# Adaptive Detection Using Whitened Data When Some of the Training Samples Undergo Covariance Mismatch

Olivier Besson

*Abstract*—**We consider adaptive detection of a signal of interest when two sets of training samples are available, one sharing the same covariance matrix as the data under test, the other set being mismatched. The approach proposed in this letter is to whiten both the data under test and the matched training samples using the sample covariance matrix of the mismatched training samples. The distribution of the whitened data is then derived and subsequently the generalized likelihood ratio test is obtained. Numerical simulations show that it performs well and is rather robust.**

*Index Terms*—**Adaptive detection, covariance mismatch, generalized likelihood ratio test, Student distribution.**

## I. Introduction and Problem Statement

ONE of the key tasks of a radar system is to detect a target, with known space and/or time signature, among disturbance (clutter, thermal noise, interference) whose parameters are generally unknown [1], [2]. A fundamental contribution to addressing and solving this problem was made by Kelly [3]–[6] who formulated it as a composite hypothesis problem where data is split into primary data (data under test where the presence of a target is sought) and secondary data -often referred to as training samples- which contains disturbance only and enable one to learn the disturbance in the data under test. Kelly derived and thoroughly analyzed the generalized likelihood ratio test (GLRT) related to various composite hypotheses testing problems, under the Gaussian assumption for the data, a known target signature and the same disturbance statistics between primary and secondary data. Kelly's GLRT and the adaptive matched filter (AMF) [7] constitute the ubiquitous references for this canonical framework.

In practice however there might exist some deviation from this canonical model. Discrepancies can come from mismatches or uncertainties about the target signatures, viz. the actual target signature differs from the assumed target signature or the target signature is not perfectly known. It can also be due to a covariance mismatch between the primary data and the training samples, which is the case of interest in the present letter. When one is confronted with mismatches, the first approach that comes to mind is to assess the robustness of canonical detectors in order to evaluate the performance loss due to operating under assumptions that are not met. For instance, references [8]–[10] analyze the GLRT, the AMF and the adaptive coherence estimator (ACE) under some technical assumptions regarding the difference between the two covariance matrices, e.g., that the generalized eigenrelation is satisfied [10]. Recently [11] analyzed the performance of the AMF when the two covariance matrices have the same eigenvectors but different eigenvalues.

When performance loss is deemed too damaging, then one usually takes into account the mismatch at the design stage of the detector, i.e., in the underlying assumptions, so that the detection scheme is from the start planned to accommodate potential discrepancies. As for covariance mismatches, a widely studied case is when the two covariance matrices are proportional to each other which leads to the adaptive coherence estimator (ACE) also referred to as a normalized AMF [12], [13]. Mismatch can also be due to additional components in the training samples. Thereby, Gerlach considered scenarios where some training samples are contaminated by signal-like components [14] or outliers [15]. An arbitrary rank-one modification is considered in [16] whereas the generalized eigenrelation is assumed to hold in [17]. In [18], two training data sets are assumed to be available, one containing thermal noise and jamming, the other one containing clutter in addition.

A very interesting approach was recently proposed in [19]. In this paper, three sets of data are available. The test data $\mathbf{X}_a$ consists of a single vector where presence of the target is sought. Additionally, two sets of target-free samples are available: one set of vectors $\mathbf{X}_b$ (called the reference vectors) shares the same covariance matrix as $\mathbf{X}_a$ while another set of vectors $\mathbf{X}_c$ has a different covariance matrix. $\mathbf{X}_b$ can be viewed as the "good" training samples while $\mathbf{X}_c$ contains mismatched training samples. The originality in [19] lies in considering $(\mathbf{X}_a, \mathbf{X}_b)$ as the primary data and $\mathbf{X}_c$ as the set of (mismatched) training samples. More precisely, it is assumed that $\mathbf{X}_a \overset{d}{=} \mathbb{C}\mathcal{N}(\mu\mathbf{v}\alpha^H, \mathbf{R}, \mathbf{I}_{T_a})$[1] and $\mathbf{X}_b \overset{d}{=} \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{R}, \mathbf{I}_{T_b})$ where $\mathbf{v}$ stands for the target signature and $\alpha$ denotes its amplitude. On the other hand $\mathbf{X}_c \overset{d}{=} \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{C}, \mathbf{I}_{T_c})$ with $\mathbf{C} \neq \mathbf{R}$. In a first step $\mathbf{C}$ is assumed to be known and cell averaging (CA) is used on the whitened data $(\mathbf{C}^{-1/2}\mathbf{X}_a, \mathbf{C}^{-1/2}\mathbf{X}_b)$. Then the sample covariance matrix (SCM) of $\mathbf{X}_c$ is substituted for its theoretical value. Actually cell

The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Guang Hua.

The author is with the University of Toulouse, ISAE-SUPAERO, 31055 Toulouse, France (e-mail: olivier.besson@isae-supaero.fr).

Digital Object Identifier 10.1109/LSP.2020.2991893

[1] $\mathbb{C}\mathcal{N}(\bar{\mathbf{X}}, \mathbf{\Sigma}, \mathbf{\Psi})$ stands for the complex matrix-variate Gaussian distribution whose density is $p(\mathbf{X}) = \pi^{-MT}|\mathbf{\Sigma}|^{-T}|\mathbf{\Psi}|^{-M}\operatorname{etr}\{-(\mathbf{X}-\bar{\mathbf{X}})^H\mathbf{\Sigma}^{-1}(\mathbf{X}-\bar{\mathbf{X}})\mathbf{\Psi}^{-1}\}$.

averaging corresponds to the GLRT for deciding whether $\mu = 0$ or $\mu \neq 0$ in the following composite hypothesis testing problem:

$$\begin{aligned}\mathbf{v}^H\mathbf{S}_c^{-1}\mathbf{X}_a|\mathbf{S}_c &\stackrel{d}{=} \mathbb{C}\mathcal{N}\left(\mu(\mathbf{v}^H\mathbf{S}_c^{-1}\mathbf{v})\alpha^H, \gamma\mathbf{I}_{T_a}\right)\\ \mathbf{v}^H\mathbf{S}_c^{-1}\mathbf{X}_b|\mathbf{S}_c &\stackrel{d}{=} \mathbb{C}\mathcal{N}\left(\mathbf{0}, \gamma\mathbf{I}_{T_b}\right) \end{aligned} \tag{1}$$

where $\gamma = \mathbf{v}^H\mathbf{S}_c^{-1}\mathbf{R}\mathbf{S}_c^{-1}\mathbf{v}$. Both $\gamma$ and $\alpha$ are unknowns and the cell-averaging test statistic for this problem can be written as $t_{CA} = \|\mathbf{v}^H\mathbf{S}_c^{-1}\mathbf{X}_a\|^2/\|\mathbf{v}^H\mathbf{S}_c^{-1}\mathbf{X}_b\|^2$. Note that cell-averaging uses the adaptively whitened data $\mathbf{Y}_a = \mathbf{L}_c^{-1}\mathbf{X}_a$ and $\mathbf{Y}_b = \mathbf{L}_c^{-1}\mathbf{X}_b$ with $\mathbf{L}_c$ a square-root of $\mathbf{S}_c$. However, the hypotheses testing problem in (1) is based on a conditional distribution, since the marginal distributions of $\mathbf{v}^H\mathbf{S}_c^{-1}\mathbf{X}_a$ and $\mathbf{v}^H\mathbf{S}_c^{-1}\mathbf{X}_b$ are obviously not Gaussian. In this letter, we start with the same model but we derive the exact (unconditional) distributions of $\mathbf{Y}_a = \mathbf{L}_c^{-1}\mathbf{X}_a$ and $\mathbf{Y}_b = \mathbf{L}_c^{-1}\mathbf{X}_b$, under some assumptions about the relation between $\mathbf{R}$ and $\mathbf{C}$. Then, based on this distribution, $\mathrm{GLR}(\mathbf{Y}_a, \mathbf{Y}_b)$ is obtained. This approach proceeds along the same philosophy as in [20] where we showed that the AMF is a conditional GLRT and where we derived the GLRT based on the marginal distribution of the adaptively whitened data. It was shown that this approach yields equivalent or slightly better results than the plain GLRT. The present letter thus extends the results of [20] to a scenario with three sets of samples, one of which undergoes a covariance mismatch.

## II. Detection Using Whitening of Test and Reference Data

As discussed previously, we assume that we wish to decide about the presence/absence of a target in the primary data $\tilde{\mathbf{X}}_a$, assuming that two sets of training samples are available, one $\tilde{\mathbf{X}}_b$ which shares the same covariance matrix as the noise covariance matrix in $\tilde{\mathbf{X}}_a$, another set $\tilde{\mathbf{X}}_c$ whose covariance matrix is different. Herein, we consider Swerling I-II type targets [21], [22] for which the target amplitude is assumed to be random and to follow a Gaussian distribution. Consequently the available data can be statistically described as $\tilde{\mathbf{X}}_a \stackrel{d}{=} \mathbb{C}\mathcal{N}(\mathbf{0}, \tilde{\mathbf{R}} + P\tilde{\mathbf{v}}\tilde{\mathbf{v}}^H, \mathbf{I}_{T_a})$, $\tilde{\mathbf{X}}_b \stackrel{d}{=} \mathbb{C}\mathcal{N}(\mathbf{0}, \tilde{\mathbf{R}}, \mathbf{I}_{T_b})$ and $\tilde{\mathbf{X}}_c \stackrel{d}{=} \mathbb{C}\mathcal{N}(\mathbf{0}, \tilde{\mathbf{C}}, \mathbf{I}_{T_c})$ where $\tilde{\mathbf{v}}$ stands for the target signature and $P$ denotes its power. $\tilde{\mathbf{R}}$ is the disturbance covariance matrix in $\tilde{\mathbf{X}}_a$ and $\tilde{\mathbf{X}}_b$, $\tilde{\mathbf{C}}$ the disturbance covariance matrix in $\tilde{\mathbf{X}}_c$. $T_a$, $T_b$ and $T_c$ are the number of observations in $\tilde{\mathbf{X}}_a$, $\tilde{\mathbf{X}}_b$ and $\tilde{\mathbf{X}}_c$, and we let $T_p = T_a + T_b$. Our problem is to decide between $H_0 : P = 0$ and $H_1 : P > 0$. In the sequel we assume that $\|\tilde{\mathbf{v}}\| = 1$ and we define the unitary matrix $\mathbf{Q} = \begin{bmatrix}\tilde{\mathbf{V}}_\perp & \tilde{\mathbf{v}}\end{bmatrix}$ along with the transformed data $\mathbf{X}_a = \mathbf{Q}^H\tilde{\mathbf{X}}_a$, $\mathbf{X}_b = \mathbf{Q}^H\tilde{\mathbf{X}}_b$, $\mathbf{X}_c = \mathbf{Q}^H\tilde{\mathbf{X}}_c$ and covariance matrices $\mathbf{R} = \mathbf{Q}^H\tilde{\mathbf{R}}\mathbf{Q}$ and $\mathbf{C} = \mathbf{Q}^H\tilde{\mathbf{C}}\mathbf{Q}$. Then the problem is equivalent to that of deciding whether $P = 0$ or $P > 0$ from

$$\begin{aligned}\mathbf{X}_a &\stackrel{d}{=} \mathbb{C}\mathcal{N}\left(\mathbf{0}, \mathbf{R} + P\mathbf{e}_M\mathbf{e}_M^H, \mathbf{I}_{T_a}\right)\\ \mathbf{X}_b &\stackrel{d}{=} \mathbb{C}\mathcal{N}\left(\mathbf{0}, \mathbf{R}, \mathbf{I}_{T_b}\right)\\ \mathbf{X}_c &\stackrel{d}{=} \mathbb{C}\mathcal{N}\left(\mathbf{0}, \mathbf{C}, \mathbf{I}_{T_c}\right)\end{aligned} \tag{2}$$

with $\mathbf{e}_M = \mathbf{Q}^H\tilde{\mathbf{v}} = [0 \ \ldots \ 0 \ 1]^T$. As said before, our approach, similarly to [19] is to whiten $\mathbf{X}_a$ and $\mathbf{X}_b$ using $\mathbf{X}_c$,

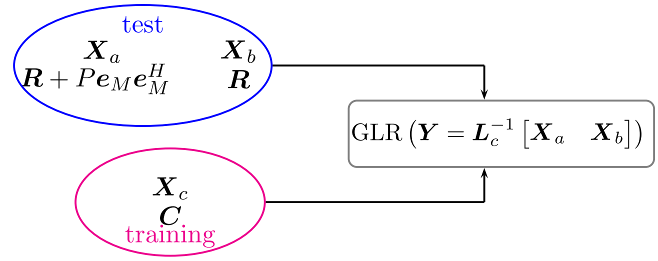


Fig. 1. Illustration of the approach. $\mathbf{X}_a$ and $\mathbf{X}_b$ are adaptively whitened by $\mathbf{L}_c^{-1}$ ($\mathbf{L}_c$ the Cholesky factor of $\mathbf{X}_c\mathbf{X}_c^H$) to produce $\mathbf{Y}$ and the GLRT based on $\mathbf{Y}$ is derived under the assumption that $\mathbf{R} = \mathbf{G}\mathbf{G}^H$ and $\mathbf{C} = \mathbf{G}\mathbf{D}^{-1}\mathbf{G}^H$.

and then to derive the GLRT from the whitened data. This is illustrated in Figure 1.

In order to derive the GLRT based on whitening of $\mathbf{X}_a$ and $\mathbf{X}_b$, we need to assume some relation between $\mathbf{R}$ and $\mathbf{C}$. Otherwise, if $\mathbf{R}$ and $\mathbf{C}$ are unknown and arbitrary, then the problem is not even identifiable unless $T_b \geq M$ and $T_c \geq M$ and, anyway, $\mathbf{X}_c$ would be useless if there is no relation between $\mathbf{R}$ and $\mathbf{C}$ as one could not infer anything about $\mathbf{R}$ from $\mathbf{X}_c$. Therefore, we must assume that there is some common information between $\mathbf{R}$ and $\mathbf{C}$. Let $\mathbf{R} = \mathbf{G}\mathbf{G}^H$ be the Cholesky decomposition of $\mathbf{R}$ where $\mathbf{G} = \mathrm{chol}(\mathbf{R})$ is a lower triangular matrix with real-valued positive diagonal entries. We assume here that $\mathrm{chol}(\mathbf{C}) = \mathbf{G}\mathbf{D}^{-1/2}$ where $\mathbf{D}$ is a diagonal matrix. The case of no covariance mismatch between $\mathbf{X}_b$ and $\mathbf{X}_c$ corresponds to $\mathbf{D} = \mathbf{I}_M$. When $\mathbf{D}$ is proportional to the identity matrix, one recovers a partially homogeneous environment where $\mathbf{R}$ and $\mathbf{C}$ differ only by a scaling factor. Note that considering a diagonal $\mathbf{D}$ is not equivalent to assuming that $\mathbf{R}$ and $\mathbf{C}$ share the same eigenvectors but have different eigenvalues, yet it is somewhat similar.

Let us partition $\mathbf{G}$ and $\mathbf{D}$ as

$$\mathbf{G} = \begin{pmatrix}\mathbf{G}_{11} & \mathbf{0}\\ \mathbf{G}_{21} & G_{22}\end{pmatrix}; \ \mathbf{D} = \begin{pmatrix}\mathbf{D}_1 & \mathbf{0}\\ \mathbf{0} & d_2\end{pmatrix} \tag{3}$$

where $\mathbf{G}_{11}$ and $\mathbf{D}_1$ are $(M-1)\times(M-1)$ matrices, and let us observe that if $\mathbf{F} = \mathrm{chol}(\mathbf{R} + P\mathbf{e}_M\mathbf{e}_M^H)$ then $\mathbf{G}^{-1}\mathbf{F} = \left(\begin{smallmatrix}\mathbf{I}_{M-1} & \mathbf{0}\\ \mathbf{0} & \lambda^{1/2}\end{smallmatrix}\right)$ where $\lambda = 1 + PG_{22}^{-2} = 1 + P\tilde{\mathbf{v}}^H\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{v}}$ [20]. Now, let $\mathbf{S}_c = \mathbf{X}_c\mathbf{X}_c^H$ and $\mathbf{L}_c = \mathrm{chol}(\mathbf{S}_c)$. Then $\mathbf{L}_c \stackrel{d}{=} \mathrm{chol}(\mathbf{C})\mathbf{T}$ where $\mathbf{T} = \mathrm{chol}(\mathbf{W})$ with $\mathbf{W} \stackrel{d}{=} \mathbb{C}\mathcal{W}(T_c, \mathbf{I}_M)$. We partition $\mathbf{T}$ as in (3). Since $\mathbf{X} = \begin{bmatrix}\mathbf{X}_a & \mathbf{X}_b\end{bmatrix} \stackrel{d}{=} \begin{bmatrix}\mathbf{F}\mathbf{N}_a & \mathbf{G}\mathbf{N}_b\end{bmatrix}$ with $\mathbf{N} = \begin{bmatrix}\mathbf{N}_a & \mathbf{N}_b\end{bmatrix} \stackrel{d}{=} \mathbb{C}\mathcal{N}(\mathbf{0}, \mathbf{I}_M, \mathbf{I}_{T_p})$, one can write the whitened data $\mathbf{Y} = \mathbf{L}_c^{-1}\mathbf{X}$ as

$$\begin{aligned}\mathbf{Y} &= \mathbf{L}_c^{-1}\mathbf{X}\\ &\stackrel{d}{=} \mathbf{T}^{-1}\mathbf{D}^{1/2}\mathbf{G}^{-1}\begin{bmatrix}\mathbf{F}\mathbf{N}_a & \mathbf{G}\mathbf{N}_b\end{bmatrix}\\ &= \mathbf{T}^{-1}\mathbf{D}^{1/2}\left[\begin{pmatrix}\mathbf{I}_{M-1} & \mathbf{0}\\ \mathbf{0} & \lambda^{1/2}\end{pmatrix}\mathbf{N}_a \quad \mathbf{N}_b\right]\end{aligned} \tag{4}$$

Let us partition $\mathbf{N} = \begin{pmatrix} \mathbf{N}_{a1} & \mathbf{N}_{b1} \\ \mathbf{N}_{a2} & \mathbf{N}_{b2} \end{pmatrix} = \begin{pmatrix} \mathbf{N}_1 \\ \mathbf{N}_2 \end{pmatrix}$. where $\mathbf{N}_{a1}, \mathbf{N}_{b1}$ contains the $M-1$ first rows of $\mathbf{N}_a, \mathbf{N}_b$. Then,

$$\begin{aligned} \mathbf{Y} &\overset{d}{=} \begin{pmatrix} \mathbf{T}_{11}^{-1}\mathbf{D}_1^{1/2} & \mathbf{0} \\ -T_{22}^{-1}\mathbf{T}_{21}\mathbf{T}_{11}^{-1}\mathbf{D}_1^{1/2} & T_{22}^{-1}d_2^{1/2} \end{pmatrix} \begin{pmatrix} \mathbf{N}_{a1} & \mathbf{N}_{b1} \\ \lambda^{1/2}\mathbf{N}_{a2} & \mathbf{N}_{b2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{T}_{11}^{-1}\mathbf{D}_1^{1/2}\mathbf{N}_1 \\ -T_{22}^{-1}\mathbf{T}_{21}\mathbf{T}_{11}^{-1}\mathbf{D}_1^{1/2}\mathbf{N}_1 + T_{22}^{-1}d_2^{1/2}\mathbf{N}_2\mathbf{\Omega}^{1/2} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{a1} & \mathbf{Y}_{b1} \\ \mathbf{Y}_{a2} & \mathbf{Y}_{b2} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_a & \mathbf{Y}_b \end{pmatrix} \end{aligned} \tag{5}$$

where $\mathbf{\Omega} = \begin{pmatrix} \lambda\mathbf{I}_{T_a} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{T_b} \end{pmatrix}$. Therefore, we obtain the following stochastic representation:

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{T}_{11}^{-1}\mathbf{D}_1^{1/2}\mathbf{N}_1 \\ \mathbf{Y}_2 &= T_{22}^{-1}(-\mathbf{T}_{21}\mathbf{Y}_1 + d_2^{1/2}\mathbf{N}_2\mathbf{\Omega}^{1/2}) \end{aligned} \tag{6}$$

Now $T_{22}^2 \overset{d}{=} \mathbb{C}\chi_L^2$ with $L = T_c - M + 1$ and $\mathbf{T}_{21} \overset{d}{=} \mathbb{C}\mathcal{N}(\mathbf{0}, 1, \mathbf{I}_{M-1})$ and these variables are independent [23], and independent of $\mathbf{N}_2 \overset{d}{=} \mathbb{C}\mathcal{N}(\mathbf{0}, 1, \mathbf{I}_{T_p})$. It ensues that

$$-\mathbf{T}_{21}\mathbf{Y}_1 + d_2^{1/2}\mathbf{N}_2\mathbf{\Omega}^{1/2}|\mathbf{Y}_1 \overset{d}{=} \mathbb{C}\mathcal{N}\left(\mathbf{0}, 1, d_2\mathbf{\Omega} + \mathbf{Y}_1^H\mathbf{Y}_1\right) \tag{7}$$

and therefore

$$\mathbf{Y}_2|\mathbf{Y}_1 \overset{d}{=} \frac{\mathbb{C}\mathcal{N}\left(\mathbf{0}, 1, d_2\mathbf{\Omega} + \mathbf{Y}_1^H\mathbf{Y}_1\right)}{\sqrt{\mathbb{C}\chi_L^2}} \tag{8}$$

follows a complex Student distribution, whose density is given by

$$\begin{aligned} p(\mathbf{Y}_2|\mathbf{Y}_1; d_2, \lambda) &= \frac{\Gamma(T_p + L)}{\pi^{T_p}\Gamma(L)}|d_2\mathbf{\Omega} + \mathbf{Y}_1^H\mathbf{Y}_1|^{-1} \\ &\times \left[1 + \mathbf{Y}_2(d_2\mathbf{\Omega} + \mathbf{Y}_1^H\mathbf{Y}_1)^{-1}\mathbf{Y}_2^H\right]^{-(T_p+L)} \end{aligned} \tag{9}$$

First observe that when $d_2 = 1$ and $T_b = 0$, we recover the distribution derived in [20]. Next note that this conditional density depends only on $d_2$ and $\lambda$ while *the distribution of* $\mathbf{Y}_1$ is that of $\mathbf{T}_{11}^{-1}\mathbf{D}_1^{1/2}\mathbf{N}_1$ and depends on $\mathbf{D}_1$ only, but *is the same under* $H_0$ *and* $H_1$. Consequently the generalized likelihood ratio is given by

$$\begin{aligned} \mathrm{GLR}(\mathbf{Y}) &= \frac{\max_{\mathbf{D}_1, d_2, \lambda} p(\mathbf{Y}_2|\mathbf{Y}_1; d_2, \lambda)p(\mathbf{Y}_1; \mathbf{D}_1)}{\max_{\mathbf{D}_1, d_2} p(\mathbf{Y}_2|\mathbf{Y}_1; d_2, 1)p(\mathbf{Y}_1; \mathbf{D}_1)} \\ &= \frac{\max_{d_2, \lambda} p(\mathbf{Y}_2|\mathbf{Y}_1; d_2, \lambda)}{\max_{d_2} p(\mathbf{Y}_2|\mathbf{Y}_1; d_2, 1)} \end{aligned} \tag{10}$$

and hence maximization of $p(\mathbf{Y}_1; \mathbf{D}_1)$ is not actually required. It implies that the distribution of $\mathbf{Y}_1$ does not need to be derived, only the distribution of $\mathbf{Y}_2|\mathbf{Y}_1$ is necessary. We also observe that the joint distribution of $(\mathbf{Y}_1, \mathbf{Y}_2)$ depends on $\mathbf{R}$ -through $\lambda$- and on $\mathbf{D}$. Under $H_0$ $\lambda = 1$ and hence this distribution does no longer depend on $\mathbf{R}$, which means that the GLR has a constant false alarm rate with respect to $\mathbf{R}$. However, the distribution of $\mathrm{GLR}(\mathbf{Y})$ still depends on $\mathbf{D}$. This is to be contrasted with the CA detector whose distribution does not depend on any unknown parameter under $H_0$. Finally, for calculation of the GLR, one needs to solve a 1-D maximization problem (with respect to $d_2 > 0$) under $H_0$ and a 2-D maximization problem under $H_1$ where $d_2 > 0$ and $\lambda \geq 1$ are the unknowns. Observe that $p(\mathbf{Y}_2|\mathbf{Y}_1; d_2, \lambda)$ involves the matrix $\mathbf{\Gamma} = d_2\mathbf{\Omega} + \mathbf{Y}_1^H\mathbf{Y}_1$ which is of size $T_p \times T_p$ and hence the computational complexity is not prohibitive. For instance $p(\mathbf{Y}_2|\mathbf{Y}_1; d_2, \lambda)$ can be evaluated rather simply on a grid of values of $(d_2, \lambda)$ over which the maximum is searched. Actually, in order to compute the determinant and the inverse of $\mathbf{\Gamma}$ one can partition the latter as

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_{aa} & \mathbf{\Gamma}_{ab} \\ \mathbf{\Gamma}_{ba} & \mathbf{\Gamma}_{bb} \end{pmatrix} = \begin{pmatrix} d_2\lambda\mathbf{I}_{T_a} + \mathbf{Y}_{a1}^H\mathbf{Y}_{a1} & \mathbf{Y}_{a1}^H\mathbf{Y}_{b1} \\ \mathbf{Y}_{b1}^H\mathbf{Y}_{a1} & d_2\mathbf{I}_{T_b} + \mathbf{Y}_{b1}^H\mathbf{Y}_{b1} \end{pmatrix} \tag{11}$$

and use the fact that $|\mathbf{\Gamma}| = |\mathbf{\Gamma}_{a.b}||\mathbf{\Gamma}_{bb}|$ where $\mathbf{\Gamma}_{a.b} = \mathbf{\Gamma}_{aa} - \mathbf{\Gamma}_{ab}\mathbf{\Gamma}_{bb}^{-1}\mathbf{\Gamma}_{ba}$, and $\mathbf{\Gamma}_{a.b}$ is a scalar when $T_a = 1$. Moreover,

$$\mathbf{\Gamma}^{-1} = \begin{pmatrix} \mathbf{I}_{T_a} \\ -\mathbf{\Gamma}_{bb}^{-1}\mathbf{\Gamma}_{ba} \end{pmatrix} \mathbf{\Gamma}_{a.b}^{-1} \begin{pmatrix} \mathbf{I}_{T_a} & -\mathbf{\Gamma}_{ab}\mathbf{\Gamma}_{bb}^{-1} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Gamma}_{bb}^{-1} \end{pmatrix} \tag{12}$$

so that

$$\begin{aligned} \mathbf{Y}_2\mathbf{\Gamma}^{-1}\mathbf{Y}_2^H &= \mathbf{Y}_{b2}\mathbf{\Gamma}_{bb}^{-1}\mathbf{Y}_{b2}^H \\ &+ (\mathbf{Y}_{a2} - \mathbf{Y}_{b2}\mathbf{\Gamma}_{bb}^{-1}\mathbf{Y}_{b1}^H\mathbf{Y}_{a1})\mathbf{\Gamma}_{a.b}^{-1}(\mathbf{Y}_{a2} - \mathbf{Y}_{b2}\mathbf{\Gamma}_{bb}^{-1}\mathbf{Y}_{b1}^H\mathbf{Y}_{a1})^H \end{aligned} \tag{13}$$

Note that the key matrix here is $\mathbf{Y}_{b1}^H\mathbf{Y}_{b1}$. Once its eigenvalue decomposition or the SVD of $\mathbf{Y}_{b1}$ is computed, all other quantities involves simple matrix-vector products.

*Remark 1:* It should be noted that $p(\mathbf{Y}_2|\mathbf{Y}_1; d_2, \lambda)$ is given by (9) for any matrix $\mathbf{D}_1$, i.e., *not necessarily diagonal*, since the second line of (6) holds true whatever $\mathbf{D}_1$. Therefore, the GLR is still given by (10) under the more general case of a *block-diagonal* $\mathbf{D}$. It turns out (we omit the details for lack of space) that $\mathbf{D}$ being block-diagonal is equivalent to $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{C}}$ verifying the so-called generalized eigenrelation (GER) [10], [24] which states that $\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{v}} = \gamma\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{v}}$.

## III. Numerical Analysis

Numerical simulations are now used to assess the performance of the detector derived above and to compare it with relevant detectors. We consider a scenario where $M = 32$. The disturbance consists of colored clutter plus thermal noise and has a covariance matrix of the form $\mathbf{R} = \mathbf{R}_c + \sigma^2\mathbf{I}_M$ where $\mathbf{R}_c(k, \ell) = e^{-0.5(2\pi\sigma_f|k-\ell|)^2}$ with $\sigma_f = 0.02$. The clutter to noise ratio $\mathrm{CNR} = -10\log_{10}\sigma^2$ is set to $\mathrm{CNR} = 30$ dB. The signal of interest is $\tilde{\mathbf{v}} = \mathbf{e}(f_s)$ where $\mathbf{e}(f) = 1/\sqrt{M}\begin{bmatrix} 1 & e^{2i\pi f} & \dots & e^{2i\pi(M-1)f} \end{bmatrix}^T$ and $f_s = 0.05$. The target amplitude is drawn from a $\mathbb{C}\mathcal{N}(0, P)$ distribution in accordance with (2). We will successively consider the case of a partially homogeneous environment -$\mathbf{D} = \gamma\mathbf{I}_M$-, the case of an arbitrary $\mathbf{D}$ and finally a case where $\mathbf{C} = \mathbf{G}\mathbf{W}^{-1}\mathbf{G}^H$ where $\mathbf{W}$ is not diagonal. The probability of false alarm is set to $P_{fa} = 10^{-3}$ and the probability of detection is evaluated as a function of the signal to noise ratio, defined as $\mathrm{SNR} = P\tilde{\mathbf{v}}^H\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{v}}$.

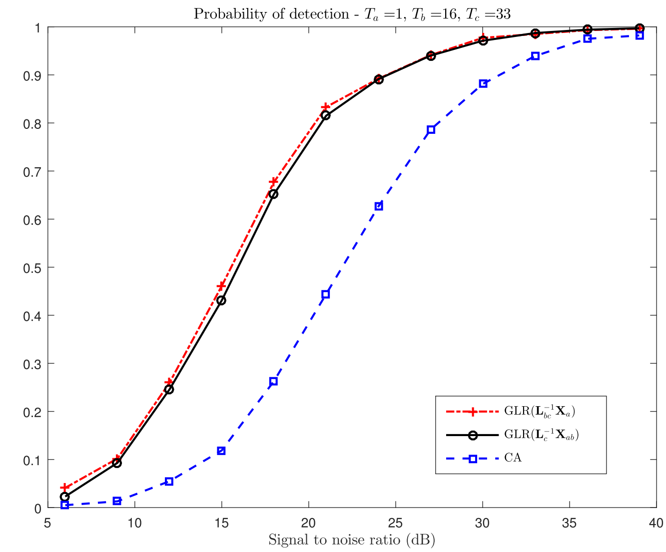


Fig. 2. Probability of detection versus SNR in a partially homogeneous environment: $\mathbf{C} = d^{-1}\mathbf{R}$ with $d^{-1} = -6$ dB. $T_a = 1$, $T_b = M/2$ and $T_c = M+1$.

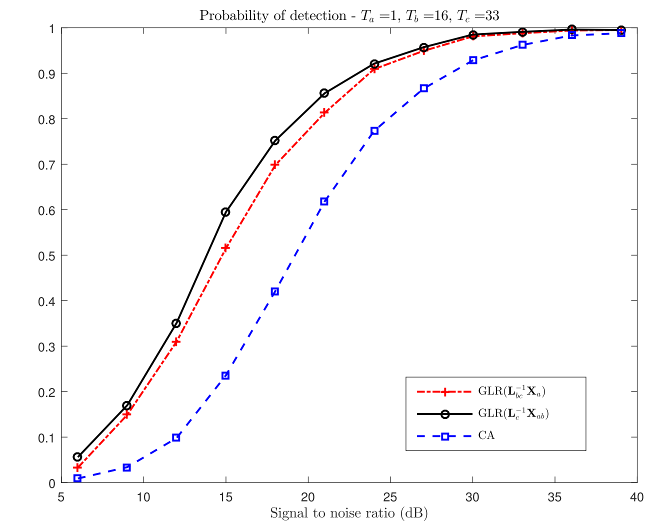


Fig. 3. Probability of detection versus SNR when $\mathbf{C} = \mathbf{G}\mathbf{D}^{-1}\mathbf{G}^H$ with $\mathbf{D} = \text{diag}(d_1, \ldots, d_M)$. $10\log_{10} d_m^{-1}$ is drawn from a uniform distribution with mean $-6$ dB and standard deviation $-6$ dB. $T_a = 1$, $T_b = M/2$ and $T_c = M+1$.

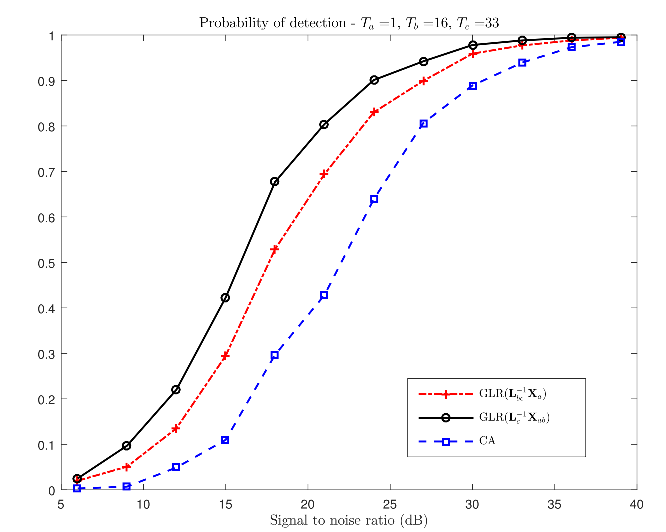


Fig. 4. Probability of detection versus SNR when $\mathbf{C} = \mathbf{G}\mathbf{W}^{-1}\mathbf{G}^H$ where $\mathbf{W}$ is drawn from a Wishart distribution $\mathbb{C}\mathcal{W}(\nu, \mu\mathbf{I}_M)$. $\nu = M+2$ and $\mathbb{E}\{[\mathbf{W}^{-1}]_{mm}\} = -6$dB. $T_a = 1$, $T_b = M/2$ and $T_c = M+1$.

The detector derived above based on whitening of $(\mathbf{X}_a, \mathbf{X}_b)$ using the SCM of $\mathbf{X}_c$ will be denoted by $\text{GLR}(\mathbf{L}_c^{-1}\mathbf{X}_{ab})$. The cell averaging detector of [19] will be denoted as CA. For comparison purposes, we also use the detector of [20] which is the GLRT based on $\mathbf{L}_{bc}^{-1}\mathbf{X}_a$ where $\mathbf{L}_{bc} = \text{chol}(\mathbf{X}_b\mathbf{X}_b^H + \mathbf{X}_c\mathbf{X}_c^H)$. This detector does not assume that there exists a mismatch between $\mathbf{X}_b$ and $\mathbf{X}_c$. Since its performance is equivalent or slightly better than that of Kelly's GLRT, we only display $\text{GLR}(\mathbf{L}_{bc}^{-1}\mathbf{X}_a)$.

### A. Partially Homogeneous Environment

We first consider the case of a partially homogeneous environment where $\mathbf{C} = d^{-1}\mathbf{R}$ with $d^{-1} = -6$dB. The results are displayed in Figure 2. It can be observed that the new detector significantly improves over cell averaging. Yet, it does not perform better than $\text{GLR}(\mathbf{L}_{bc}^{-1}\mathbf{X}_a)$ which seems quite robust. This can be attributed to the fact that $\text{GLR}(\mathbf{L}_c^{-1}\mathbf{X}_{ab})$ is based on a model which contains $M$ unknowns (the diagonal elements of $\mathbf{D}$) whereas here a single parameter $d$ explains the relation between $\mathbf{C}$ and $\mathbf{R}$.

### B. Case of Arbitrary $\mathbf{D}$

In the following simulation, we consider an arbitrary matrix $\mathbf{D}$. This matrix is fixed but $d_m^{-1}$ was generated from a uniform distribution such that the mean value of $d_m^{-1}$ is $\mathbb{E}\{d_m^{-1}\} = -6$dB, and the standard deviation is equal to the mean. This scenario fits the model used by $\text{GLR}(\mathbf{L}_c^{-1}\mathbf{X}_{ab})$ and the latter, as seen in Figure 3, provides the best performance. It is still much better than CA and slightly better than $\text{GLR}(\mathbf{L}_{bc}^{-1}\mathbf{X}_a)$.

### C. Robustness in the Case $\mathbf{C} = \mathbf{G}\mathbf{W}^{-1}\mathbf{G}^H$

Finally, we test the robustness of the detector when the matrix $\mathbf{C} = \mathbf{G}\mathbf{W}^{-1}\mathbf{G}^H$ where $\mathbf{W}$ is not diagonal but some positive definite matrix. More precisely, $\mathbf{W}$ was drawn from a complex Wishart distribution $\mathbb{C}\mathcal{W}(\nu, \mu\mathbf{I}_M)$ with $\nu = M+2$ and $\mathbb{E}\{[\mathbf{W}^{-1}]_{mm}\} = -6$dB. Note that this scenario does not correspond to the hypotheses used by $\text{GLR}(\mathbf{L}_c^{-1}\mathbf{X}_{ab})$. Interestingly enough the latter performs very well and the improvement compared to $\text{GLR}(\mathbf{L}_{bc}^{-1}\mathbf{X}_a)$ increases.

## IV. Conclusions

In this letter, we addressed the problem of detecting a Swerling I target when some of the training samples undergo a covariance mismatch. The approach is based on whitening the data under test and the matched training samples using the mismatched training samples. We derived the GLRT based on the distribution of the whitened data. Numerical simulations showed that it offers a significant improvement compared to cell averaging and is rather robust.

## REFERENCES


[1] M. A. Richards, J. A. Scheer, and W. A. Holm, Eds., *Principles of Modern Radar: Basic Principles*, vol. 1. London, U.K.: Institution Engineering Technology, 2010.

[2] W. L. Melvin and J. A. Scheer, Eds., *Principles of Modern Radar: Advanced Principles*, vol. 2. London, U.K.: Institution Engineering Technology, 2012.

[3] E. J. Kelly, "Adaptive detection in non-stationary interference, Part I and Part II," Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, MA, USA, Tech. Rep. 724, Jun. 1985.

[4] E. J. Kelly, "Adaptive detection in non-stationary interference, Part III," Massachusetts Institute of Technology, Lincoln Laboratory, 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 Institute of Technology, Lincoln Laboratory, Lexington, MA, USA, Tech. Rep. 848, Apr. 1989.

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

[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. Blum and K. F. McDonald, "Analysis of STAP algorithms for cases with mismatched steering and clutter statistics," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 301–310, Feb. 2000.

[9] K. F. McDonald and R. S. Blum, "Exact performance of STAP algorithms with mismatched steering and clutter statistics," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2750–2763, Oct. 2000.

[10] C. D. Richmond, "Performance of a class of adaptive detection algorithms in nonhomogeneous environments," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1248–1262, May 2000.

[11] R. S. Raghavan, "False alarm analysis of the AMF algorithm for mismatched training," *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 83–96, Jan. 2019.

[12] 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.

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

[14] K. Gerlach, "The effects of signal contamination on two adaptive detectors," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 1, pp. 297–309, Jan. 1995.

[15] K. Gerlach, "Outlier resistant adaptive matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 885–901, Jul. 2002.

[16] O. Besson, "Detection in the presence of surprise or undernulled interference," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 352–354, May 2007.

[17] O. Besson and D. Orlando, "Adaptive detection in non-homogeneous environments using the generalized eigenrelation," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 731–734, Oct. 2007.

[18] V. Carotenuto, A. De Maio, D. Orlando, and L. Pallotta, "Adaptive radar detection using two sets of training data," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1791–1801, Apr. 2018.

[19] R. S. Raghavan, "A CFAR detector for mismatched eigenvalues of training sample covariance matrix," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4624–4635, Sep. 2019.

[20] O. Besson, "Detection of Gaussian signal using adaptively whitened data," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 430–434, Mar. 2019.

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

[22] 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.

[23] 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.

[24] C. D. Richmond, "Statistics of adaptive nulling and use of the generalized eigenrelation (GER) for modeling inhomogeneities in adaptive processing," *IEEE Trans. Signal Process.*, vol. 48, no. 5, pp. 1263–1273, May 2000.