned} \tag{27}$$

it follows that

$$\begin{aligned}p_1(\boldsymbol{X},\boldsymbol{Y}) \propto\ & |\tilde{\boldsymbol{R}}_{22}|^{-T_t}\operatorname{etr}\left\{-\tilde{\boldsymbol{R}}_{22}^{-1}\left(\tilde{\boldsymbol{Y}}_2\tilde{\boldsymbol{Y}}_2^H+\tilde{\boldsymbol{X}}_2\tilde{\boldsymbol{X}}_2^H\right)\right\}\\ &\times\exp\left\{-\left(\boldsymbol{\beta}-\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}\right)^H\tilde{\boldsymbol{A}}_{22}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}\right)\right\}\\ &\times\tilde{R}_{1.2}^{-T_s}\left(P+\tilde{R}_{1.2}\right)^{-T_p}\exp\left\{-\tilde{A}_{1.2}\right\}.\end{aligned} \tag{28}$$

The first term in the previous equation is recognized, up to a scalar factor, as the usual complex multivariate Gaussian likelihood function with covariance matrix $\tilde{\boldsymbol{R}}_{22}$ and is maximized at $\tilde{\boldsymbol{R}}_{22}=T_t^{-1}\left(\tilde{\boldsymbol{Y}}_2\tilde{\boldsymbol{Y}}_2^H+\tilde{\boldsymbol{X}}_2\tilde{\boldsymbol{X}}_2^H\right)$ [14]. The quadratic form in the second line is minimized for $\boldsymbol{\beta}=\tilde{\boldsymbol{A}}_{22}^{-1}\tilde{\boldsymbol{A}}_{21}$ and, hence, for any P and $\tilde{R}_{1.2}$

$$\begin{aligned}\max_{\tilde{\boldsymbol{R}}_{22},\boldsymbol{\beta}} p_1(\boldsymbol{X},\boldsymbol{Y}) \propto\ & |\tilde{\boldsymbol{Y}}_2\tilde{\boldsymbol{Y}}_2^H+\tilde{\boldsymbol{X}}_2\tilde{\boldsymbol{X}}_2^H|^{-T_t}\exp\left\{-\tilde{A}_{1.2}\right\}\\ &\times\tilde{R}_{1.2}^{-T_s}\left(P+\tilde{R}_{1.2}\right)^{-T_p}.\end{aligned} \tag{29}$$

Next, observe that $\tilde{A}_{1.2}^{-1}$ is the upper-left corner of $\tilde{\boldsymbol{A}}^{-1}$ and the latter is given by

$$\begin{aligned}\tilde{\boldsymbol{A}}^{-1}&=\tilde{R}_{1.2}\left[\tilde{\boldsymbol{S}}_y+(1+P\tilde{R}_{1.2}^{-1})^{-1}\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^H\right]^{-1}\\ &=\tilde{R}_{1.2}\boldsymbol{V}^H\left[\boldsymbol{S}_y+(1+P\tilde{R}_{1.2}^{-1})^{-1}\boldsymbol{X}\boldsymbol{X}^H\right]^{-1}\boldsymbol{V}.\end{aligned} \tag{30}$$

It ensues that

$$\tilde{A}_{1.2}^{-1}=\tilde{R}_{1.2}\boldsymbol{v}^H\left[\boldsymbol{S}_y+(1+P\tilde{R}_{1.2}^{-1})^{-1}\boldsymbol{X}\boldsymbol{X}^H\right]^{-1}\boldsymbol{v}. \tag{31}$$

For the sake of notational convenience, let us introduce $a=\tilde{R}_{1.2}=\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}>0$ and $b=P\tilde{R}_{1.2}^{-1}$. Observe that $b=P\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}$ is tantamount the signal-to-noise ratio at the output of the optimal filter $\boldsymbol{R}^{-1}\boldsymbol{v}$. Then, one can rewrite (29) as

$$\begin{aligned}\max_{\tilde{\boldsymbol{R}}_{22},\boldsymbol{\beta}} p_1(\boldsymbol{X},\boldsymbol{Y}) \propto\ & |\boldsymbol{V}_\perp^H\left(\boldsymbol{S}_y+\boldsymbol{X}\boldsymbol{X}^H\right)\boldsymbol{V}_\perp|^{-T_t}\\ &\times a^{-T_t}(1+b)^{-T_p}\\ &\times\exp\left\{-a^{-1}\left[\boldsymbol{v}^H\left(\boldsymbol{S}_y+(1+b)^{-1}\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}\right]^{-1}\right\}.\end{aligned} \tag{32}$$

Let $f(a)=a^{-T_t}\exp\left\{-\xi^{-1}a^{-1}\right\}$ where $\xi=\boldsymbol{v}^H(\boldsymbol{S}_y+(1+b)^{-1}\boldsymbol{X}\boldsymbol{X}^H)^{-1}\boldsymbol{v}>0$, and note that

$$\frac{\partial\log f(a)}{\partial a}=\frac{-aT_t+\xi^{-1}}{a^2}$$

is positive for $a\leq(T_t\xi)^{-1}$ and negative otherwise. Therefore, the maximum is achieved at $a=(T_t\xi)^{-1}>0$ and is given by

$$\max_{a>0} a^{-T_t}\exp\left\{-\xi^{-1}a^{-1}\right\}=\left(\frac{e}{T_t}\right)^{-T_t}\xi^{T_t}. \tag{33}$$

It follows that

$$\max_{\tilde{\boldsymbol{R}}_{22},\boldsymbol{\beta},a} p_1(\boldsymbol{X},\boldsymbol{Y}) \propto |\boldsymbol{V}_\perp^H(\boldsymbol{S}_y+\boldsymbol{X}\boldsymbol{X}^H)\boldsymbol{V}_\perp|^{-T_t} \times (1+b)^{-T_p}\left[\boldsymbol{v}^H\left(\boldsymbol{S}_y+(1+b)^{-1}\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}\right]^{T_t}. \quad (34)$$

Let us now observe that

$$\begin{aligned}
&|\boldsymbol{V}_\perp^H(\boldsymbol{S}_y+c\boldsymbol{X}\boldsymbol{X}^H)\boldsymbol{V}_\perp| \\
&=|\boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{V}_\perp||\boldsymbol{I}_{T_p}+c\boldsymbol{X}^H\boldsymbol{V}_\perp\left(\boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{V}_\perp\right)^{-1}\boldsymbol{V}_\perp^H\boldsymbol{X}| \\
&=|\boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{V}_\perp||\boldsymbol{I}_{T_p}+c\boldsymbol{X}^H\boldsymbol{S}_y^{-1/2}\boldsymbol{P}_{\boldsymbol{S}_y^{1/2}\boldsymbol{V}_\perp}\boldsymbol{S}_y^{-1/2}\boldsymbol{X}| \\
&=|\boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{V}_\perp||\boldsymbol{I}_{T_p}+c\boldsymbol{X}^H\boldsymbol{S}_y^{-1/2}\boldsymbol{P}^\perp_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}}\boldsymbol{S}_y^{-1/2}\boldsymbol{X}| \quad (35)
\end{aligned}$$

and

$$\begin{aligned}
&|\boldsymbol{I}_{T_p}+c\boldsymbol{X}^H\boldsymbol{S}_y^{-1/2}\boldsymbol{P}^\perp_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}}\boldsymbol{S}_y^{-1/2}\boldsymbol{X}| \\
&=|\boldsymbol{I}_{T_p}+c\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}-c\frac{\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}}{\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}}| \\
&=|\boldsymbol{I}_{T_p}+c\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}| \\
&\quad\times\left[1-c\frac{\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}\left(\boldsymbol{I}_{T_p}+c\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}}{\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}}\right] \\
&=|\boldsymbol{I}_{T_p}+c\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}| \\
&\quad\times\frac{\boldsymbol{v}^H\left[\boldsymbol{S}_y^{-1}-c\boldsymbol{S}_y^{-1}\boldsymbol{X}\left(\boldsymbol{I}_{T_p}+c\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\right]\boldsymbol{v}}{\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}} \\
&=|\boldsymbol{I}_{T_p}+c\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}|\frac{\boldsymbol{v}^H\left(\boldsymbol{S}_y+c\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}}{\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}}. \quad (36)
\end{aligned}$$

Furthermore, since $\boldsymbol{V}^H\boldsymbol{S}_y\boldsymbol{V}=\begin{bmatrix}\boldsymbol{v}^H\boldsymbol{S}_y\boldsymbol{v} & \boldsymbol{v}^H\boldsymbol{S}_y\boldsymbol{V}_\perp \\ \boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{v} & \boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{V}_\perp\end{bmatrix}$, one has

$$\begin{aligned}
|\boldsymbol{S}_y| &= |\boldsymbol{V}^H\boldsymbol{S}_y\boldsymbol{V}| \\
&= |\boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{V}_\perp| \\
&\quad\times\left[\boldsymbol{v}^H\boldsymbol{S}_y\boldsymbol{v}-\boldsymbol{v}^H\boldsymbol{S}_y\boldsymbol{V}_\perp\left(\boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{v}\right)^{-1}\boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{v}\right] \\
&= |\boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{V}_\perp|\left[\boldsymbol{v}^H\boldsymbol{S}_y^{1/2}\boldsymbol{P}^\perp_{\boldsymbol{S}_y^{1/2}\boldsymbol{V}_\perp}\boldsymbol{S}_y^{1/2}\boldsymbol{v}\right] \\
&= |\boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{V}_\perp|\left[\boldsymbol{v}^H\boldsymbol{S}_y^{1/2}\boldsymbol{P}_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}}\boldsymbol{S}_y^{1/2}\boldsymbol{v}\right] \\
&= |\boldsymbol{V}_\perp^H\boldsymbol{S}_y\boldsymbol{V}_\perp|\left(\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}\right)^{-1} \quad (37)
\end{aligned}$$

so that (34) can be indifferently written as

$$\begin{aligned}
&\max_{\tilde{\boldsymbol{R}}_{22},\boldsymbol{\beta},a} p_1(\boldsymbol{X},\boldsymbol{Y}) \\
&\propto |\boldsymbol{S}_y|^{-T_t}|\boldsymbol{I}_{T_p}+\boldsymbol{X}^H\boldsymbol{S}_y^{-1/2}\boldsymbol{P}^\perp_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}}\boldsymbol{S}_y^{-1/2}\boldsymbol{X}|^{-T_t}(1+b)^{-T_p} \\
&\quad\times\frac{|\boldsymbol{I}_{T_p}+(1+b)^{-1}\boldsymbol{X}^H\boldsymbol{S}_y^{-1/2}\boldsymbol{P}^\perp_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}}\boldsymbol{S}_y^{-1/2}\boldsymbol{X}|^{T_t}}{|\boldsymbol{I}_{T_p}+(1+b)^{-1}\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}|^{T_t}} \quad (38)
\end{aligned}$$

or

$$\max_{\tilde{\boldsymbol{R}}_{22},\boldsymbol{\beta},a} p_1(\boldsymbol{X},\boldsymbol{Y}) \propto |\boldsymbol{S}_y|^{-T_t}|\boldsymbol{I}_{T_p}+\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}|^{-T_t} \times(1+b)^{-T_p}\left[\frac{\boldsymbol{v}^H\left(\boldsymbol{S}_y+(1+b)^{-1}\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}}{\boldsymbol{v}^H\left(\boldsymbol{S}_y+\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}}\right]^{T_t}. \quad (39)$$

Finally, using the previous equations along with (21), the GLR for Gaussian signals is given by

$$\begin{aligned}
GLR_{\text{Gaussian}}^{1/T_t} &= \frac{|\boldsymbol{I}_{T_p}+\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}|}{|\boldsymbol{I}_{T_p}+\boldsymbol{X}^H\boldsymbol{S}_y^{-1/2}\boldsymbol{P}^\perp_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}}\boldsymbol{S}_y^{-1/2}\boldsymbol{X}|} \\
&\quad\times\max_b\frac{|\boldsymbol{I}_{T_p}+(1+b)^{-1}\boldsymbol{X}^H\boldsymbol{S}_y^{-1/2}\boldsymbol{P}^\perp_{\boldsymbol{S}_y^{-1/2}\boldsymbol{v}}\boldsymbol{S}_y^{-1/2}\boldsymbol{X}|}{(1+b)^{T_p/T_t}|\boldsymbol{I}_{T_p}+(1+b)^{-1}\boldsymbol{X}^H\boldsymbol{S}_y^{-1}\boldsymbol{X}|} \\
&= \frac{\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}}{\boldsymbol{v}^H\left(\boldsymbol{S}_y+\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}} \\
&\quad\times\max_b\frac{\boldsymbol{v}^H\left(\boldsymbol{S}_y+(1+b)^{-1}\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\boldsymbol{v}}{(1+b)^{T_p/T_t}\left(\boldsymbol{v}^H\boldsymbol{S}_y^{-1}\boldsymbol{v}\right)}. \quad (40)
\end{aligned}$$

APPENDIX B
DERIVATION OF TWO-STEP GLRT

In this appendix, we derive the GLRT for the detection problem in (6), for the three different hypotheses on $\boldsymbol{u}$, namely $\boldsymbol{u}=\boldsymbol{0}$, $\boldsymbol{u}\perp\boldsymbol{v}$ and $\boldsymbol{R}^{-1/2}\boldsymbol{u}\perp\boldsymbol{R}^{-1/2}\boldsymbol{v}$.

A. Noise Only under H_0, $\boldsymbol{u}=\boldsymbol{0}$

When $\boldsymbol{R}$ is known and $\boldsymbol{u}=\boldsymbol{0}$, only P is unknown and therefore only $p_1(\boldsymbol{X})$ needs to be maximized. One has

$$\begin{aligned}
p_1(\boldsymbol{X}) &= \pi^{-MT_p}|\boldsymbol{R}+P\boldsymbol{v}\boldsymbol{v}^H|^{-T_p} \\
&\quad\times\operatorname{etr}\left\{-\boldsymbol{X}^H\left(\boldsymbol{R}+P\boldsymbol{v}\boldsymbol{v}^H\right)^{-1}\boldsymbol{X}\right\} \\
&= \pi^{-MT_p}|\boldsymbol{R}|^{-T_p}\left(1+P\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}\right)^{-T_p} \\
&\quad\times\operatorname{etr}\left\{-\boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{X}^H\right\} \\
&\quad\times\exp\left\{\frac{P\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{S}_x\boldsymbol{R}^{-1}\boldsymbol{v}}{1+P\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}}\right\}. \quad (41)
\end{aligned}$$

where $\boldsymbol{S}_x=\boldsymbol{X}\boldsymbol{X}^H$. One needs to maximize (41) with respect to P. Let $u=\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v}$ and $v=\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{S}_x\boldsymbol{R}^{-1}\boldsymbol{v}$ and let us define the function

$$h(P)=(1+uP)^{-T_p}\exp\left\{\frac{vP}{1+uP}\right\}. \quad (42)$$

Differentiating the logarithm of $h(P)$ yields

$$\frac{\partial\log h(P)}{\partial P}=\frac{v-uT_p-u^2T_pP}{(1+uP)^2} \quad (43)$$

which is positive for $P\le(v-uT_p)/(u^2T_p)$ and negative otherwise. Now since P is necessarily positive, the maximum of $h(P)$ is achieved at $P=0$ if $v\le uT_p$ and at $(v-uT_p)/(u^2T_p)$

otherwise. Therefore, one has

$$\max_{P\geq 0} h(P) = \begin{cases} 1 & v/(uT_p) \leq 1 \\ g\left(\frac{v}{uT_p}\right) & \text{otherwise} \end{cases} \tag{44}$$

where

$$g(x) = x^{-T_p} \exp\left\{(x-1)T_p\right\}. \tag{45}$$

It then follows that

$$\max_{P\geq 0} p_1(\boldsymbol{X}) = \pi^{-MT_p} |\boldsymbol{R}|^{-T_p} \operatorname{etr}\left\{-\boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{X}^H\right\} \times g\left(\max\left[\frac{\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{S}_x\boldsymbol{R}^{-1}\boldsymbol{v}}{(\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v})T_p}, 1\right]\right) \tag{46}$$

and hence the GLR for known $\boldsymbol{R}$ is given by

$$GLR^{\boldsymbol{u}=\boldsymbol{0}}_{|\boldsymbol{R}}(\boldsymbol{X}) = g\left(\max\left[\frac{\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{S}_x\boldsymbol{R}^{-1}\boldsymbol{v}}{(\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v})T_p}, 1\right]\right). \tag{47}$$

B. $\boldsymbol{u}$ Orthogonal to $\boldsymbol{v}$

Let us assume now that $\boldsymbol{u} \perp \boldsymbol{v}$. There is no difference from the previous case for $p_1(\boldsymbol{X})$ but now $p_0(\boldsymbol{X})$ should be maximized with respect to $\boldsymbol{u}$ under the constraint that $\boldsymbol{u} \perp \boldsymbol{v}$. It means that $\boldsymbol{u} = \boldsymbol{V}_\perp\boldsymbol{\eta}$ for some vector $\boldsymbol{\eta} \in \mathbb{C}^{M-1}$. The pdf of $\boldsymbol{X}$ under H_0 is now

$$\begin{aligned} p_0(\boldsymbol{X}) &= \pi^{-MT_p} |\boldsymbol{R} + \boldsymbol{V}_\perp\boldsymbol{\eta}\boldsymbol{\eta}^H\boldsymbol{V}_\perp^H|^{-T_p} \\ &\quad\times \operatorname{etr}\left\{-\boldsymbol{X}^H\left(\boldsymbol{R} + \boldsymbol{V}_\perp\boldsymbol{\eta}\boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\right)^{-1}\boldsymbol{X}\right\} \\ &= \pi^{-MT_p}|\boldsymbol{R}|^{-T_p}\left(1 + \boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\boldsymbol{R}^{-1}\boldsymbol{V}_\perp\boldsymbol{\eta}\right)^{-T_p} \\ &\quad\times \operatorname{etr}\left\{-\boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{X}^H\right\} \\ &\quad\times \exp\left\{\frac{\boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\boldsymbol{R}^{-1}\boldsymbol{S}_x\boldsymbol{R}^{-1}\boldsymbol{V}_\perp\boldsymbol{\eta}}{1 + \boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\boldsymbol{R}^{-1}\boldsymbol{V}_\perp\boldsymbol{\eta}}\right\}. \end{aligned} \tag{48}$$

Let

$$f(\boldsymbol{\eta}) = \left(1 + \boldsymbol{\eta}^H\boldsymbol{G}\boldsymbol{\eta}\right)^{-T_p}\exp\left\{\frac{\boldsymbol{\eta}^H\boldsymbol{F}\boldsymbol{\eta}}{1 + \boldsymbol{\eta}^H\boldsymbol{G}\boldsymbol{\eta}}\right\} \tag{49}$$

where $\boldsymbol{G} > \boldsymbol{0}$. Differentiating the logarithm of $f(.)$ yields

$$\begin{aligned} \frac{\partial \log f(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} = &-T_p\frac{\boldsymbol{G}\boldsymbol{\eta}}{1 + \boldsymbol{\eta}^H\boldsymbol{G}\boldsymbol{\eta}} \\ &+ \frac{\boldsymbol{F}\boldsymbol{\eta}}{1 + \boldsymbol{\eta}^H\boldsymbol{G}\boldsymbol{\eta}} - \frac{(\boldsymbol{\eta}^H\boldsymbol{F}\boldsymbol{\eta})\boldsymbol{G}\boldsymbol{\eta}}{(1 + \boldsymbol{\eta}^H\boldsymbol{G}\boldsymbol{\eta})^2}. \end{aligned} \tag{50}$$

Equating the result to $\boldsymbol{0}$ results in $\boldsymbol{G}\boldsymbol{\eta} \propto \boldsymbol{F}\boldsymbol{\eta}$, which means that $\boldsymbol{\eta}$ is a generalized eigenvector of the matrix pencil $(\boldsymbol{F}, \boldsymbol{G})$. Let $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_Q, \cdots, \boldsymbol{e}_{M-1}$ be the generalized eigenvectors of $(\boldsymbol{F}, \boldsymbol{G})$ associated with generalized eigenvalues $\lambda_1 \geq \cdots \geq \lambda_Q > 0, \cdots, 0$ where $Q = \min(T_p, M-1)$. Note that one can choose $\boldsymbol{e}_q$ such that $\boldsymbol{e}_q^H\boldsymbol{G}\boldsymbol{e}_q = 1$. Then, $\boldsymbol{\eta} = \beta\boldsymbol{e}_q$ for some q. It follows that

$$f(\beta\boldsymbol{e}_q) = \left(1 + |\beta|^2\right)^{-T_p}\exp\left\{\frac{\lambda_q|\beta|^2}{1 + |\beta|^2}\right\}. \tag{51}$$

Let us thus study the function $h(x, \lambda_q) = (1+x)^{-T_p}\exp\left\{\frac{\lambda_q x}{1+x}\right\}$ for $x > 0$. One has

$$\frac{\partial \log h(x, \lambda_q)}{\partial x} = \frac{\lambda_q - T_p - xT_p}{(1+x)^2}. \tag{52}$$

This derivative is positive for $x \leq (\lambda_q - T_p)/T_p$, negative otherwise. It follows that $h(x, \lambda_q)$ achieves its maximum at $(\lambda_q - T_p)/T_p$ if the latter is positive, at 0 otherwise. Some simple calculations enable one to show that

$$\max_{x\geq 0} h(x, \lambda_q) = \begin{cases} 1 & \lambda_q/T_p \leq 1 \\ g\left(\frac{\lambda_q}{T_p}\right) & \text{otherwise} \end{cases} \tag{53}$$

where $g(.)$ is defined in (45). Now, it is easily verified that $g(x)$ is monotonically increasing for $x > 1$, which means that $f(\boldsymbol{\eta})$ will be maximized when $\boldsymbol{\eta}$ is proportional to $\boldsymbol{e}_1$. Gathering the previous findings, we find out that

$$\max_{\boldsymbol{\eta}} p_0(\boldsymbol{X}) = \pi^{-MT_p}|\boldsymbol{R}|^{-T_p}\operatorname{etr}\left\{-\boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{X}^H\right\} \times g\left(\max\left[\frac{\lambda_{\max}\left(\boldsymbol{V}_\perp^H\boldsymbol{R}^{-1}\boldsymbol{S}_x\boldsymbol{R}^{-1}\boldsymbol{V}_\perp, \boldsymbol{V}_\perp^H\boldsymbol{R}^{-1}\boldsymbol{V}_\perp\right)}{T_p}, 1\right]\right). \tag{54}$$

The GLR for known $\boldsymbol{R}$ and $\boldsymbol{u} \perp \boldsymbol{v}$ is thus given by

$$GLR^{\boldsymbol{u}\perp\boldsymbol{v}}_{|\boldsymbol{R}}(\boldsymbol{X}) = \frac{g\left(\max\left[\frac{\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{S}_x\boldsymbol{R}^{-1}\boldsymbol{v}}{(\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v})T_p}, 1\right]\right)}{g\left(\max\left[\frac{\lambda_{\max}\left(\boldsymbol{V}_\perp^H\boldsymbol{R}^{-1}\boldsymbol{S}_x\boldsymbol{R}^{-1}\boldsymbol{V}_\perp, \boldsymbol{V}_\perp^H\boldsymbol{R}^{-1}\boldsymbol{V}_\perp\right)}{T_p}, 1\right]\right)}. \tag{55}$$

C. $\boldsymbol{R}^{-1/2}\boldsymbol{u}$ Orthogonal to $\boldsymbol{R}^{-1/2}\boldsymbol{v}$

We finally consider orthogonality in the whitened space. We now have $\boldsymbol{R}^{-1/2}\boldsymbol{u} = \boldsymbol{R}^{1/2}\boldsymbol{V}_\perp\boldsymbol{\eta}$ for some vector $\boldsymbol{\eta} \in \mathbb{C}^{M-1}$. The covariance matrix under H_0 can be written as

$$\begin{aligned} &\boldsymbol{R} + \boldsymbol{R}\boldsymbol{V}_\perp\boldsymbol{\eta}\boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\boldsymbol{R} \\ &= \boldsymbol{R}^{1/2}\left[\boldsymbol{I}_M + \boldsymbol{R}^{1/2}\boldsymbol{V}_\perp\boldsymbol{\eta}\boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\boldsymbol{R}^{1/2}\right]\boldsymbol{R}^{1/2} \end{aligned} \tag{56}$$

whose inverse is

$$\left(\boldsymbol{R} + \boldsymbol{R}\boldsymbol{V}_\perp\boldsymbol{\eta}\boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\boldsymbol{R}\right)^{-1} = \boldsymbol{R}^{-1} - \frac{\boldsymbol{V}_\perp\boldsymbol{\eta}\boldsymbol{\eta}^H\boldsymbol{V}_\perp^H}{1 + \boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\boldsymbol{R}\boldsymbol{V}_\perp\boldsymbol{\eta}}. \tag{57}$$

The pdf of $\boldsymbol{X}$ under H_0 thus becomes

$$\begin{aligned} p_0(\boldsymbol{X}) &= \pi^{-MT_p}|\boldsymbol{R}|^{-T_p}\left(1 + \boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\boldsymbol{R}\boldsymbol{V}_\perp\boldsymbol{\eta}\right)^{-T_p} \\ &\quad\times \operatorname{etr}\left\{-\boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{X}^H\right\} \\ &\quad\times \exp\left\{\frac{\boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\boldsymbol{S}_x\boldsymbol{V}_\perp\boldsymbol{\eta}}{1 + \boldsymbol{\eta}^H\boldsymbol{V}_\perp^H\boldsymbol{R}\boldsymbol{V}_\perp\boldsymbol{\eta}}\right\}. \end{aligned} \tag{58}$$

Obviously (58) bears much resemblance with (48) and one can avail the previous derivations to simply show that

$$\max_{\boldsymbol{\eta}} p_0(\boldsymbol{X}) = \pi^{-MT_p} |\boldsymbol{R}|^{-T_p} \operatorname{etr}\left\{-\boldsymbol{R}^{-1}\boldsymbol{X}\boldsymbol{X}^H\right\} \times g\left(\max\left[\frac{\lambda_{\max}\left(\boldsymbol{V}_\perp^H \boldsymbol{S}_x \boldsymbol{V}_\perp, \boldsymbol{V}_\perp^H \boldsymbol{R}\boldsymbol{V}_\perp\right)}{T_p}, 1\right]\right). \quad (59)$$

The GLR for known $\boldsymbol{R}$ and $\boldsymbol{R}^{-1/2}\boldsymbol{u} \perp \boldsymbol{R}^{-1/2}\boldsymbol{v}$ is thus given by

$$GLR_{|\boldsymbol{R}}^{\boldsymbol{R}^{-1/2}\boldsymbol{u}\perp\boldsymbol{R}^{-1/2}\boldsymbol{v}}(\boldsymbol{X}) = \frac{g\left(\max\left[\frac{\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{S}_x\boldsymbol{R}^{-1}\boldsymbol{v}}{(\boldsymbol{v}^H\boldsymbol{R}^{-1}\boldsymbol{v})T_p}, 1\right]\right)}{g\left(\max\left[\frac{\lambda_{\max}\left(\boldsymbol{V}_\perp^H\boldsymbol{S}_x\boldsymbol{V}_\perp, \boldsymbol{V}_\perp^H\boldsymbol{R}\boldsymbol{V}_\perp\right)}{T_p}, 1\right]\right)}. \quad (60)$$

References

[1] E. Conte, A. De Maio, and G. Ricci, "GLRT-based adaptive detection algorithm for range-spread targets," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1336–1348, Jul. 2001.

[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 I and Part II," Massachusetts Inst. Technol., Lincoln Lab., Lexington, MA, USA, Tech. Rep. 724, Jun. 1985.

[4] E. J. Kelly, "Adaptive detection in non-stationary interference, Part III," , Massachusetts Inst. Technol., Lincoln Lab., Lexington, MA, USA, *Tech. Rep. 761*, 24 Aug. 1987.

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

[6] E. J. Kelly, "Performance of an adaptive detection algorithm; rejection of unwanted signals," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 25, no. 2, pp. 122–133, Apr. 1989.

[7] 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.

[8] R. S. Raghavan, H. F. Qiu, and D. J. McLaughlin, "CFAR detection in clutter with unknown correlation properties," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 2, pp. 647–657, Apr. 1995.

[9] P. Swerling, "Probability of detection for fluctuating targets," *IRE Trans. Inf. Theory*, vol. 6, no. 2, pp. 269–308, Apr. 1960.

[10] P. Swerling, "Radar probability of detection for some additional fluctuating target cases," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 2, pp. 698–709, Apr. 1997.

[11] N. B. Pulsone and C. M. Rader, "Adaptive beamformer orthogonal rejection test," *IEEE Trans. Signal Process.*, vol. 49, no. 3, pp. 521–529, Mar. 2001.

[12] F. Bandiera, O. Besson, and G. Ricci, "An ABORT-like detector with improved mismatched signals rejection capabilities," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 14–25, Jan. 2008.

[13] F. Bandiera, D. Orlando, and G. Ricci, *Advanced Radar Detection Schemes Under Mismatched Signal Models*, ser. Synthesis lectures on signal processing, J. Moura, Ed. Williston, VT, USA: Morgan & Claypool, 2009.

[14] 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.

Olivier Besson received the Ph.D. degree in signal processing from Institut National Polytechnique, Toulouse, France, in 1992. He is currently working with Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-Supaéro), Toulouse, France. His research interests include statistical array processing, adaptive detection and estimation, with application to radar.

Angelo Coluccia received the Engineering degree in telecommunication engineering (summa cum laude) in 2007 and the Ph.D. degree in information engineering in 2011, both from the University of Salento. Lecce, Italy. He is currently an Assistant Professor in the Dipartimento di Ingegneria dell'Innovazione, University of Salento, where he teaches the course of Telecommunication Systems. Since 2008, he has been engaged in research projects on traffic analysis, security, and anomaly detection in operational cellular networks. He has been a Research Fellow at Forschungszentrum Telekommunikation Wien, Vienna, Austria, and Visiting Scholar at Institut Supérieur de l'Aéronautique et de l'Espace ISAE-Supaéro, Toulouse, France. His research interests include the area of multichannel and multiagent signal processing, including cooperative sensing and estimation in wireless networks, detection, and localization. Relevant application fields are radar, ad-hoc (sensor, overlay, social) networks, and intelligent transportation systems.

Eric Chaumette received the Engineer degree in electronics from ENAC, and the M.Sc. degree in signal processing from ENSEEIHT, University of Toulouse, France, both in 1989. He received the Ph.D. degree from CNRS, ENS de Cachan, France, in 2004. From 1990 to 2013, he worked as a Research Engineer involved in radar detection and estimation performance, first with Thales (1990–2007) and second with the French Aerospace Lab (ONERA, 2008–2013). From 2000 to 2014, he was also an Associate Researcher in laboratory SATIE, CNRS, ENS de Cachan, France, where he received the accreditation to supervise research (HDR) in 2014. He is currently a Professor in the Electronics, Optronics and Signal Department (DEOS), ISAE-Supaéro, Toulouse, France. His main domains of interest are related to detection and estimation theory applied to radar, GNSS, and navigation.

Giuseppe Ricci was born in Naples, Italy, on February 15, 1964. He received the Doctorate degree and the Ph.D. degree, both in electronic engineering, from the University of Naples "Federico II," Naples, Italy, in 1990 and 1994, respectively. Since 1995, he has been with the University of Salento (formerly University of Lecce), first as an Assistant Professor of telecommunications and, since 2002, as a Professor. His research interests include field of statistical signal processing with emphasis on radar processing, localization algorithms, and CDMA systems. More precisely, he has focused on high-resolution radar clutter modeling, detection of radar signals in Gaussian and nonGaussian disturbance, oil spill detection from SAR data, track-before-detect algorithms fed by space-time radar data, localization in wireless sensor networks, multiuser detection in overlay CDMA systems, and blind multiuser detection. He has held visiting positions at the University of Colorado at Boulder, CO, USA, in 1997–1998 and in April/May 2001, at the Colorado State University, CO, USA, in July/September 2003, March 2005, September 2009, and March 2011, at Ensica, Toulouse, France, in March 2006, and at the University of Connecticut, Storrs CT, USA, in September 2008.

François Vincent received the Engineer degree in electronics and signal processing from ENSEEIHT, University of Toulouse, France, the PhD. Degree in signal processing from the University of Toulouse and the "habilitation à diriger des recherches" from INPT, University of Toulouse, in 1995, 1999 and 2009, respectively. From 1999 to 2001, he was a Research Engineer with Siemens, Toulouse, France. He joined ISAE-Supaéro, Toulouse, France, in 2001 and he is currently a Professor. His research interests include radar and navigation signal processing.