Adaptive Radar Detection Using Two Sets of Training Data

Vincenzo Carotenuto, *Member, IEEE*, Antonio De Maio, *Fellow, IEEE*, Danilo Orlando, *Senior Member, IEEE*, and Luca Pallotta, *Member, IEEE*

***Abstract*— This paper deals with adaptive radar detection of a point-like target in a homogeneous environment characterized by the presence of clutter, jamming, and radar internal noise. At the design stage two training datasets, whose gathering is carefully motivated in the paper, are considered to get receiver adaptation. Hence, the maximum likelihood estimator of the interference covariance matrix for the cell under test is computed exploiting both the available secondary sets. This estimate is then used to build two adaptive decision rules based on the two-step generalized likelihood ratio test and Rao test criteria. Remarkably, they are not limited by the conventional constraint on the cardinality of the classic training dataset. At the analysis stage, the detection performances of the newly proposed detectors are compared with those of the analogous conventional counterparts and the interplay among the different parameters of the problem is thoroughly studied.**

***Index Terms*— Adaptive Radar Detection, Generalized Likelihood Ratio Test, Interference Covariance Matrix, Maximum Likelihood Estimation, Rao test.**

I. Introduction

In radar detection based upon array of sensors or pulse trains [1]–[3], the Interference Covariance Matrix (ICM) might comprise several contributions associated with different interference sources. The most common sources are

- the electronic devices generating thermal noise, which is always present and leads to an ICM component modeled as $\sigma_n^2\boldsymbol{I}$, where $\sigma_n^2 > 0$ is the thermal noise power and $\boldsymbol{I}$ is the identity matrix;
- the specific operating environment, whose backscattering gives rise to the clutter component;
- jamming systems which can significantly modify the ICM depending on the jamming type and the radar signal processing. More precisely, if the radar performs spatial processing using N sensors, then K_J barrage noise jammers [4] illuminating the victim radar from the angle of arrival $\theta_{J,k}$, $k = 1, \ldots, K_J$, would yield the following ICM component

$$\sum_{k=1}^{K_J} \sigma_{J,k}^2 \boldsymbol{v}(\theta_{J,k})\boldsymbol{v}^\dagger(\theta_{J,k}), \tag{1}$$

where $(\cdot)^\dagger$ denotes conjugate transpose, $\sigma_{J,k}^2 > 0$ and

$$\boldsymbol{v}(\theta_{J,k}) = [1 \; e^{j2\pi\nu_s(\theta_{J,k})} \; \ldots \; e^{j2\pi\nu_s(\theta_{J,k})(N-1)}]^T \tag{2}$$

are the power and the spatial steering vector of the k-th jammer. In the last equation, $(\cdot)^T$ denotes transpose and $\nu_s(\cdot)$ is the spatial frequency and is a function of $\theta_{J,k}$. In the case of temporal processing, a noise jammer usually produces an ICM contribution equal to $\sigma_J^2\boldsymbol{I}$, where $\sigma_J^2 > 0$ is the overall jammer power. Finally, still in the context of temporal processing, equation (1) can be modified to account for the effect of multi-tone continuous-wave jammers replacing $\boldsymbol{v}(\theta_{J,k})$ with

$$\boldsymbol{v}(\nu_{J,k}) = [1 \; e^{j2\pi\nu_{J,k}} \; \ldots \; e^{j2\pi\nu_{J,k}(N-1)}]^T, \tag{3}$$

where K_J is the number of transmitted tones and $\nu_{J,k}$ is the normalized Doppler frequency of the kth tone. Such jammers can emulate the radar cross section of Swerling 1 targets.

Note that the physical mechanisms involved in the mentioned interference sources are significantly different. Specifically, thermal noise component, generated by the electronic devices, affects all data and is always present in the course of radar operation. On the other hand, the mentioned jammer component depends on the presence of jamming systems in the surveillance area and (when it is present) contaminates all the range cells processed by the radar[1]. Finally, clutter component originates from echoes backscattered by objects which are not of interest to the radar [6]. Clutter strength and, consequently, its range extension depends on the carrier frequency, the transmitted power, the environment morphology, and the grazing angle [6]. It follows that the clutter component is in general range-dependent and tied up to the transmitted waveform.

Based upon the physics behind these phenomena, it is possible to acquire data free of clutter components and affected by the remaining interference sources only. For instance, consider a system employing pulse-to-pulse frequency agility which transmits one pulse using N antennas (spatial processing only case), then it is clear that there do not exist clutter echoes before transmitting the pulse waveform and, at this stage, it is possible to record clutter-free data (see Figure 1). In addition, due to frequency agility, the bandpass filter, tuned around the current frequency, would eliminate any residual clutter

A. De Maio is with the Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, Università degli Studi di Napoli "Federico II", via Claudio 21, I-80125 Napoli, Italy. E-mail: ademaio@unina.it.

D. Orlando is with the Engineering Faculty of Università degli Studi "Niccolò Cusano", via Don Carlo Gnocchi 3, 00166 Roma, Italy. E-mail: danilo.orlando@unicusano.it.

V. Carotenuto and L. Pallotta are with CNIT, viale G.P. Usberti, n. 181/A - 43124 Parma, c/o udr Università "Federico II", via Claudio 21, I-80125 Napoli, Italy. E-mail: vincenzo.carotenuto@unina.it,luca.pallotta@unina.it

[1]Observe that jamming systems are also capable of generating coherent echoes which are not range distributed and occupy a specific range cell. These electronic countermeasures are contrasted by means of SideLobe Blanker [5] techniques which are not the object of this paper.

component produced by previous transmissions. Another example of practical interest concerns radar systems transmitting coherent pulse trains. In this case, if the Pulse Repetition Interval (PRI) is sufficiently high, then data collected before transmitting the next pulse within the same burst and at high ranges (or after the instrumental range), result free of clutter contribution (see Figure 2). These data (free of clutter components) can be suitably exploited to achieve a more reliable estimate of the ICM, which could be useful in many radar signal processing applications (e.g., adaptive interference cancellation, beamforming, adaptive detection). As a matter of fact, in conventional radar systems, the ICM is unknown and must be estimated from data collecting training samples in the proximity of the cell under test (CUT) as corroborated by a plethora of examples in the open literature. For instance in [1], [2], it is assumed that a set of training data is available at the receiver with the constraint that the number of data is greater than the vector size to ensure the estimate invertibility with probability 1. This constraint can be relaxed using a priori information about the clutter, the system geometry, or diagonal loading as done in [7]–[12]. Other examples can be found in [13]–[23, and references therein].

In this paper, we propose an alternative approach relying on the joint exploitation of two sets of secondary data. Specifically, the first set is acquired either before transmission of the pulse (spatial processing) or from far range bins (temporal processing with a high PRI), whereas, the second set comes from the conventional radar reference window surrounding the CUT. Notice that data belonging to the reference window share the same ICM components as the CUT including the clutter component, while the ICM of data gathered during the first training phase do not contain the clutter component. Hence, the detection problem is formulated as a binary hypothesis test in the presence of two sets of secondary data each of them representative of different interference components. This problem is solved resorting to the two-step Generalized Likelihood Ratio Test (GLRT) design procedure [2] and the Rao test [24] to come up with two adaptive decision schemes capable of taking advantage of the information carried by the additional training data set. Both architectures exploit the ICM Maximum Likelihood Estimate (MLE) whose closed form expression is derived using the aforementioned two secondary data sets. This derivation represents the main technical innovation of this paper from the signal processing point of view.

Remarkably, it is proven that the expression of the MLE does not suppose that the cardinality of the classic training set is greater than the vector size. This constraint is required by conventional adaptive decision schemes, which use only this training set for estimation purposes, in order to make the sample covariance matrix invertible with probability 1. However, this requirement might not be met in practical scenarios, because the volume of homogeneous data belonging to the reference window can be limited by the heterogeneity of the surveillance area [25]–[27]. On the other hand, the formation of the first data set is not conditioned by this limitation, since, in the course of the acquisition, the factors producing the heterogeneity do not come into play. As a consequence, the proposed architectures share the capability of operating in sample-starved environments [28], [29].

The performance analysis is aimed at investigating the behavior of the proposed architectures in terms of Probability of Detection P_D versus the main parameters of interest for a preassigned Probability of False Alarm P_{fa}. Moreover, comparisons with conventional detectors are also considered to corroborate the double training idea.

The remainder of this paper is organized as follows. Section II is devoted to the problem formulation. The adaptive architectures are devised in Section III, where the two-step GLRT and the Rao test design criteria are applied. Section IV provides the performance analysis by means of numerical examples. Finally, some concluding remarks and future research tracks are given in Section V. Proofs and derivations are confined in the appendix.

Notation

Vectors and matrices are denoted by boldface lower-case and upper-case letters, respectively. Symbols $\det(\cdot)$ and $\mathrm{Tr}(\cdot)$ denote the determinant and the trace of a square matrix, respectively. Symbols $\boldsymbol{I}$ and $\mathbf{0}$ represent the identity matrix and the null vector or matrix of suitable dimensions, respectively. The imaginary unit is denoted by j. Given a vector $\boldsymbol{a}$, $\mathbf{diag}\,(\boldsymbol{a})$ indicates the diagonal matrix whose ith diagonal element is the ith entry of $\boldsymbol{a}$. The curled inequality symbol $\succeq$ (and its strict form $\succ$) is used to denote generalized matrix inequality: for any hermitian matrix $\boldsymbol{A}$, $\boldsymbol{A} \succeq \mathbf{0}$ means that $\boldsymbol{A}$ is a positive semi-definite matrix ($\boldsymbol{A} \succ \mathbf{0}$ for positive definiteness). As to the numerical sets, $\mathbb{R}$ is the set of real numbers, $\mathbb{R}^{N\times M}$ is the set of $(N \times M)$-dimensional real matrices (or vectors if $M = 1$), $\mathbb{C}$ is the set of complex numbers, and $\mathbb{C}^{N\times M}$ is the set of $(N \times M)$-dimensional complex matrices (or vectors if $M = 1$). The (k, l)-entry (or l-entry) of a generic matrix $\boldsymbol{A}$ (or vector $\boldsymbol{a}$) is denoted by $\boldsymbol{A}(k, l)$ (or $\boldsymbol{a}(l)$). Symbols $\Re\{z\}$ and $\Im\{z\}$ indicate the real and imaginary parts of $z \in \mathbb{C}$, respectively. We use $(\cdot)^T$ and $(\cdot)^\dagger$ to denote transpose and conjugate transpose, respectively. The acronym iid means independent and identically distributed. The Signal-to-Noise plus Interference Ratio, the Clutter-to-Noise Ratio, and the Jammer-to-Noise Ratio are denoted by SINR, CNR, and JNR, respectively. Finally, we write $\boldsymbol{x} \sim \mathcal{CN}_N(\boldsymbol{m}, \boldsymbol{M})$ if $\boldsymbol{x}$ is a complex circular N-dimensional normal vector with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{M} \succ \mathbf{0}$, whereas $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M] \sim \mathcal{CN}_{N,M}(\boldsymbol{m}, \boldsymbol{M}, \boldsymbol{I})$ if $\boldsymbol{x}_i \sim \mathcal{CN}_N(\boldsymbol{m}, \boldsymbol{M})$, $i = 1, \ldots, M$, and iid.

II. Problem Formulation

As stated in the previous section, adaptive radar detection requires the ICM to be somehow estimated from data. To this end, the system collects a set of secondary data $\boldsymbol{Z}_2 = [\boldsymbol{z}_{2,1}, \ldots, \boldsymbol{z}_{2,M}] \in \mathbb{C}^{N\times M}$ with N the number of system channels, in the proximity of the CUT, $\boldsymbol{z} \in \mathbb{C}^{N\times 1}$ say. Before proceeding, we make clear that the subscript 2 means that the corresponding vector and/or matrix belongs to what we have called second (or classic) training data set.

Now, the conventional detection problem is often cast as the following binary hypothesis test

$$\begin{cases} H_1 : \begin{cases} \boldsymbol{z} = \alpha \boldsymbol{v} + \boldsymbol{n}, \\ \boldsymbol{z}_{2,m} = \boldsymbol{n}_{2,m}, \quad m = 1, \ldots, M, \end{cases} \\ H_0 : \begin{cases} \boldsymbol{z} = \boldsymbol{n}, \\ \boldsymbol{z}_{2,m} = \boldsymbol{n}_{2,m}, \quad m = 1, \ldots, M, \end{cases} \end{cases} \tag{4}$$

where

- $\boldsymbol{n}, \boldsymbol{n}_{2,m} \sim \mathcal{CN}_N(\boldsymbol{0}, \boldsymbol{M}_2)$, $m = 1, \ldots, M$, are iid vectors representing the interference echoes;
- $\alpha = \alpha_r + j\alpha_i$, with $\alpha_r, \alpha_i \in \mathbb{R}$, is an amplitude factor accounting for target response and propagation effects;
- $\boldsymbol{v}$ is the nominal steering vector.

In the open literature, there exist several solutions to the above detection problem. The seminal and most cited architectures are Kelly's GLRT [1] and the Adaptive Matched Filter (AMF) [2], whose expressions are

$$\frac{|\boldsymbol{z}^\dagger \boldsymbol{S}_2^{-1} \boldsymbol{v}|^2}{\boldsymbol{v}^\dagger \boldsymbol{S}_2^{-1} \boldsymbol{v} (1 + \boldsymbol{z}^\dagger \boldsymbol{S}_2^{-1} \boldsymbol{z})} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{5}$$

and

$$\frac{|\boldsymbol{z}^\dagger \boldsymbol{S}_2^{-1} \boldsymbol{v}|^2}{\boldsymbol{v}^\dagger \boldsymbol{S}_2^{-1} \boldsymbol{v}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{6}$$

respectively, where η is the detection threshold[2] set to ensure the desired P_{fa} and $\boldsymbol{S}_2 = \boldsymbol{Z}_2 \boldsymbol{Z}_2^\dagger$ is M-times the Sample Covariance Matrix (SCM) of the conventional secondary data[3]. However, the SCM is representative of the sum of the different ICM components and requires $M \geq N$.

The proposed approach consists in exploiting, along with $\boldsymbol{Z}_2$ without any constraint on M, an additional set of training data $\boldsymbol{Z}_1 = [\boldsymbol{z}_{1,1}, \ldots, \boldsymbol{z}_{1,K}]$ statistically independent of $\boldsymbol{Z}_2$ and collected as described in Section I. With the above modifications in mind, the detection problem at hand becomes

$$\begin{cases} H_1 : \begin{cases} \boldsymbol{z} = \alpha \boldsymbol{v} + \boldsymbol{n} \\ \boldsymbol{z}_{1,k} = \boldsymbol{n}_{1,k}, \quad k = 1, \ldots, K, \\ \boldsymbol{z}_{2,m} = \boldsymbol{n}_{2,m}, \quad m = 1, \ldots, M, \end{cases} \\ H_0 : \begin{cases} \boldsymbol{z} = \boldsymbol{n} \\ \boldsymbol{z}_{1,k} = \boldsymbol{n}_{1,k}, \quad k = 1, \ldots, K, \\ \boldsymbol{z}_{2,m} = \boldsymbol{n}_{2,m}, \quad m = 1, \ldots, M, \end{cases} \end{cases} \tag{7}$$

where $\boldsymbol{n}_{1,k} \sim \mathcal{CN}_N(\boldsymbol{0}, \boldsymbol{M}_1)$, $k = 1, \ldots, K \geq N$, are[4] iid vectors representing the contribution of thermal noise and of a prospective jammer. Moreover, the above vectors are statistically independent of $\boldsymbol{n}$.

An important remark is now in order. Matrices $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ result from the contributions of different interference sources, more precisely the following cases can likely occur:

- the first training data set contains thermal noise only, while in the second one a clutter component is also present, namely
$$\boldsymbol{M}_1 = \sigma_n^2 \boldsymbol{I} \quad \text{and} \quad \boldsymbol{M}_2 = \sigma_n^2 \boldsymbol{I} + \boldsymbol{M}_c, \tag{8}$$
where $\boldsymbol{M}_c \succeq \boldsymbol{0}$ is the clutter component;
- the first training data set is representative of thermal noise and jammer, while the second one also contains a clutter component, namely
$$\boldsymbol{M}_1 = \sigma_n^2 \boldsymbol{I} + \boldsymbol{M}_J, \tag{9}$$
$$\boldsymbol{M}_2 = \sigma_n^2 \boldsymbol{I} + \boldsymbol{M}_c + \boldsymbol{M}_J, \tag{10}$$
where $\boldsymbol{M}_J \succeq \boldsymbol{0}$ is the jamming component covariance.

Thus, in each case, the inequality $\boldsymbol{M}_2 - \boldsymbol{M}_1 \succeq \boldsymbol{0}$ holds, or, equivalently, $\boldsymbol{M}_2 = \boldsymbol{M}_1 + \boldsymbol{R}$ with $\boldsymbol{R} \succeq \boldsymbol{0}$.

III. Detector Design

In this section, two adaptive decision rules are devised resorting to the two-step GLRT and to the Rao test design criteria[5]. As preliminary step towards the receiver derivations, some useful definitions are required. Specifically, the joint probability density function (pdf) of $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ under both hypotheses is

$$f(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \boldsymbol{M}_1, \boldsymbol{R}) = \left(\frac{1}{\pi^N}\right)^{K+M} \times \frac{\exp\left\{-\text{tr}\left[\boldsymbol{M}_1^{-1} \boldsymbol{S}_1\right] - \text{tr}\left[(\boldsymbol{M}_1 + \boldsymbol{R})^{-1} \boldsymbol{S}_2\right]\right\}}{\det(\boldsymbol{M}_1)^K \det(\boldsymbol{M}_1 + \boldsymbol{R})^M}, \tag{11}$$

where $\boldsymbol{S}_1 = \boldsymbol{Z}_1 \boldsymbol{Z}_1^\dagger$. On the other hand, the pdf of $\boldsymbol{z}$ depends on which hypothesis is true. Thus, its expression under the H_i, $i = 0, 1$, is given by

$$f_i(\boldsymbol{z}; i\alpha, \boldsymbol{M}_1, \boldsymbol{R}, H_i) = \frac{\exp\left\{-\text{tr}\left[(\boldsymbol{M}_1 + \boldsymbol{R})^{-1} (\boldsymbol{z} - i\alpha\boldsymbol{v})(\boldsymbol{z} - i\alpha\boldsymbol{v})^\dagger\right]\right\}}{\pi^N \det(\boldsymbol{M}_1 + \boldsymbol{R})}. \tag{12}$$

Finally, the joint pdf of $\boldsymbol{z}$, $\boldsymbol{Z}_1$, and $\boldsymbol{Z}_2$ under H_i, $i = 0, 1$, has the following expression

$$f_i(\boldsymbol{z}, \boldsymbol{Z}_1, \boldsymbol{Z}_2; i\alpha, \boldsymbol{M}_1, \boldsymbol{R}, H_i) = f_i(\boldsymbol{z}; i\alpha, \boldsymbol{M}_1, \boldsymbol{R}, H_i) f(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \boldsymbol{M}_1, \boldsymbol{R}). \tag{13}$$

A. GLRT-based Architecture

The following rationale is adopted to design the decision rule

- first, compute the GLRT assuming that $\boldsymbol{M}_1 + \boldsymbol{R}$ is known;
- second, replace $\boldsymbol{M}_1 + \boldsymbol{R}$ with a suitable estimate.

Thus, according to the first step, the GLRT for known $\boldsymbol{M}_1 + \boldsymbol{R}$ is

$$\max_\alpha \frac{f_1(\boldsymbol{z}; \alpha, \boldsymbol{M}_1, \boldsymbol{R}, H_1)}{f_0(\boldsymbol{z}; 0, \boldsymbol{M}_1, \boldsymbol{R}, H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{14}$$

[2]Hereafter, η is used to denote the detection threshold or any proper modification of it for all the considered receivers.

[3]Note that the SCM is the ICM MLE based upon $\boldsymbol{Z}_2$.

[4]Notice that the constraint $K \geq N$ is less difficult than $M \geq N$ since the acquisition of $\boldsymbol{Z}_1$ is not influenced by the environment heterogeneity.

[5]For the problem at hand, the derivation of the plain GLRT is currently under investigation.

Following the lead of [2], it is not difficult to prove that (14) is statistically equivalent to

$$\frac{\left|\boldsymbol{z}^{\dagger}\left(\boldsymbol{M}_1+\boldsymbol{R}\right)^{-1}\boldsymbol{v}\right|^2}{\boldsymbol{v}^{\dagger}\left(\boldsymbol{M}_1+\boldsymbol{R}\right)^{-1}\boldsymbol{v}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{15}$$

Now, instead of using $\boldsymbol{S}_2$ in place of $(\boldsymbol{M}_1+\boldsymbol{R})$, the latter is estimated exploiting the additional training data set to account for the specific components of the ICM. Specifically, the following theorem provides the expression of the MLE of $\boldsymbol{M}_2$ based upon $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$.

Theorem 3.1: Let $\boldsymbol{Z}_1 \sim \mathcal{CN}_{N,K}(\boldsymbol{0},\boldsymbol{M}_1,\boldsymbol{I})$ with $K \geq N$, $\boldsymbol{Z}_2 \sim \mathcal{CN}_{N,M}(\boldsymbol{0},\boldsymbol{M}_2,\boldsymbol{I})$ where $\boldsymbol{M}_2 = \boldsymbol{M}_1 + \boldsymbol{R}$ with $\boldsymbol{R} \succeq \boldsymbol{0}$, and denote by $\lambda_{S,1} \geq \ldots \geq \lambda_{S,N} \geq 0$ the eigenvalues of $\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$ with $\boldsymbol{U}_S \in \mathbb{C}^{N\times N}$ a unitary matrix containing the corresponding eigenvectors. Then, the MLE of $\boldsymbol{M}_1+\boldsymbol{R}$ is

$$\widehat{\boldsymbol{M}}_2 = \boldsymbol{S}_1^{1/2}\boldsymbol{U}_S\boldsymbol{D}_A\widehat{\boldsymbol{\Lambda}}\boldsymbol{D}_A\boldsymbol{U}_S^{\dagger}\boldsymbol{S}_1^{1/2}, \tag{16}$$

where $\boldsymbol{S}_i$, $i=1,2$, have been previously defined, $a = 2K+2M$, $\widehat{\boldsymbol{\Lambda}} = \textbf{diag}\left(\frac{K}{M}\lambda_{S,1},\ldots,\frac{K}{M}\lambda_{S,\hat{r}},1,\ldots,1\right)$, and

$$\boldsymbol{D}_A = \textbf{diag}\left(\frac{1}{\sqrt{K}},\ldots,\frac{1}{\sqrt{K}},\sqrt{\frac{\lambda_{S,\hat{r}+1}+1}{a/2}},\ldots,\sqrt{\frac{\lambda_{S,N}+1}{a/2}}\right) \tag{17}$$

with $\hat{r}$ the number of $\lambda_{S,i}$ greater than M/K.

Proof: See the appendix. ∎

Thus, the adaptivity is achieved replacing $(\boldsymbol{M}_1+\boldsymbol{R})$ in (15) with (16) to come up with

$$\frac{\left|\boldsymbol{z}^{\dagger}\left(\boldsymbol{S}_1^{1/2}\boldsymbol{U}_S\boldsymbol{D}_A\widehat{\boldsymbol{\Lambda}}\boldsymbol{D}_A\boldsymbol{U}_S^{\dagger}\boldsymbol{S}_1^{1/2}\right)^{-1}\boldsymbol{v}\right|^2}{\boldsymbol{v}^{\dagger}\left(\boldsymbol{S}_1^{1/2}\boldsymbol{U}_S\boldsymbol{D}_A\widehat{\boldsymbol{\Lambda}}\boldsymbol{D}_A\boldsymbol{U}_S^{\dagger}\boldsymbol{S}_1^{1/2}\right)^{-1}\boldsymbol{v}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{18}$$

The above architecture is referred to in the following as Double Trained AMF (DT-AMF). Note that the estimate provided by *Theorem 3.1* is invertible irrespective of the rank of $\boldsymbol{S}_2$ (recall also that $K \geq N$ and, hence, $\boldsymbol{S}_1$ is full-rank with probability 1) since $\widehat{\boldsymbol{\Lambda}}$ and $\boldsymbol{D}_A$ are always nonsingular even when $M < N$. It follows that this architecture can be used in sample starved scenarios.

As final remark, it is of interest to discuss the behavior of the DT-AMF under two different operating conditions. More precisely, let us firstly assume that K and $M \geq N$ are such that $\lambda_{S,i} > \frac{M}{K}$, $i=1,\ldots,N$, then (18) becomes statistically equivalent to the AMF (see equation (6)). Note that the condition on $\lambda_{S,i}$ implies that the additional dataset is ignored and, hence, $\boldsymbol{S}_1$ is not used in the estimate. On the opposite side, in the second case, assume that K and $M \geq N$ are such that $\lambda_{S,i} \leq \frac{M}{K}$, $i=1,\ldots,N$, then

$$\boldsymbol{D}_A\widehat{\boldsymbol{\Lambda}}\boldsymbol{D}_A = \frac{1}{K+M}(\boldsymbol{I}+\boldsymbol{\Lambda}_S) \tag{19}$$

and the additional training set is used to load $\boldsymbol{S}_2$, namely

$$\widehat{\boldsymbol{M}}_2 = \frac{1}{K+M}(\boldsymbol{S}_1+\boldsymbol{S}_2), \tag{20}$$

which is the SCM based over the two sets of secondary data. Of course depending on the size of the two training datasets and the values of the $\lambda_{S,i}$s, intermediate situations can be experienced and the estimator automatically accounts for them.

B. Rao Test

As first step, define $\boldsymbol{S}_3 = \boldsymbol{S}_2 + \boldsymbol{z}\boldsymbol{z}^{\dagger}$, $M_1 = M+1$, $a_1 = 2(K+M+1)$, and the vector of parameters of interest $\boldsymbol{\theta}_r = [\alpha_r\ \alpha_i]^T$. Moreover, observe that $\boldsymbol{M}_2$ contains N_s nuisance parameters which can be grouped in a given univocal way into a vector $\boldsymbol{\theta}_s \in \mathbb{R}^{N_s\times 1}$. Now, partition the Fisher Information Matrix as follows

$$\boldsymbol{J}(\boldsymbol{\theta}) = \begin{bmatrix}\boldsymbol{J}_{rr}(\boldsymbol{\theta}) & \boldsymbol{J}_{rs}(\boldsymbol{\theta})\\ \boldsymbol{J}_{sr}(\boldsymbol{\theta}) & \boldsymbol{J}_{ss}(\boldsymbol{\theta})\end{bmatrix}, \tag{21}$$

where $\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\theta}_r^T\ \boldsymbol{\theta}_s^T\end{bmatrix}^T$ and $\boldsymbol{J}_{rr}(\boldsymbol{\theta}) \in \mathbb{R}^{2\times 2}$. The Rao test for problem (7) is given by [24]

$$\begin{aligned}&\left[\frac{\partial}{\partial\boldsymbol{\theta}_r}\log f_1\left(\boldsymbol{z},\boldsymbol{Z}_1,\boldsymbol{Z}_2;\alpha,\boldsymbol{M}_1,\boldsymbol{R},H_1\right)\right]^T_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_0}\\ &\qquad\times\left[\boldsymbol{J}^{-1}\left(\widehat{\boldsymbol{\theta}}_0\right)\right]_{rr}\\ &\times\left[\frac{\partial}{\partial\boldsymbol{\theta}_r}\log f_1\left(\boldsymbol{z},\boldsymbol{Z}_1,\alpha,\boldsymbol{Z}_2;\boldsymbol{M}_1,\boldsymbol{R},H_1\right)\right]_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}_0} \underset{H_0}{\overset{H_1}{\gtrless}} \eta,\end{aligned} \tag{22}$$

where $\widehat{\boldsymbol{\theta}}_0 = [\boldsymbol{0}^T\ \widehat{\boldsymbol{\theta}}_{s,0}^T]^T$ is the MLE of $\boldsymbol{\theta}$ under H_0 and

$$\left[\boldsymbol{J}^{-1}(\boldsymbol{\theta})\right]_{rr} = \left[\boldsymbol{J}_{rr}(\boldsymbol{\theta}) - \boldsymbol{J}_{rs}(\boldsymbol{\theta})\boldsymbol{J}_{ss}^{-1}(\boldsymbol{\theta})\boldsymbol{J}_{sr}(\boldsymbol{\theta})\right]^{-1}. \tag{23}$$

As for the MLE of $\boldsymbol{\theta}_s$ under H_0 denoted by $\widehat{\boldsymbol{\theta}}_{s,0}$, it can be obtained exploiting the following theorem, which provides the MLE of $\boldsymbol{M}_2$ under H_0.

Theorem 3.2: The MLE of $\boldsymbol{M}_1+\boldsymbol{R}$ under the null hypothesis of (7) is

$$\widehat{\boldsymbol{M}}_2 = \boldsymbol{S}_1^{1/2}\boldsymbol{V}_S\boldsymbol{C}_A\widehat{\boldsymbol{\Phi}}\boldsymbol{C}_A\boldsymbol{V}_S^{\dagger}\boldsymbol{S}_1^{1/2}, \tag{24}$$

where $\boldsymbol{S}_1$ has been already defined, $\gamma_{S,1} \geq \ldots \geq \gamma_{S,N} \geq 0$ are the eigenvalues of $\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_3\boldsymbol{S}_1^{-1/2}$ with $\boldsymbol{V}_S \in \mathbb{C}^{N\times N}$ a unitary matrix containing the corresponding eigenvectors,

$$\widehat{\boldsymbol{\Phi}} = \textbf{diag}\left(\frac{K}{M_1}\gamma_{S,1},\ldots,\frac{K}{M_1}\gamma_{S,\hat{r}_1},1,\ldots,1\right), \tag{25}$$

and

$$\begin{aligned}\boldsymbol{C}_A = \textbf{diag}\Bigg(&\frac{1}{\sqrt{K}},\ldots,\frac{1}{\sqrt{K}},\sqrt{\frac{\gamma_{S,\hat{r}_1+1}+1}{a_1/2}},\\ &\ldots,\sqrt{\frac{\gamma_{S,N}+1}{a_1/2}}\Bigg)\end{aligned} \tag{26}$$

with $\hat{r}_1$ the number of $\gamma_{S,i}$ greater than M_1/K.

Proof: Exploit Theorem 3.1 with M_1 and $\boldsymbol{S}_3$ in place of M and $\boldsymbol{S}_2$, respectively. ∎

Now, it is not difficult to show that [30]

$$\frac{\partial}{\partial \boldsymbol{\theta}_r} \log f\left(\boldsymbol{z}, \boldsymbol{Z}_1, \boldsymbol{Z}_2; \boldsymbol{M}_1, \boldsymbol{R}, \alpha, H_1\right) = \begin{bmatrix} 2\Re\left\{\boldsymbol{v}^\dagger (\boldsymbol{M}_1+\boldsymbol{R})^{-1}(\boldsymbol{z}-\alpha\boldsymbol{v})\right\} \\ 2\Im\left\{\boldsymbol{v}^\dagger (\boldsymbol{M}_1+\boldsymbol{R})^{-1}(\boldsymbol{z}-\alpha\boldsymbol{v})\right\} \end{bmatrix} \quad (27)$$

while $\left[\boldsymbol{J}^{-1}(\boldsymbol{\theta})\right]_{rr} = \left[2\boldsymbol{v}^\dagger(\boldsymbol{M}_1+\boldsymbol{R})^{-1}\boldsymbol{v}\right]^{-1}\boldsymbol{I}$. Gathering the above results, we obtain the following detector

$$\frac{\left|\boldsymbol{z}^\dagger \left(\boldsymbol{S}_1^{1/2}\boldsymbol{V}_S\boldsymbol{C}_A\widehat{\boldsymbol{\Phi}}\boldsymbol{C}_A\boldsymbol{V}_S^\dagger\boldsymbol{S}_1^{1/2}\right)^{-1}\boldsymbol{v}\right|^2}{\boldsymbol{v}^\dagger \left(\boldsymbol{S}_1^{1/2}\boldsymbol{V}_S\boldsymbol{C}_A\widehat{\boldsymbol{\Phi}}\boldsymbol{C}_A\boldsymbol{V}_S^\dagger\boldsymbol{S}_1^{1/2}\right)^{-1}\boldsymbol{v}} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad (28)$$

referred to as Double Trained Rao Detector (DT-RAOD). Finally, we point out that the same final considerations as for the DT-AMF apply to the DT-RAOD replacing $\boldsymbol{S}_2$ with $\boldsymbol{S}_3$ and M with M_1.

Before illustrating the performance analysis, we note that the proposed architectures are more time consuming than the conventional counterparts. In fact, resorting to the usual Landau notation, the AMF and RAOD [30] involve the evaluation and the inversion of $\boldsymbol{S}_2$ which are $\mathcal{O}(MN^2)$ and $\mathcal{O}(N^3)$, respectively. On the other hand, the DT-AMF and DT-RAOD need

- the computation of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ (or $\boldsymbol{S}_3$) which is $\mathcal{O}((K+M)N^2)$ (or $\mathcal{O}((K+M+1)N^2)$);
- the inversion of $\boldsymbol{S}_1$ which is $\mathcal{O}(N^3)$;
- the computation of $\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$ which is $\mathcal{O}(N^3)$;
- the eigendecomposition of $\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$ which is $\mathcal{O}(N^3)$. controllato su golub per SVD e' O(n3), pag 254

IV. Numerical Examples and Discussion

This section presents the performance analysis of the proposed architectures in comparison with the AMF [2], the RAOD [30], the Matched Filter (MF) given by (15), and the AMF evaluated using transformed data as proposed in [31]. Specifically, the latter architecture is constructed as follows. Whiten $\boldsymbol{z}$, $\boldsymbol{Z}_2$, and $\boldsymbol{v}$ using $\boldsymbol{S}_1$, namely

$$\tilde{\boldsymbol{z}} = \boldsymbol{S}_1^{-1/2}\boldsymbol{z},\ \tilde{\boldsymbol{Z}}_2 = \boldsymbol{S}_1^{-1/2}\boldsymbol{Z}_2,\ \tilde{\boldsymbol{v}} = \boldsymbol{S}_1^{-1/2}\boldsymbol{v}. \quad (29)$$

and let $r < N-1$ the number of eigenvalues of the actual clutter covariance matrix greater than the noise power level. Then, consider the first r singular vectors, $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r$, of $\tilde{\boldsymbol{Z}}_2$ and form matrix $\bar{\boldsymbol{B}} = [\boldsymbol{v}, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_r] \in \mathbb{C}^{N\times L}$, where $L = r+1 < N$. Finally, transform data through $\bar{\boldsymbol{B}}$ as $\bar{\boldsymbol{z}} = \bar{\boldsymbol{B}}^\dagger\tilde{\boldsymbol{z}}$, $\bar{\boldsymbol{Z}}_2 = \bar{\boldsymbol{B}}^\dagger\tilde{\boldsymbol{Z}}_2$, $\bar{\boldsymbol{v}} = \bar{\boldsymbol{B}}^\dagger\tilde{\boldsymbol{v}}$, and compute the AMF using the transformed data. The resulting receiver is referred to in the following as Reduced Dimensions AMF (RD-AMF). A few remarks are now in order. First, observe that the RD-AMF is clairvoyant since it requires the knowledge of the actual clutter covariance matrix and of the noise power level. Second, the RD-AMF also requires the constraint $M \geq L$ to ensure that $\bar{\boldsymbol{S}}_2 = \bar{\boldsymbol{Z}}\bar{\boldsymbol{Z}}^\dagger$ is invertible with probability 1.

The metric used to assess the performances is the P_D estimated by means of standard Monte Carlo counting techniques over 1000 independent trials. The thresholds are set in order to guarantee $P_{fa} = 10^{-4}$ and are evaluated using $100/P_{fa}$ independent trials. The following study is structured to provide a picture as complete as possible of the parameters which influence the detection performances.

The interference is modeled as circular complex normal random vectors with zero mean and covariance matrix

$$\boldsymbol{M}_1 = \boldsymbol{I} + \sum_{i=1}^{2} \text{JNR}\ \boldsymbol{v}(\nu_{J,i})\boldsymbol{v}(\nu_{J,i})^\dagger, \quad (30)$$

for the first set of training data, and

$$\boldsymbol{M}_2 = \boldsymbol{M}_1 + \text{CNR}\ \boldsymbol{M}_c, \quad (31)$$

for the second set of training data. In (30), $\boldsymbol{I}$ represents the thermal noise component and the last addendum is due to a dual-tone continuous-wave jammer with $\nu_{J,1} = 0.3$ and $\nu_{J,2} = 0.45$; in (31), $\boldsymbol{M}_c$ accounts for the clutter contribution with $\boldsymbol{M}_c(m,n) = \rho^{|m-n|}$ and $\rho = 0.95$. Finally, the SINR is defined as $\text{SINR} = |\alpha|^2\boldsymbol{v}^\dagger\boldsymbol{M}_2^{-1}\boldsymbol{v}$ with $\boldsymbol{v} = [1\ \ldots\ 1]^T$.

The analysis presented starts from the evaluation of the sensitivity to the CNR and the JNR using low numbers of training samples. Then, the performance are investigated maintaining the CNR and the JNR constant and varying the cardinality of the secondary data sets or the vector size. Finally, sample starved scenarios are considered.

Figures 3-5 consider the effects of both the CNR and the JNR on the performance assuming $N = 8$ and $M = 9 = K$. More precisely, Figure 3 considers CNR$=$ 20 dB and JNR$=$ 30 dB, then the CNR value increases by 10 dB in Figure 4 and, finally, in Figure 5 the JNR is set to 40 dB. From inspection of these figures, it turns out that the Rao-based architectures exhibit very poor performance in the presence of small sets of secondary data due to the inclusion of the CUT in the computation of the SCM, while the DT-AMF guarantees the best detection performance with a maximum loss of about 7 dB at $P_D = 0.9$ with respect to the MF. Another aspect is that the newly proposed architectures suffer performance degradation as the CNR and/or JNR increase. Finally, for the simulated covariance structure, observe that r increases for high values of the CNR and, hence, when $L = r+1 = N$ the RD-AMF cannot be constructed. For this reason, it is not reported in Figures 4 and 5.

The curves reported in Figure 6 refer to the same parameters setting as in Figure 3 but for $K = 16$. The DT-AMF confirms its superiority with respect to the considered competitors (for the considered values of the parameters) followed by the RD-AMF whose loss with respect to the DT-AMF is about 5 dB at $P_d = 0.9$. Note that the performance of the DT-AMF improves of about 1 dB at $P_D = 0.9$ with respect to the analogous curve of Figure 3. On the opposite side, the DT-RAOD and the RAOD continue to exhibit poor performances. In Figures 7 and 8, the performance is investigated when a small number of training data is available to estimate $\boldsymbol{M}_2$. Specifically, Figure 7 considers the case $L < M < N$ where the AMF and the RAOD cannot be used. Inspection of the figure highlights a significant gain (greater than 10 dB) of the DT-AMF over the RD-AMF at $P_D = 0.9$. In Figure 8

the constraint on M becomes more restricting, since $M < L$. Under this assumption, the RD-AMF is not well-defined and, hence, its performance curve is not reported while the DT-AMF is still capable of achieving P_D values greater than 0.9 for the considered parameter setting.

Figure 9 shows the influence of N on the performances. As expected, increasing N (and consequently M and K) leads to an enhancement of the detection performance for all the architectures. The ranking observed in Figure 3 is repeated with the top step of the podium occupied by the DT-AMF, which exhibits a gain of about 1 dB at $P_d = 0.9$ over the RD-AMF. Finally, in Figure 10, we analyze the behaviors of the considered detection rules when the so-called RMB rule is satisfied for the estimation of $\boldsymbol{M}_2$ [32], namely $M = 2N$. More precisely, when the RMB rule holds, the performances are compressed and all curves belong to an interval of about 3 dB.

Summarizing, P_D curves have singled out the DT-AMF as the recommended detector in the presence of a double training data set at least for the considered scenarios and parameter setting and for matched signals. As for the DT-RAOD, it is completely unuseful when a small number of training data is available. As a matter of fact, the analysis in the presence of mismatched signals represents a future line of research allowing to observe the behavior of the proposed architectures also in terms of rejection of unwanted signals [33].

V. Conclusions

In this paper, we have addressed a new radar detection problem, which exploits two sets of training samples. The first set is acquired before transmitting the pulse waveform (spatial processing only) or at high ranges (pulse trains with high PRI), while the second data set is formed using the conventional reference window centered at the CUT. Then, the problem has been formulated as a binary hypothesis test and is solved using the two-step GLRT and the Rao test design procedures. Remarkably, closed-form expression for the MLE estimate of the ICM is provided. This result represents the main technical contribution of the paper. The performance analysis has highlighted that the advantage of using two training data sets is twofold. On one side, data from the first training set are less exposed to inhomogeneity and provide reliable estimates of some ICM components, which lead to increased detection performances. On the other side, decision schemes exploiting both training sets are capable of operating in sample starved conditions. The numerical examples presented here have singled out the DT-AMF as the recommended decision scheme at least for the considered scenarios.

Future research tracks might concern the derivation of the plain GLRT, the analysis of two-stage architectures where the first stage suitably selects the training data [34] while the second stage is responsible for the detection, or the exploitation of two training sets when the interference power level of the second data set is different from that in the CUT.

Appendix
Proof of Theorem 3.1

Let us evaluate the ML estimate of $(\boldsymbol{M}_1 + \boldsymbol{R})$ optimizing the joint likelihood of $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ with respect to $\boldsymbol{M}_1$ and $\boldsymbol{R}$, namely

$$\max_{\boldsymbol{M}_1} \max_{\boldsymbol{R}} f\left(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \boldsymbol{M}_1, \boldsymbol{R}\right). \tag{32}$$

As first step of the derivation, consider the logarithm of the joint pdf

$$\begin{aligned} &\log f\left(\boldsymbol{Z}_1, \boldsymbol{Z}_2; \boldsymbol{M}_1, \boldsymbol{R}\right) \approx -K \log\left[\det\left(\boldsymbol{M}_1\right)\right] \\ &- M \log\left[\det\left(\boldsymbol{M}_1 + \boldsymbol{R}\right)\right] - \mathrm{Tr}\left[\boldsymbol{M}_1^{-1} \boldsymbol{S}_1\right] \\ &- \mathrm{Tr}\left[\left(\boldsymbol{M}_1 + \boldsymbol{R}\right)^{-1} \boldsymbol{S}_2\right] = h(\boldsymbol{M}_1, \boldsymbol{R}; \boldsymbol{Z}_1, \boldsymbol{Z}_2). \end{aligned} \tag{33}$$

Now, denote by $\lambda_1 \geq \ldots \geq \lambda_N \geq 1$ the eigenvalues of $\boldsymbol{M}_1^{-1/2}\left(\boldsymbol{M}_1 + \boldsymbol{R}\right)\boldsymbol{M}_1^{-1/2}$, whose eigendecomposition is

$$\boldsymbol{M}_1^{-1/2}\left(\boldsymbol{M}_1 + \boldsymbol{R}\right)\boldsymbol{M}_1^{-1/2} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\dagger}, \tag{34}$$

where $\boldsymbol{U} \in \mathbb{C}^{N \times N}$ such that[6] $\boldsymbol{U}^{\dagger}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^{\dagger} = \boldsymbol{I}$, and define $\boldsymbol{A} = \boldsymbol{M}_1^{1/2}\boldsymbol{U}$. It is straightforward to show that

$$\boldsymbol{A}\boldsymbol{A}^{\dagger} = \boldsymbol{M}_1, \quad \boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}^{\dagger} = \boldsymbol{M}_1 + \boldsymbol{R}. \tag{35}$$

Observe the set of eigenvalues $\{\lambda_1, \ldots, \lambda_N\}$ can be partitioned as $\{\Omega_1,\ \Omega_2\}$, where Ω_1 is the subset containing the eigenvalues strictly greater than 1 while Ω_2 is the subset whose elements are equal to 1. Moreover, the cardinality of Ω_1 is the unknown rank of $\boldsymbol{R}$.

Using the last equalities, it is possible to recast the log likelihood as

$$\begin{aligned} &h(\boldsymbol{M}_1, \boldsymbol{R}; \boldsymbol{Z}_1, \boldsymbol{Z}_2) = h_1(\boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2) \\ &= -2K \log\det\left(\boldsymbol{A}\right) - 2M \log\det\left(\boldsymbol{A}\right) - M \log\det\left(\boldsymbol{\Lambda}\right) \\ &- \mathrm{Tr}\left(\left(\boldsymbol{A}\boldsymbol{A}^{\dagger}\right)^{-1}\boldsymbol{S}_1\right) - \mathrm{Tr}\left(\left(\boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}^{\dagger}\right)^{-1}\boldsymbol{S}_2\right) \\ &= -\left(2K + 2M\right)\log\det\left(\boldsymbol{A}\right) - M\log\det\left(\boldsymbol{\Lambda}\right) \\ &- \mathrm{Tr}\left(\left(\boldsymbol{A}^{\dagger}\right)^{-1}\boldsymbol{A}^{-1}\boldsymbol{S}_1\right) - \mathrm{Tr}\left(\left(\boldsymbol{A}^{\dagger}\right)^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{A}^{-1}\boldsymbol{S}_2\right). \end{aligned} \tag{36}$$

Now, let us set $\bar{\epsilon}$ such that $0 < \bar{\epsilon} < M/K$ and $r(\bar{\epsilon}) = \hat{r}$, where $r(\bar{\epsilon})$ is the number of eigenvalues of $\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$, $\lambda_{S,i}$ say, greater than $\frac{M}{K} - \bar{\epsilon}$ and $\hat{r}$ is the number of $\lambda_{S,i}$, $i = 1, \ldots, N$, greater than M/K. Moreover, let us introduce the auxiliary variable $0 < \epsilon < \bar{\epsilon}$, then we can define $\boldsymbol{S}(\epsilon) = \boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2} + \epsilon\boldsymbol{I}$, whose eigendecomposition is $\boldsymbol{S}(\epsilon) = \boldsymbol{U}_S\boldsymbol{\Lambda}_S(\epsilon)\boldsymbol{U}_S^{\dagger}$, where $\boldsymbol{\Lambda}_S(\epsilon) = \boldsymbol{\Lambda}_S + \epsilon\boldsymbol{I}$, $\boldsymbol{\Lambda}_S \in \mathbb{R}^{N \times N}$ is diagonal and contains the eigenvalues of $\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$, and $\boldsymbol{U}_S \in \mathbb{C}^{N \times N}$ is a unitary matrix of the corresponding eigenvectors of $\boldsymbol{S}_1^{-1/2}\boldsymbol{S}_2\boldsymbol{S}_1^{-1/2}$. Two remarks are now in order. First, observe that $\boldsymbol{U}_S$ does not depend on ϵ; second, $\boldsymbol{S}(\epsilon)$ is full-rank irrespective of the rank of $\boldsymbol{S}_2$ which is controlled by the value of M.

[6]Given $\boldsymbol{X} \succeq 0$ and its eigenvalue decomposition $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\dagger}$, without loss of generality we assume that $\det(\boldsymbol{U}) = 1$.

The next step consists in introducing the following function

$$\begin{aligned}&\bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2)\\&= -(2K+2M)\log\det(\boldsymbol{A}) - M\log\det(\boldsymbol{\Lambda})\\&- \mathrm{Tr}\left(\left(\boldsymbol{A}^\dagger\right)^{-1}\boldsymbol{A}^{-1}\boldsymbol{S}_1\right) - \mathrm{Tr}\left(\left(\boldsymbol{A}^\dagger\right)^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{A}^{-1}\boldsymbol{S}_2(\epsilon)\right),\end{aligned} \tag{37}$$

where $\boldsymbol{S}_2(\epsilon) = \boldsymbol{S}_2 + \epsilon\boldsymbol{S}_1$. It is important to observe that

$$\lim_{\epsilon\to 0^+}\bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2) = h_1(\boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2). \tag{38}$$

Now, recast $\boldsymbol{S}_1$ and $\boldsymbol{S}_2(\epsilon)$ as functions of $\boldsymbol{U}_S$ and $\boldsymbol{\Lambda}_S(\epsilon)$. To this end, denote by $\boldsymbol{B} = \boldsymbol{S}_1^{1/2}\boldsymbol{U}_S$, then $\boldsymbol{S}_1$ and $\boldsymbol{S}_2(\epsilon)$ can be decomposed as $\boldsymbol{S}_1 = \boldsymbol{B}\boldsymbol{B}^\dagger$ and $\boldsymbol{S}_2(\epsilon) = \boldsymbol{B}\boldsymbol{\Lambda}_S(\epsilon)\boldsymbol{B}^\dagger$. It follows that

$$\begin{aligned}&\bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2)\\&= -(2K+2M)\log\det(\boldsymbol{A}) - M\log\det(\boldsymbol{\Lambda})\\&- \mathrm{Tr}\left[\left(\boldsymbol{A}^\dagger\right)^{-1}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{B}^\dagger\right]\\&- \mathrm{Tr}\left[\left(\boldsymbol{A}^\dagger\right)^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{\Lambda}_S(\epsilon)\boldsymbol{B}^\dagger\right]\\&= -(2K+2M)\log\det(\boldsymbol{A}) - M\log\det(\boldsymbol{\Lambda})\\&- \mathrm{Tr}\left[\left(\boldsymbol{A}^{-1}\boldsymbol{B}\right)^\dagger\boldsymbol{\Lambda}_S^{1/2}(\epsilon)\boldsymbol{\Lambda}_S^{-1}(\epsilon)\boldsymbol{\Lambda}_S^{1/2}(\epsilon)\left(\boldsymbol{A}^{-1}\boldsymbol{B}\right)\right]\\&- \mathrm{Tr}\left[\boldsymbol{\Lambda}^{-1}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{\Lambda}_S^{1/2}(\epsilon)\boldsymbol{\Lambda}_S^{1/2}(\epsilon)\left(\boldsymbol{A}^{-1}\boldsymbol{B}\right)^\dagger\right]\\&= -(2K+2M)\log\det(\boldsymbol{A}) - M\log\det(\boldsymbol{\Lambda})\\&- \mathrm{Tr}\left[\left(\boldsymbol{\Lambda}_S^{1/2}(\epsilon)\boldsymbol{A}^{-1}\boldsymbol{B}\right)^\dagger\boldsymbol{\Lambda}_S^{-1}(\epsilon)\left(\boldsymbol{\Lambda}_S^{1/2}(\epsilon)\boldsymbol{A}^{-1}\boldsymbol{B}\right)\right]\\&- \mathrm{Tr}\left[\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{\Lambda}_S^{1/2}(\epsilon)\right)\left(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{\Lambda}_S^{1/2}(\epsilon)\right)^\dagger\right].\end{aligned} \tag{39}$$

Now, note that the optimization of $\bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2)$ with respect to $\boldsymbol{M}_1$ and $\boldsymbol{R}$ is tantamount to

$$\max_{\boldsymbol{\Lambda}}\max_{\boldsymbol{A}}\bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2). \tag{40}$$

In order solve the above problem, let us exploit the singular value decomposition of

$$\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{\Lambda}_S^{1/2}(\epsilon) = \boldsymbol{P}_\epsilon\boldsymbol{D}_\epsilon\boldsymbol{Q}_\epsilon \tag{41}$$

where $\boldsymbol{P}_\epsilon, \boldsymbol{Q}_\epsilon \in \mathbb{C}^{N\times N}$ are unitary and $\det(\boldsymbol{P}_\epsilon\boldsymbol{Q}_\epsilon) = 1$, while $\boldsymbol{D}_\epsilon \in \mathbb{R}^{N\times N}$ is diagonal. Then, replacing (41) into (39) leads to

$$\begin{aligned}&\bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2)\\&= (2K+2M)\log\det\left(\boldsymbol{P}_\epsilon\boldsymbol{D}_\epsilon\boldsymbol{Q}_\epsilon\boldsymbol{\Lambda}_S(\epsilon)^{-1/2}\boldsymbol{B}^{-1}\right)\\&- M\log\det(\boldsymbol{\Lambda})\\&- \mathrm{Tr}\left[\left(\boldsymbol{\Lambda}_S^{1/2}(\epsilon)\boldsymbol{P}_\epsilon\boldsymbol{D}_\epsilon\boldsymbol{Q}_\epsilon\boldsymbol{\Lambda}_S^{-1/2}(\epsilon)\right)^\dagger\boldsymbol{\Lambda}_S^{-1}(\epsilon)\right.\\&\left.\times\left(\boldsymbol{\Lambda}_S^{1/2}(\epsilon)\boldsymbol{P}_\epsilon\boldsymbol{D}_\epsilon\boldsymbol{Q}_\epsilon\boldsymbol{\Lambda}_S^{-1/2}(\epsilon)\right)\right]\\&- \mathrm{Tr}\left[\boldsymbol{\Lambda}^{-1}\boldsymbol{P}_\epsilon\boldsymbol{D}_\epsilon\boldsymbol{Q}_\epsilon\left(\boldsymbol{P}_\epsilon\boldsymbol{D}_\epsilon\boldsymbol{Q}_\epsilon\right)^\dagger\right]\\&= (2K+2M)\left\{\log\det(\boldsymbol{D}_\epsilon) - \frac{1}{2}\log\det(\boldsymbol{\Lambda}_S(\epsilon))\right.\\&\left.- \log\det(\boldsymbol{B})\right\} - M\log\det(\boldsymbol{\Lambda})\\&- \mathrm{Tr}\left[\boldsymbol{\Lambda}_S^{-1}(\epsilon)\boldsymbol{Q}_\epsilon^\dagger\boldsymbol{D}_\epsilon^2\boldsymbol{Q}_\epsilon\right] - \mathrm{Tr}\left[\boldsymbol{\Lambda}^{-1}\boldsymbol{P}_\epsilon\boldsymbol{D}_\epsilon^2\boldsymbol{P}_\epsilon^\dagger\right]\\&= -(K+M)\log\det(\boldsymbol{\Lambda}_S(\epsilon)) - (2K+2M)\log\det(\boldsymbol{B})\\&+ (2K+2M)\log\det(\boldsymbol{D}_\epsilon) - M\log\det(\boldsymbol{\Lambda})\\&- \mathrm{Tr}\left[\boldsymbol{D}_\epsilon^2\boldsymbol{Q}_\epsilon\boldsymbol{\Lambda}_S^{-1}(\epsilon)\boldsymbol{Q}_\epsilon^\dagger\right] - \mathrm{Tr}\left[\boldsymbol{D}_\epsilon^2\boldsymbol{P}_\epsilon^\dagger\boldsymbol{\Lambda}^{-1}\boldsymbol{P}_\epsilon\right]\\&= \bar{h}_2(\epsilon, \boldsymbol{D}_\epsilon, \boldsymbol{P}_\epsilon, \boldsymbol{Q}_\epsilon, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2).\end{aligned} \tag{42}$$

It follows that problem (40) can be recast as

$$\max_{\boldsymbol{D}_\epsilon}\max_{\boldsymbol{P}_\epsilon}\max_{\boldsymbol{Q}_\epsilon}\max_{\boldsymbol{\Lambda}}\bar{h}_2(\epsilon, \boldsymbol{D}_\epsilon, \boldsymbol{P}_\epsilon, \boldsymbol{Q}_\epsilon, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2). \tag{43}$$

Maximization with respect to $\boldsymbol{P}_\epsilon$ can be accomplished noticing that the nonzero entries of $\boldsymbol{D}_\epsilon^2$ are in descending order while the nonzero entries of $\boldsymbol{\Lambda}^{-1}$ appear in ascending order. Let $\boldsymbol{X}_P = \boldsymbol{P}_\epsilon^\dagger\boldsymbol{\Lambda}^{-1}\boldsymbol{P}_\epsilon$, then exploiting *Theorem 1* of [35] the following inequality holds

$$-\mathrm{Tr}\left[\boldsymbol{D}_\epsilon^2\boldsymbol{X}_P\right] \le -\sum_{i=1}^N d_{\epsilon,i}^2\lambda_i^{-1}, \tag{44}$$

where $d_{\epsilon,i}^2$, $i = 1, \ldots, N$, are the main diagonal entries of $\boldsymbol{D}_\epsilon^2$. Under the constraint that $\boldsymbol{P}_\epsilon$ is unitary, the equality can be satisfied if $\widehat{\boldsymbol{P}}_\epsilon = e^{j\phi_P}\boldsymbol{I}$ for arbitrary $\phi_P \in [0,\ 2\pi]$. Following the same line of reasoning, it is possible to show that optimization with respect to $\boldsymbol{Q}_\epsilon$ leads to $\widehat{\boldsymbol{Q}}_\epsilon = e^{j\phi_Q}\boldsymbol{I}$ for arbitrary $\phi_Q \in [0,\ 2\pi]$. Thus, choosing for simplicity $\phi_P = \phi_Q = 0$, the following inequality holds

$$\begin{aligned}&\bar{h}_2(\epsilon, \boldsymbol{D}_\epsilon, \boldsymbol{I}, \boldsymbol{I}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2) \ge \bar{h}_2(\epsilon, \boldsymbol{D}_\epsilon, \boldsymbol{P}_\epsilon, \boldsymbol{Q}_\epsilon, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2)\\&= \bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2).\end{aligned} \tag{45}$$

It still remains to perform the maximization with respect to $\boldsymbol{D}_\epsilon$ and $\boldsymbol{\Lambda}$. The terms involved in this optimization are

$$\begin{aligned}&\underbrace{(2K+2M)}_{a}\log\det(\boldsymbol{D}_\epsilon) - M\log\det(\boldsymbol{\Lambda})\\&- \mathrm{Tr}\left[\boldsymbol{D}_\epsilon^2\boldsymbol{\Lambda}_S^{-1}(\epsilon)\right] - \mathrm{Tr}\left[\boldsymbol{D}_\epsilon^2\boldsymbol{\Lambda}^{-1}\right]\\&= \frac{a}{2}\sum_{i=1}^N\log d_{\epsilon,i}^2 - M\sum_{i=1}^N\log\lambda_i - \sum_{i=1}^N\frac{d_{\epsilon,i}^2}{\lambda_{S,i}+\epsilon} - \sum_{i=1}^N\frac{d_{\epsilon,i}^2}{\lambda_i}\\&= g_\epsilon\left(d_{\epsilon,1}, \ldots, d_{\epsilon,N}, \lambda_1, \ldots, \lambda_N\right) = g_\epsilon(\boldsymbol{p}),\end{aligned} \tag{46}$$

where $\boldsymbol{p} = [d_{\epsilon,1}, \ldots, d_{\epsilon,N}, \lambda_1, \ldots, \lambda_N]^T$ and $\lambda_{S,i}$, $i = 1, \ldots, N$, are the nonzero elements of $\boldsymbol{\Lambda}_S$. Therefore, the stationary points of $g(\cdot, \ldots, \cdot, \ldots, \cdot)$ can be obtained setting to zero its gradient, namely $\frac{\partial}{\partial \boldsymbol{p}}[g_\epsilon(\boldsymbol{p})] = \boldsymbol{0}$. To solve the above equation, we exploit the linearity of the derivative operator and focus on the pair $(\lambda_i, d^2_{\epsilon,i})$. More precisely, equating to zero the first derivative of $g_\epsilon(\boldsymbol{p})$ with respect to λ_i and $d^2_{\epsilon,i}$ leads to the following system of equations

$$\begin{cases} -\dfrac{M}{\lambda_i} + \dfrac{d^2_{\epsilon,i}}{\lambda_i^2} = 0, \\ \dfrac{a}{2}\dfrac{1}{d^2_{\epsilon,i}} - \left(\dfrac{1}{\lambda_{S,i}+\epsilon} + \dfrac{1}{\lambda_i}\right) = 0. \end{cases} \tag{47}$$

It follows that optimizers of (46) are

$$\widehat{d}^2_{\epsilon,i} = \begin{cases} K(\lambda_{S,i}+\epsilon), & \text{if } \lambda_{S,i} > \frac{M}{K} - \epsilon, \\ \dfrac{a}{2}\dfrac{\lambda_{S,i}+\epsilon}{\lambda_{S,i}+\epsilon+1}, & \text{if } 0 \le \lambda_{S,i} \le \frac{M}{K} - \epsilon, \end{cases} \tag{48}$$

and

$$\widehat{\lambda}_i = \begin{cases} \dfrac{K}{M}(\lambda_{S,i}+\epsilon), & \text{if } \lambda_{S,i} > \frac{M}{K} - \epsilon, \\ 1, & \text{if } 0 \le \lambda_{S,i} \le \frac{M}{K} - \epsilon, \end{cases} \tag{49}$$

The above equations allow to construct the ML estimates $\widehat{\boldsymbol{D}}_\epsilon$, $\widehat{\boldsymbol{\Lambda}}_\epsilon$, and $\widehat{\boldsymbol{A}}_\epsilon$. The expressions of the above quantities for $\epsilon < \bar{\epsilon}$ are given by

$$\widehat{\boldsymbol{D}}_\epsilon = \textbf{diag}\left(\sqrt{K(\lambda_{S,1}+\epsilon)}, \ldots, \sqrt{K(\lambda_{S,\hat{r}}+\epsilon)}, \sqrt{\frac{a}{2}\frac{\lambda_{S,\hat{r}+1}+\epsilon}{\lambda_{S,\hat{r}+1}+\epsilon+1}}, \ldots, \sqrt{\frac{a}{2}\frac{\lambda_{S,N}+\epsilon}{\lambda_{S,N}+\epsilon+1}}, \right), \tag{50}$$

$$\widehat{\boldsymbol{\Lambda}}_\epsilon = \textbf{diag}\left(\frac{K}{M}(\lambda_{S,1}+\epsilon), \ldots, \frac{K}{M}(\lambda_{S,\hat{r}}+\epsilon), 1, \ldots, 1\right), \tag{51}$$

$$\widehat{\boldsymbol{A}}_\epsilon = \boldsymbol{B}\ \textbf{diag}\left(\frac{1}{\sqrt{K}}, \ldots, \frac{1}{\sqrt{K}}, \sqrt{\frac{\lambda_{S,\hat{r}+1}+\epsilon+1}{a/2}}, \ldots, \sqrt{\frac{\lambda_{S,N}+\epsilon+1}{a/2}}\right). \tag{52}$$

Replacing (51) and (52) into (37), it turns out that

$$\bar{h}_1(\epsilon, \widehat{\boldsymbol{A}}_\epsilon, \widehat{\boldsymbol{\Lambda}}_\epsilon; \boldsymbol{Z}_1, \boldsymbol{Z}_2) \ge \bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2), \tag{53}$$

for the considered values of ϵ, and that there exists and is finite the limit of $\bar{h}_1(\epsilon, \widehat{\boldsymbol{A}}_\epsilon, \widehat{\boldsymbol{\Lambda}}_\epsilon; \boldsymbol{Z}_1, \boldsymbol{Z}_2)$ and $\bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2)$ for $\epsilon \to 0^+$. In fact

$$\widehat{\boldsymbol{A}} = \lim_{\epsilon\to 0^+} \widehat{\boldsymbol{A}}_\epsilon = \boldsymbol{B}\ \textbf{diag}\left(\frac{1}{\sqrt{K}}, \ldots, \frac{1}{\sqrt{K}}, \sqrt{\frac{\lambda_{S,\hat{r}+1}+1}{a/2}}, \ldots, \sqrt{\frac{\lambda_{S,N}+1}{a/2}}\right) = \boldsymbol{B}\boldsymbol{D}_A \tag{54}$$

$$\widehat{\boldsymbol{\Lambda}} = \lim_{\epsilon\to 0^+} \widehat{\boldsymbol{\Lambda}}_\epsilon = \textbf{diag}\left(\frac{K}{M}\lambda_{S,1}, \ldots, \frac{K}{M}\lambda_{S,\hat{r}}, 1, \ldots, 1\right), \tag{55}$$

and exploiting the continuity of $\bar{h}_1(\epsilon, \widehat{\boldsymbol{A}}_\epsilon, \widehat{\boldsymbol{\Lambda}}_\epsilon; \boldsymbol{Z}_1, \boldsymbol{Z}_2)$, we can claim that

$$\lim_{\epsilon\to 0^+} \bar{h}_1(\epsilon, \widehat{\boldsymbol{A}}_\epsilon, \widehat{\boldsymbol{\Lambda}}_\epsilon; \boldsymbol{Z}_1, \boldsymbol{Z}_2) = h_1(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{\Lambda}}; \boldsymbol{Z}_1, \boldsymbol{Z}_2). \tag{56}$$

Besides, recall that

$$\lim_{\epsilon\to 0^+} \bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2) = h_1(\boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2). \tag{57}$$

As a consequence, the following inequality holds

$$\begin{aligned} &\lim_{\epsilon\to 0^+} \bar{h}_1(\epsilon, \widehat{\boldsymbol{A}}_\epsilon, \widehat{\boldsymbol{\Lambda}}_\epsilon; \boldsymbol{Z}_1, \boldsymbol{Z}_2) = h_1(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{\Lambda}}; \boldsymbol{Z}_1, \boldsymbol{Z}_2) \\ &\ge \lim_{\epsilon\to 0^+} \bar{h}_1(\epsilon, \boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2) = h_1(\boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2), \end{aligned} \tag{58}$$

which implies that

$$(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{\Lambda}}) = \arg\max_{\boldsymbol{A},\boldsymbol{\Lambda}} h_1(\boldsymbol{A}, \boldsymbol{\Lambda}; \boldsymbol{Z}_1, \boldsymbol{Z}_2). \tag{59}$$

Thus, the ML estimates of $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ are

$$\widehat{\boldsymbol{M}}_1 = \widehat{\boldsymbol{A}}\widehat{\boldsymbol{A}}^\dagger = \boldsymbol{S}_1^{1/2}\boldsymbol{U}_S\boldsymbol{D}_A^2\boldsymbol{U}_S^\dagger\boldsymbol{S}_1^{1/2} \tag{60}$$

and

$$\widehat{\boldsymbol{M}}_2 = \widehat{\boldsymbol{A}}\widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{A}}^\dagger = \boldsymbol{S}_1^{1/2}\boldsymbol{U}_S\boldsymbol{D}_A^2\widehat{\boldsymbol{\Lambda}}\ \boldsymbol{U}_S^\dagger\boldsymbol{S}_1^{1/2}, \tag{61}$$

respectively.

REFERENCES

[1] E. J. Kelly, "An adaptive detection algorithm," *IEEE Transactions on Aerospace and Electronic Systems*, , no. 2, pp. 115–127, 1986.

[2] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, 1992.

[3] A. De Maio and M. S. Greco, Eds., *Modern Radar Detection Theory*, Radar, Sonar and Navigation. Institution of Engineering and Technology, 2015.

[4] F. Neri, *Introduction to Electronic Defense Systems*, Artech House radar library. SciTech Publishing, Incorporated, 2006.

[5] A. Farina, "Eccm techniques," in *Radar Handbook*, M. I. Skolnik, Ed., chapter 24. McGraw-Hill, 2008.

[6] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Basic Principles*, Raleigh, NC, 2010.

[7] A. De Maio, D. Orlando, C. Hao, and G. Foglia, "Adaptive detection of point-like targets in spectrally symmetric interference," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3207–3220, 2016.

[8] G. Pailloux, P. Forster, J. P. Ovarlez, and F. Pascal, "Persymmetric adaptive radar detectors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 2376–2390, 2011.

[9] R. Nitzberg, "Application of maximum likelihood estimation of persymmetric covariance matrices to adaptive processing," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 16, no. 1, pp. 124–127, 1980.

[10] C. Hao, S. Gazor, G. Foglia, B. Liu, and C. Hou, "Persymmetric adaptive detection and range estimation of a small target," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 4, pp. 2590–2604, 2015.

[11] H. Li, J. Li, and P. Stoica, "Persymmetric parametric adaptive matched filter for multichannel adaptive signal detection," *IEEE Transactions on Signal Processing*, vol. 46, no. 7, pp. 1954–1966, 1998.

[12] Y. I. Abramovich, N. K. Spencer, and Gorokhov A. Y., "Modified GLRT and AMF framework for adaptive detectors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 1017–1051, July 2007.

[13] J. Liu, H. Li, and B. Himed, "Persymmetric adaptive target detection with distributed MIMO radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 372–382, 2015.

[14] J. Liu, G. Cui, H. Li, and B. Himed, "On the performance of a persymmetric adaptive matched filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 4, pp. 2605–2614, 2015.

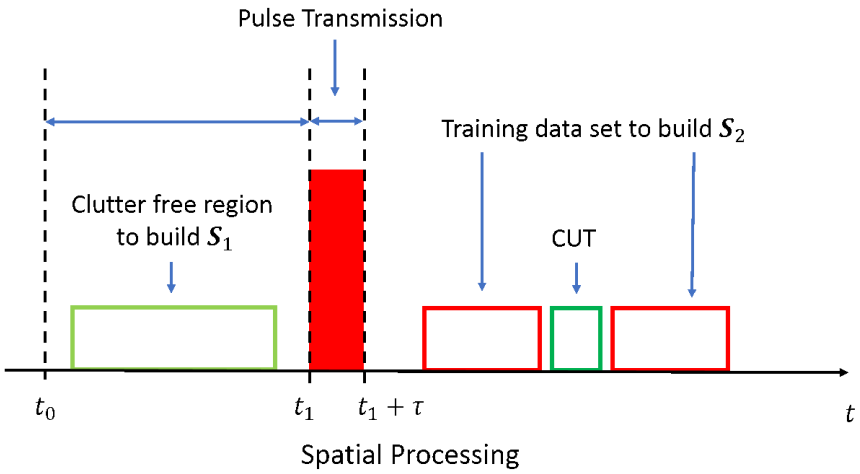

Fig. 1. Acquisition strategy in case of spatial processing.

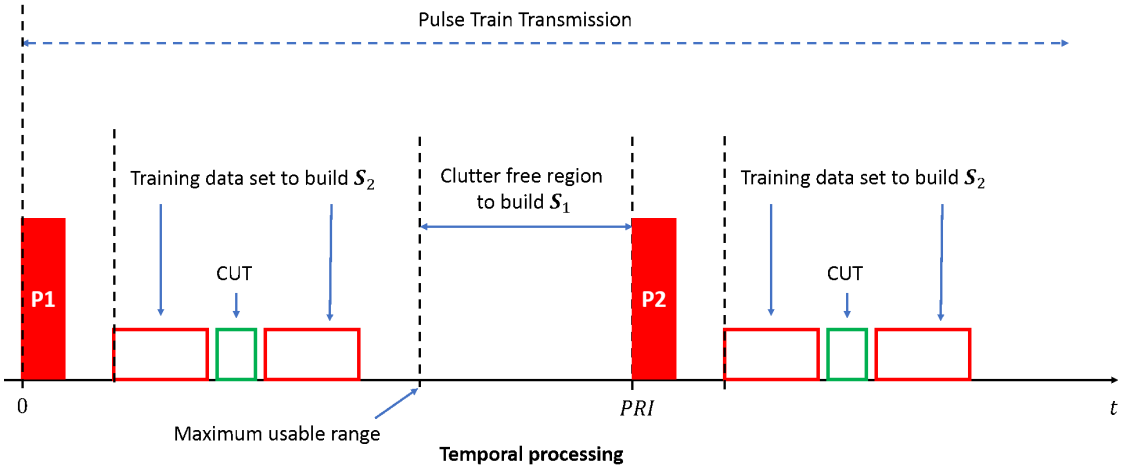

Fig. 2. Acquisition strategy in case of temporal processing.

[15] J. Liu, W. Liu, H. Liu, B. Chen, X. G. Xia, and F. Dai, "Average SINR calculation of a persymmetric sample matrix inversion beamformer," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 2135–2145, April 2016.

[16] W. Liu, Y. L. Wang, J. Liu, W. Xie, H. Chen, and R. Li, "Design and performance analysis of adaptive subspace detectors in orthogonal interference and Gaussian noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 5, pp. 2068–2079, October 2016.

[17] J. Liu, Z. Zhang, Y. Yang, and H. Liu, "A CFAR adaptive subspace detector for first-order or second-order Gaussian signals based on a single observation," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5126–5140, 2011.

[18] P. Wang, H. Li, and B. Himed, "Knowledge-aided parametric tests for multichannel adaptive signal detection," *IEEE Transactions on Signal Processing*, vol. 59, no. 12, pp. 5970–5982, 2011.

[19] J. Liu, W. Liu, B. Chen, H. Liu, H. Li, and C. Hao, "Modified Rao test for multichannel adaptive signal detection," *IEEE Transactions on Signal Processing*, vol. 64, no. 3, pp. 714–725, Feb 2016.

[20] Z. Chen, H. Li, and M. Rangaswamy, "Conjugate gradient adaptive matched filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 1, pp. 178–191, January 2015.

[21] J. Liu, H. Zhao, W. Liu, H. Li, and H Liu, "Adaptive detection using both the test and training data for noise correlation estimation," *Signal Processing*, vol. 137, pp. 309–318, 2017.

[22] W. Liu, W. Xie, J. Liu, and Y. Wang, "Adaptive double subspace signal detection in Gaussian background-Part I: homogeneous environments," *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2345–2357, 2014.

[23] W. Liu, W. Xie, J. Liu, and Y. Wang, "Adaptive double subspace signal detection in gaussian background-Part II: partially homogeneous environments," *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2358–2369, 2014.

[24] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 2, 1998.

[25] W. L. Melvin, "Space-time adaptive radar performance in heterogeneous clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 621–633, 2000.

[26] F. Gini and A. Farina, "Vector subspace detection in compound-Gaussian clutter Part I: Survey and new results," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 4, pp. 1295–1311, 2002.

[27] F. Gini and M. Greco, "Suboptimum approach to adaptive coherent radar detection in compound-Gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 3, pp. 1095–1103, 1999.

[28] Y. I. Abramovich and O. Besson, "On the expected likelihood approach for assessment of regularization covariance matrix," *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 777–781, 2014.

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

[30] A. De Maio, "Rao test for adaptive detection in Gaussian interference with unknown covariance matrix," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3577–3584, 2007.

[31] D. F. Marshall, "A two step adaptive interference nulling algorithm for use with airborne sensor arrays," in *IEEE Seventh SP Workshop on Statistical Signal and Array Processing*, June 1994, pp. 301–304.
[32] I.S. Reed, J.D. Mallett, and L.E. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 10, no. 4, pp. 853–863, 1974.
[33] F. Bandiera, D. Orlando, and G. Ricci, *Advanced Radar Detection Schemes Under Mismatched Signal Models*, San Rafael, US, 2009.
[34] D. J. Rabideau and S. M. Kogon, "A signal processing architecture for space-based gmti radar," in *Proceedings of the 1999 IEEE Radar Conference. Radar into the Next Millennium (Cat. No.99CH36249)*, 1999, pp. 96–101.
[35] L. Mirsky, "On the trace of matrix products," *Mathematische Nachrichten*, vol. 20, pp. 171–174, 1959.

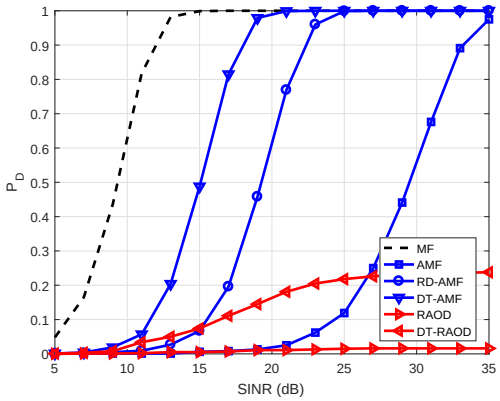

Fig. 3. P_D versus SINR for the MF, DT-AMF, RD-AMF, DT-RAOD, RAOD, and AMF assuming $N = 8$, $K = M = 9$, CNR= 20 dB, and JNR= 30 dB.

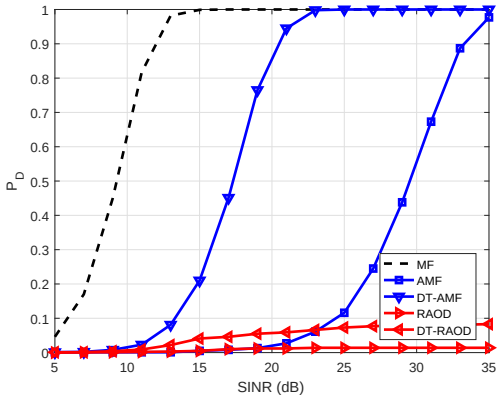

Fig. 4. P_D versus SINR for the MF, DT-AMF, DT-RAOD, RAOD, and AMF assuming $N = 8$, $K = M = 9$, CNR= 30 dB, and JNR= 30 dB.

Vincenzo Carotenuto Vincenzo Carotenuto (S'12, M'16) received the M.Sc. degree in telecommunication engineering and the Ph.D. degree in electronic and telecommunication engineering from the University of Naples Federico II, Naples, Italy, in in 2010 and 2015, respectively. His research interest lies in the field of statistical signal processing, with an emphasis on radar signal processing.

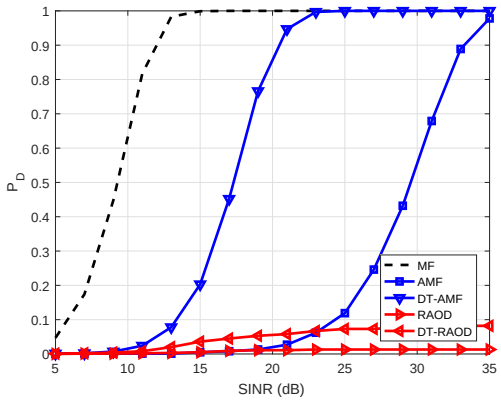

Fig. 5. P_D versus SINR for the MF, DT-AMF, DT-RAOD, RAOD, and AMF assuming $N = 8$, $K = M = 9$, CNR= 30 dB, and JNR= 40 dB.

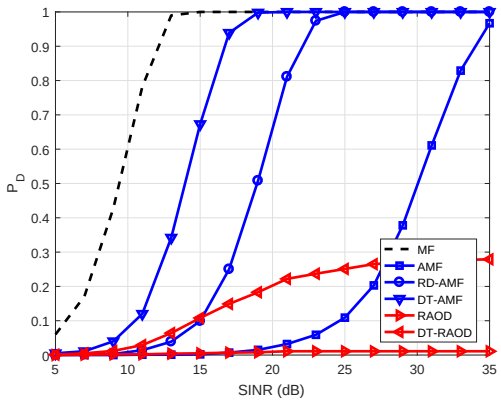

Fig. 6. P_D versus SINR for the MF, DT-AMF, RD-AMF, DT-RAOD, RAOD, and AMF assuming $N = 8$, $M = 9$, $K = 16$, CNR= 20 dB, and JNR= 30 dB.

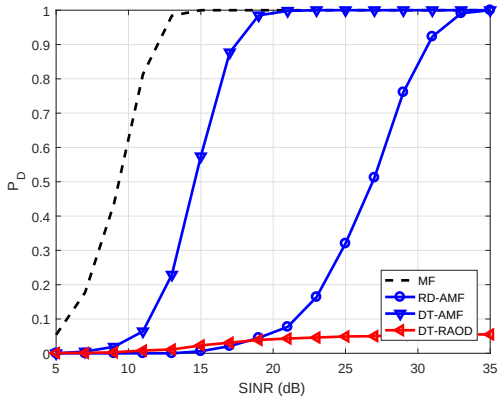

Fig. 7. P_D versus SINR for the MF, DT-AMF, RD-AMF, and DT-RAOD assuming $N = 8$, $M = 7$, $K = 16$, CNR= 20 dB, and JNR= 30 dB.

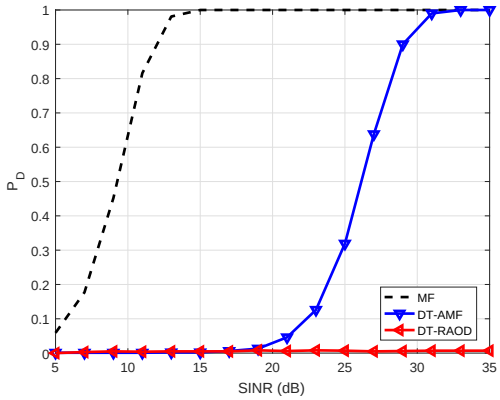

Fig. 8. P_D versus SINR for the MF, DT-AMF, and DT-RAOD assuming $N = 8$, $M = 5$, $K = 16$, CNR= 20 dB, and JNR= 30 dB.

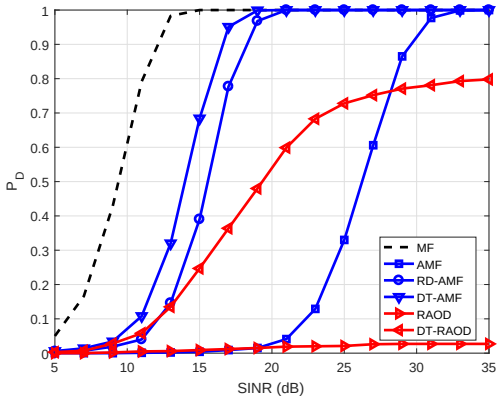

Fig. 9. P_D versus SINR for the MF, DT-AMF, RD-AMF, DT-RAOD, RAOD, and AMF assuming $N = 16$, $M = 18$, $K = 18$, CNR= 20 dB, and JNR= 30 dB.

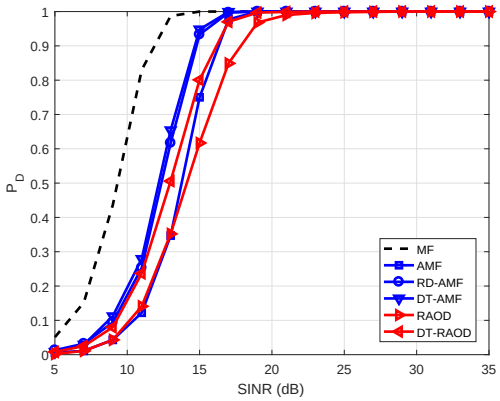

Fig. 10. P_D versus SINR for the MF, DT-AMF, RD-AMF, DT-RAOD, RAOD, and AMF assuming $N = 16$, $M = 32$, $K = 18$, CNR= 20 dB, and JNR= 30 dB.

Antonio De Maio Antonio De Maio (S'01-A'02-M'03-SM'07-F'13) was born in Sorrento, Italy, on June 20, 1974. He received the Dr.Eng. degree (with honors) and the Ph.D. degree in information engineering, both from the University of Naples Federico II, Naples, Italy, in 1998 and 2002, respectively. From October to December 2004, he was a Visiting Researcher with the U.S. Air Force Research Laboratory, Rome, NY. From November to December 2007, he was a Visiting Researcher with the Chinese University of Hong Kong, Hong Kong. Currently, he is an Associate Professor with the University of Naples Federico II. His research interest lies in the field of statistical signal processing, with emphasis on radar detection, optimization theory applied to radar signal processing, and multiple-access communications. Dr. De Maio is a Fellow member of IEEE and the recipient of the 2010 IEEE Fred Nathanson Memorial Award as the young (less than 40 years of age) AESS Radar Engineer 2010 whose performance is particularly noteworthy as evidenced by contributions to the radar art over a period of several years, with the following citation for "robust CFAR detection, knowledge-based radar signal processing, and waveform design and diversity"

Danilo Orlando Danilo Orlando (SM' 13) was born in Gagliano del Capo, Italy, on August 9, 1978. He received the Dr. Eng. Degree (with honors) in computer engineering and the Ph.D. degree (with maximum score) in information engineering, both from the University of Salento (formerly University of Lecce), Italy, in 2004 and 2008, respectively. From July 2007 to July 2010, he has worked with the University of Cassino (Italy), engaged in a research project on algorithms for track-before-detect of multiple targets in uncertain scenarios. From September to November 2009, he has been visiting scientist at the NATO Undersea Research Centre (NURC), La Spezia (Italy) where he collaborated with Dr. F. Ehlers on track-before-detect strategies for multistatic sonars. From September 2011 to April 2015, he has worked at Elettronica SpA engaged as system analyst in the field of Electronic Warfare. In May 2015, he joined Università degli Studi "Niccolò Cusano", where he is currently an Associate Professor. His main research interests are in the field of statistical signal processing and image processing with more emphasis on adaptive detection and tracking of multiple targets in multisensor scenarios. He has held visiting positions with the department of Avionics and Systems of ENSICA (now Institut Supérieur de l'Aéronautique et de l'Espace, ISAE), Toulouse (France) in February-March 2007, where he has worked with Prof. O. Besson on adaptive radar detection in presence of mismatched signals. He is Senior Member of IEEE and author or co-author of more than 80 scientific publications in international journals, conferences, and books.

Luca Pallotta Luca Pallotta (S'12, M'15) received the Laurea Specialistica degree (cum laude) in telecommunication engineering in 2009 from the University of Sannio, Benevento, Italy, and the Ph.D. degree in electronic and telecommunication engineering in 2014 from the University of Naples Federico II, Naples, Italy. His research interest lies in the field of statistical signal processing, with emphasis on radar signal processing and radar targets classification. Dr. Pallotta won the Student Paper Competition at the IEEE Radar Conference 2013.