Adaptive Radar Detection in the Presence of Multiple Alternative Hypotheses Using Kullback-Leibler Information Criterion-Part I: Detector Designs

Pia Addabbo, *Senior Member, IEEE*, Sudan Han, Filippo Biondi, *Member, IEEE*, Gaetano Giunta, *Senior Member, IEEE*, and Danilo Orlando, *Senior Member, IEEE*

***Abstract*—In this paper, we develop a new elegant systematic framework relying on the Kullback-Leibler Information Criterion to approach the design of one-stage adaptive detection architectures for multiple hypothesis testing problems in radar. Specifically, at the design stage, we assume that one out of several alternative hypotheses may be in force and that only one null hypothesis exists. Then, starting from the case where all the parameters are known and proceeding towards the adaptivity with respect to the entire parameter set, we come up with decision schemes for multiple alternative hypotheses consisting of the sum between the compressed log-likelihood ratio based upon the available data and a penalty term accounting for the number of unknown parameters. Such a term arises from suitable approximations of the Kullback-Leibler Divergence between the true and a candidate probability density function. Interestingly, under specific constraints, the proposed decision schemes can share the constant false alarm rate property by virtue of the Invariance Principle. Finally, we also show that the new architectures can be viewed as the result of a suitable regularization of the log-likelihood.**

***Index Terms*—Adaptive radar detection, constant false alarm rate, generalized likelihood ratio test, kullback-leibler information criterion, model order selection, multiple alternative hypothesis testing.**

I. Introduction

NOWADAYS, modern radar systems incorporate sophisticated signal processing algorithms which take advantage of the computational power made available by recent advances in technology. This growth in complexity is dictated by the fact that these systems have to face more and more challenging scenarios where conventional algorithms might fail or exhibit poor performance. For instance, in target-rich environments, structured echoes may contaminate data used to estimate the spectral properties of the interference (also known as training or secondary data) leading to a dramatic attenuation of the signal of interest and, hence, to a nonnegligible number of missed detections [1]. In such a case, radar systems should be endowed with signal processing schemes capable of detecting and suppressing the outliers in order to make the training data set homogeneous [2]–[5]. Another important example concerns high resolution radars, which can resolve a target into a number of different scattering centers depending on the radar bandwidth and the range extent of the target [6]. The classic approach to the detection of range-spread targets consists in processing one range bin at a time despite the fact that contiguous cells contain target energy and, then, in applying suitable integration procedures. As a consequence, classic detection algorithms do not collect as much energy as possible to increase the Signal-to-Interference-plus-Noise Ratio (SINR). To overcome this drawback, architectures capable of detecting distributed targets by exploiting a preassigned number of contiguous range bins have been developed [7]–[11]. These energy issues also hold for multiple point-like targets. In fact, detection algorithms which can take advantage of the cumulative energy associated with all the point-like targets are highly desirable.

In the open literature, the existing examples concerning the detection of either multiple point-like or range-spread targets share the assumption that the number of scatterers (or at least an upper bound on it) is known and are based upon the Maximum Likelihood (ML) approach [11]–[13]. However, in scenarios of practical value, such a priori information is not often available especially when the radar system is operating in search mode. Moreover, the problem of jointly detecting multiple point-like targets is very difficult since target positions and, more importantly, target number are unknown parameters that must be estimated. Thus, conversely to the conventional detection problems that comprise two hypotheses, which are the noise-only (or null) and the signal-plus-noise (or alternative) hypothesis, this lack of a priori information naturally leads to multiple alternative hypotheses with the consequence that the radar engineer has to face composite multiple hypothesis tests.

Besides high-resolution radars operating in target-dense environments, another situation leading to multiple hypothesis tests is related to possible electronic attacks from adversary forces (jammers). These attacks comprise active techniques aimed at

Manuscript received October 31, 2020; revised April 9, 2021; accepted June 2, 2021. Date of publication June 15, 2021; date of current version July 29, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. David Ramrez *(Corresponding author: Sudan Han.)*

Pia Addabbo is with the Università degli studi "Giustino Fortunato," Benevento, 00166 Roma, Italy (e-mail: p.addabbo@unifortunato.eu).

Sudan Han is with the Defense Innovation Institute, Beijing 100071, China (e-mail: xiaoxiaosu0626@163.com).

Filippo Biondi is with the Department of Engineering, University of L'Aquila, 67100 L'Aquila, Italy (e-mail: biopippo@gmail.com).

Gaetano Giunta is with the Department of Engineering, University of Roma Tre, 00146 Roma, Italy (e-mail: gaetano.giunta@uniroma3.it).

Danilo Orlando is with the Faculty of Engineering, Università degli Studi "Niccolò Cusano", 00166 Roma, Italy (e-mail: danilor78@gmail.com).

Digital Object Identifier 10.1109/TSP.2021.3089440

protecting a platform from being detected and tracked by the radar [6] through two approaches: masking and deception. More precisely, Noncoherent Jammers or Noise-Like Jammers (NLJs) attempt to mask targets generating nondeceptive interference which blends into the thermal noise of the radar receiver degrading the radar sensitivity due to an increase of the Constant False Alarm Rate (CFAR) threshold [6], [14], [15]. On the other hand, the Coherent Jammers (CJs) illuminate the victim radar by means of low duty-cycle signals with specific parameters that, when estimated by the radar processor, force the latter to allocate hardware resources for false target handling. In fact, CJs are equipped with electronic apparatuses capable of receiving, modifying, amplifying, and retransmitting the radar's own signal to create false targets with radar's range, Doppler, and angle far away from the true position of the platform under protection [6], [15].

A possible way to react to this kind of interference relies on the use of decision schemes devised by modifying the conventional detection problem with additional hypotheses accounting for such threats [16]–[18]. In [16], adaptive detection and discrimination between useful signals and CJs in the presence of thermal noise, clutter, and possible NLJs is addressed by considering an additional hypothesis under which data contain the CJs only. In addition, CJ signals are assumed to lie in the orthogonal complement of the subspace spanned by the nominal steering vector (after whitening by the true covariance matrix of the composite disturbance). The resulting multiple hypothesis test is solved by resorting to an approach based upon the ML principle and a generalized Neyman-Pearson criterion [19]–[21], which maximizes the overall probability of correct decision leading to architectures whose behavior depends on the combinations of different detection thresholds. This aspect is a bless and a curse since it lends such architectures a flexibility in terms of directivity but threshold setting might become very time-demanding (especially in the presence of a high number of alternative hypotheses) due to the infinite number of threshold combinations ensuring the same value for the Probability of False Alarm (P_{fa}) and yielding different detection performance. More importantly, such a solution is only effective when the multiple hypotheses are not nested. As a matter of fact, in the presence of nested hypotheses, the ML approach and, hence, the Generalized Likelihood Ratio Test (GLRT), may fail because the likelihood function monotonically increases with the number of unknown parameters (or model order). As a consequence, the ML approach experiences a natural inclination to overestimate the hypothesis order. An alternative approach consists in looking for detection schemes that incorporate the expedients of the so-called Model Order Selection (MOS) rules [22]–[25], which account for the parameter number diversity to moderate the overestimation attitude of the ML approach. In [17], the authors follow this last approach and conceive heuristic two-stage detection architectures for multiple NLJs whose number is unknown. Specifically, the first stage exploits the MOS rules to provide an estimate of the number of NLJs, whereas the second stage consists of a jammer detector that addresses a binary hypothesis test where the alternative hypothesis is selected by the first stage. Finally, it is important to observe that MOS rules can be adapted to accomplish detection tasks by also considering the model order "0" which is associated with the null hypothesis [26]–[28]. However, in this case, it is not possible to set any threshold in order to guarantee a preassigned P_{fa} and, more importantly, the CFAR property, which is of primary concern in radar, cannot be a priori stated.

In the first part of this work, we develop an elegant systematic framework relying on the Kullback-Leibler Information Criterion (KLIC) [29] to approach multiple hypothesis testing problems where there exist many alternative hypotheses (a preliminary work on KLIC-based detectors assuming the subspace model for the target can be found in [30]). It is important to recall here that in [17], this problem has been addressed resorting to heuristic design criteria that rely on the cascade of multiple stages, while the architectures developed within the herein proposed framework follow from a theoretical principle. In this sense, notice that a natural competitor for them is represented by the so-called Multifamily Likelihood Ratio Test (MFLRT) [31]. Moreover, the proposed framework provides an important interpretation of both the Likelihood Ratio Test (LRT) and the GLRT from an information theoretic standpoint and, remarkably, it lays the theoretical foundation for the design of new one-stage decision schemes for multiple hypothesis tests (this result represents the main technical contribution of this paper). Interestingly, the new detection architectures share a common structure that is given by the sum of a conventional decision statistic and a penalty term[1] and unlike the architectures devised in [16], such decision schemes compare a decision statistic with one threshold only.

The starting point of the developed framework is the case where all the parameters are known showing that under suitable regularity conditions the LRT approximates a test which selects the hypothesis with the associated Probability Density Function (PDF) minimizing the Kullback-Leibler Divergence (KLD) with respect to the true data distribution. In addition, the LRT coincides with such a test when the KLD is measured with respect to the Empirical Data Probability Density (EDPD) [32]. Then, we guide the reader towards more difficult scenarios where the distribution parameters are no longer known. Specifically, we resort to Taylor series approximations of the KLD, that are also used to derive the MOS rules [23], in order to come up with decision schemes capable of moderating the overfitting inclination of the ML approach. From a different perspective, the same results can be obtained by regularizing the PDF under the generic alternative hypothesis through a suitable prior for the unknown model order and applying a procedure that combines the ML approach and the Bayesian estimation [23], [33]. Finally, the proposed theoretical framework is completed with the investigation of the CFAR behavior of these decision schemes framing the analysis in the more general context of statistical invariance [34]. To this end, two propositions are provided that allow us to state when the newly proposed decision architectures are invariant with respect to a given group of transformations and, possibly, ensure the CFAR property. Before concluding this

[1]Such a structure is reminiscent of that of KLIC-based MOS rules, which consists of the compressed log-likelihood plus a penalty term.

section, we highlight that in the companion paper [35], which is the second part of this work, the focus is on the application of this new design framework to both real and synthetic aperture radar problems also in comparison with the MFLRT and architectures based upon the heuristic two-stage paradigm.

The remainder of this paper is organized as follows: the next section contains preliminary definitions and formulates the detection problem at hand in terms of a multiple hypothesis test. Section III describes the proposed framework and introduces the new architectures, whereas in Section IV some discussions about the CFAR properties exploiting the Principle of Invariance are also provided. Finally, concluding remarks and future research tracks are given in Section V. Some derivations and the description of the alternative framework leading to the same architectures are confined to the appendices.

A. Notation

In the sequel, vectors and matrices are denoted by boldface lower-case and upper-case letters, respectively. The symbols $\det(\cdot)$, $(\cdot)$, $(\cdot)^T$, $(\cdot)^\dagger$, and $(\cdot)^{-1}$ denote the determinant, trace, transpose, conjugate transpose, and inverse, respectively. As to numerical sets, $\mathbb{R}$ is the set of real numbers, $\mathbb{R}^{N\times M}$ is the Euclidean space 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 Euclidean space of $(N \times M)$-dimensional complex matrices (or vectors if $M = 1$). The Dirac delta function is indicated by $\delta(\cdot)$. The acronym IID means Independent and Identically Distributed, while symbol $E_f[\cdot]$ denotes the statistical expectation with respect to the PDF f. If A and B are two continuous random variables, $f(A|B)$ is the conditional PDF of A given B. Finally, $\boldsymbol{x} \sim f(\boldsymbol{x};\boldsymbol{\theta})$ means that f is the PDF of $\boldsymbol{x}$ with parameter vector $\boldsymbol{\theta}$.

II. Problem Formulation and Preliminary Definitions

Let us consider a radar system, equipped with N identical space and/or time channels, which collects a data matrix $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K] \in \mathbb{C}^{N\times K}$ whose[2] columns can be modeled as statistically independent random vectors whose distribution belongs to a preassigned family. For simplicity and in order not to burden the notation (a point better explained below), we assume that these vectors share the same distribution parameters and that their joint unknown PDF is denoted by $\bar{f}(\boldsymbol{Z};\boldsymbol{\theta})$ with $\boldsymbol{\theta} \in \mathbb{R}^{p\times 1}$ the parameter vector taking on value in a specific parameter space, $\Theta \subseteq \mathbb{R}^{p\times 1}$ say. Now, a conventional binary decision problem partitions the latter into two subsets Θ_0 and Θ_1 corresponding to the null (H_0) and the alternative (H_1) hypothesis, respectively. Thus, denoting by $\boldsymbol{\theta}_0$ the elements of Θ_0 and by $\boldsymbol{\theta}_1$ the elements of Θ_1, then the PDF of $\boldsymbol{Z}$ under the ith hypothesis can be written as $f_i(\boldsymbol{Z};\boldsymbol{\theta}_i)$, $i = 0, 1$. Hereafter, it is important to distinguish between the actual and unknown PDF of $\boldsymbol{Z}$, denoted by $\bar{f}(\boldsymbol{Z};\boldsymbol{\theta})$, and the PDF of $\boldsymbol{Z}$ when[3] $\boldsymbol{\theta} = \boldsymbol{\theta}_i \in \Theta_i$ (namely, $f_i(\boldsymbol{Z};\boldsymbol{\theta}_i)$).

Generally speaking, in many detection applications, the actual value of the parameter vector is unknown or, at most, partially known. In addition, not all the entries of $\boldsymbol{\theta}$ are useful to take a decision for the specific problem at hand. As a consequence, we can split $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = [\boldsymbol{\theta}_r^T, \boldsymbol{\theta}_s^T]^T$, where $\boldsymbol{\theta}_r \in \Theta_r \subseteq \mathbb{R}^{p_r\times 1}$ contains the parameters of interest, while the components of $\boldsymbol{\theta}_s \in \Theta_s \subseteq \mathbb{R}^{p_s\times 1}$ represent the parameters that do not enter into the decision process[4] and are called *nuisance parameters* [37]. From a more general perspective, it would be possible to associate a parameter vector of interest to each column of $\boldsymbol{Z}$ with the consequence that $\boldsymbol{z}_k$s are not identically distributed. However, as stated before, we prefer to proceed assuming that $\boldsymbol{z}_k$s share the same parameter vector in order to maintain an easy notation and because the extension to the more general case is straightforward as we will show in the examples introduced in the companion paper [35] (Part II). Thus, with the above remarks in mind, a conventional binary hypothesis testing problem can be expressed as

$$\begin{cases} H_0 : \boldsymbol{Z} \sim f_0(\boldsymbol{Z};\boldsymbol{\theta}_0) = f_0(\boldsymbol{Z};\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s) \\ \qquad\qquad\qquad = \displaystyle\prod_{k=1}^{K} g_0(\boldsymbol{z}_k;\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s), \\ H_1 : \boldsymbol{Z} \sim f_1(\boldsymbol{Z};\boldsymbol{\theta}_1) = f_1(\boldsymbol{Z};\boldsymbol{\theta}_{r,1},\boldsymbol{\theta}_s) \\ \qquad\qquad\qquad = \displaystyle\prod_{k=1}^{K} g_1(\boldsymbol{z}_k;\boldsymbol{\theta}_{r,1},\boldsymbol{\theta}_s), \end{cases} \tag{1}$$

where $\boldsymbol{\theta}_i = [\boldsymbol{\theta}_{r,i}^T, \boldsymbol{\theta}_s^T]^T \in \mathbb{R}^{p\times 1}$, $i = 0, 1$, with $\boldsymbol{\theta}_{r,i} \in \Theta_r^i \subseteq \mathbb{R}^{p_r\times 1}$ the parameter vector of interest under H_i, and $g_i(\cdot;\cdot)$, $i = 0, 1$, is the PDF of $\boldsymbol{z}_k$, $k = 1, \ldots, K$, under H_i.

Two remarks are now in order. First, note that $\{\Theta_r^0, \Theta_r^1\}$ is a partition of Θ_r. Second, the above problem assumes that p_r, p_s, and, hence, the total number of unknown parameters p, are perfectly known. However, in many radar (and, generally speaking, signal processing) applications, it is common for p to be unknown under the alternative hypothesis due to the fact that the size of $\boldsymbol{\theta}_{r,1}$ might depend on the specific operating scenario [23]. For instance, radar systems might face situations where an unknown number of targets are present in the region of interest [12], [13] or be under attack from an unknown number of jammers [17], [18], [38] (further examples can be found in the second part of this work). Under this assumption, the hypothesis test can be modified as

$$\begin{cases} \text{under } H_0 : \ \boldsymbol{\theta}_{r,0} \in \Theta_r^0 \subseteq \mathbb{R}^{p_{r,0}\times 1}, \\ \text{under } H_1 : \ \boldsymbol{\theta}_{r,1}^m \in \Theta_r^m \subseteq \mathbb{R}^{p_{r,m}\times 1}, \ p_{r,m} \in \Omega_r, \end{cases} \tag{2}$$

where only $\Omega_r = \{p_{r,1}, \ldots, p_{r,M}\}$ with $p_{r,1} \leq \ldots \leq p_{r,M}$, $M \in \mathbb{N}$, and $p_{r,0}$ are known. Moreover, M represents an upper bound on the number of the alternative hypotheses that can be

[2]Observe that in the case of Space-Time Adaptive Processing (STAP), N represents the number of space-time channels, whereas when the system transmits either a single pulse through an array of sensors or exploits a single antenna to transmit a pulse train, N represents the number of either the elements of the spatial array or the transmitted pulses, respectively [36]. Finally, K may represent the number of range bins or the number of pulses when a slice of the STAP datacube is processed.

[3]In what follows, we assume correctly specified models, namely that data distribution family is known.

[4]Note that p_r is the number of the parameters of interest and p_s is the number of the nuisance parameters. It follows that $p = p_r + p_s$.

set, for instance, according to the operating and/or computational requirements. It follows that the PDF of $\boldsymbol{Z}$ (and, hence, that of $\boldsymbol{z}_k$) under H_1 depends on $p_{r,m}$ and, more importantly, the uncertainty on the latter leads to a testing problem formed by multiple (possibly nested) H_1 hypotheses, i.e.,

$$\begin{cases} H_0: & \boldsymbol{Z} \sim f_0(\boldsymbol{Z}; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s), \\ H_{1,1}: & \boldsymbol{Z} \sim f_{1,1}(\boldsymbol{Z}; \boldsymbol{\theta}^1_{r,1}, \boldsymbol{\theta}_s, p_{r,1}), \\ \vdots & \vdots \\ H_{1,M}: & \boldsymbol{Z} \sim f_{1,M}(\boldsymbol{Z}; \boldsymbol{\theta}^M_{r,1}, \boldsymbol{\theta}_s, p_{r,M}), \end{cases} \tag{3}$$

where

$$f_{1,m}(\boldsymbol{Z}; \boldsymbol{\theta}^m_{r,1}, \boldsymbol{\theta}_s, p_{r,m}) = \prod_{k=1}^{K} g_{1,m}(\boldsymbol{z}_k; \boldsymbol{\theta}^m_{r,1}, \boldsymbol{\theta}_s, p_{r,m}) \tag{4}$$

is the PDF of $\boldsymbol{Z}$ under $H_{1,m}$, namely when $\boldsymbol{\theta}^m_{r,1} \in \Theta^m_r$, $m = 1, \ldots, M$, with $g_{1,m}(\boldsymbol{z}_k; \boldsymbol{\theta}^m_{r,1}, \boldsymbol{\theta}_s, p_{r,m})$ the PDF of $\boldsymbol{z}_k$ under $H_{1,m}$. In what follows, we denote by $\bar{m}$ the true model order under the alternative hypothesis.

In the next section, we propose an *Information-Theoretic* based approach to deal with problem (3) exploiting the KLIC [29]. Specifically, this criterion relies on the measurement of a certain distance, the so-called KLD, between a candidate distribution belonging to the family of densities $\mathcal{F} = \{f_0,\ f_{1,m},\ m \in \{1, \ldots, M\}\}$ and the actual distribution of $\boldsymbol{Z}$, which is assumed to lie in $\mathcal{F}$ and denoted by

$$\bar{f}(\boldsymbol{Z}; \boldsymbol{\theta}) = \prod_{k=1}^{K} \bar{g}(\boldsymbol{z}_k; \boldsymbol{\theta}), \tag{5}$$

where $\bar{g}(\boldsymbol{z}_k; \boldsymbol{\theta})$ is the true PDF of $\boldsymbol{z}_k$, $k = 1, \ldots, K$. Besides, we suppose that the inequalities required to invoke the Kintchine's Strong Law of Large Numbers [39] are valid, namely

$$|E_{\bar{g}}[\log g_{1,m}(\boldsymbol{z}_k; \boldsymbol{\theta}^m_{r,1}, \boldsymbol{\theta}_s, p_{r,m})]| < +\infty, \tag{6}$$

$$|E_{\bar{g}}[\log g_0(\boldsymbol{z}_k; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s)]| < +\infty, \tag{7}$$

as well as the following "regularity conditions" [23]

$$\begin{aligned} &\frac{1}{\mathcal{T}} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \left[\log \prod_{k=1}^{K} g_{1,m}(\boldsymbol{z}_k; \boldsymbol{\theta}^m_{r,1}, \boldsymbol{\theta}_s, p_{r,m}) \right] \\ &\quad \stackrel{\mathcal{T} \to \infty}{\longrightarrow} \frac{1}{\mathcal{T}} E \left\{ \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \left[\log \prod_{k=1}^{K} g_{1,m}(\boldsymbol{z}_k; \boldsymbol{\theta}^m_{r,1}, \boldsymbol{\theta}_s, p_{r,m}) \right] \right\}, \end{aligned} \tag{8}$$

and $p/\mathcal{T} \stackrel{\mathcal{T} \to \infty}{\longrightarrow} 0$, which are required to suitably approximate the KLD. In the above equations, $\mathcal{T}$ represents the total number of real-valued observations, that, for the problem at hand, is equal to $2NK$ since we are dealing with complex vectors. In addition, $\mathcal{T}$ becomes infinite in the limit of large N and/or K values.

Notice that a "minimum information distance" selection criterion has already been successfully applied for model order estimation giving rise to the so-called MOS rules [23], [40]. The resulting selection architectures share the same structure consisting of a fitting term (the compressed log-likelihood function) plus an adjustment which also depends on the number of parameters. As a consequence, the parameter number diversity comes into play to moderate the overfitting inclination of the compressed likelihood in the case of nested hypotheses. In fact, in this context, the KLIC-based MOS rules can provide satisfactory classification performance whereas the ML approach may fail because the likelihood function monotonically increases with $p_{r,m}$ and the Maximum Likelihood Estimate (MLE) of p_r will always be $p_{r,M}$ (or, equivalently, the MLE of p will be $p_{r,M} + p_s$).

To conclude this preliminary section, for the reader ease, we recall that the KLD [29] (also called relative entropy) between $\bar{f}$ and $f_{1,m}$, namely, the PDF of a generic candidate model under H_1 (or under $H_{1,m}$), can be written as [41]

$$D(\bar{f} \| f_{1,m}) = \int_{-\infty}^{\infty} \bar{f}(\boldsymbol{Z}) \log \frac{\bar{f}(\boldsymbol{Z})}{f_{1,m}(\boldsymbol{Z}; \boldsymbol{\theta}^m_{r,1}, \boldsymbol{\theta}_s, p_{r,m})} d\boldsymbol{Z}, \tag{9}$$

where $d\boldsymbol{Z} = dz^r_{1,1} dz^i_{1,1} \ldots dz^r_{N,K} dz^i_{N,K}$ with $z^r_{n,k}$ and $z^i_{n,k}$ the real and imaginary parts of the nth component of $\boldsymbol{z}_k$, and can be decomposed into the sum of two terms[5]

$$\begin{aligned} D(\bar{f} \| f_{1,m}) &= \int_{-\infty}^{\infty} \bar{f}(\boldsymbol{Z}) \log \bar{f}(\boldsymbol{Z}) d\boldsymbol{Z} \\ &\quad - \int_{-\infty}^{\infty} \bar{f}(\boldsymbol{Z}) \log f_{1,m}(\boldsymbol{Z}; \boldsymbol{\theta}^m_{r,1}, \boldsymbol{\theta}_s, p_{r,m}) d\boldsymbol{Z} \\ &= -h(\bar{f}) + h(\bar{f}, f_{1,m}), \end{aligned} \tag{10}$$

where $h(\bar{f})$ is the differential entropy of $\bar{f}$ and $h(\bar{f}, f_{1,m})$ is the cross entropy between $\bar{f}$ and $f_{1,m}$. Note that unlike $h(\bar{f})$, $h(\bar{f}, f_{1,m})$ depends on the mth model (or alternative hypothesis). Analogously, we can write the KLD with respect to the PDF under H_0 as

$$D(\bar{f} \| f_0) = -h(\bar{f}) + h(\bar{f}, f_0). \tag{11}$$

Recall that $D(\cdot \| \cdot)$ is not a true distance between distributions since it is not symmetric and does not satisfy the triangle inequality [41]. Nonetheless, it is often useful to think of the KLD as a "distance" between distributions. Finally, the KLD can be interpreted as the information loss when either $f_{1,m}$ or f_0 is used to approximate $\bar{f}$ [22].

III. KLIC-Based Decision Rules

In this section, we exploit the KLIC to devise decision schemes for problem (3). The motivation behind this choice is twofold. On one side, it is linked to the fact that, as shown in [23], starting from Kullback-Leibler (KL) divergence it is possible to obtain the MOS rules and, hence, it is worth investigating whether or not KL divergence allows us to come up with decision schemes with model order selection capabilities. On the other side, as shown below, KL divergence leads to tractable mathematics and provides a derivation of well known decision theory tools from an information theory standpoint. In what follows, we proceed by considering the case where all the parameters are known and then evolve to more difficult cases where the parameters become unknown. It is important to

[5]We assume that the considered PDFs exist with respect to a Lebesgue measure.

observe that we define here an information theoretic framework where well-established decision rules such as the LRT and the GLRT are suitably encompassed as special cases.

A. KLIC-Based Detectors for Known Model and Parameters

In this case, the number of alternative hypotheses is reduced to 1 and problem (3) turns into problem (1) with the additional assumption that $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are known. Moreover, the number of parameters of interest is denoted by $p_{r,\bar{m}}$ and is assumed known. As a consequence, data distribution is completely determined by the hypotheses and the true PDF of $\boldsymbol{Z}$ belongs to the following family $\mathcal{F}_{\theta,p} = \{f_0,\ f_{1,\bar{m}}\}$. Thus, a natural test based upon KLIC would decide for the hypothesis with the associated PDF (computed at $\boldsymbol{Z}$) exhibiting the "minimum distance" from $\bar{f}$. Otherwise stated, such a test can be formulated as

$$D(\bar{f}\|f_0) \underset{H_0}{\overset{H_1}{\gtrless}} D(\bar{f}\|f_{1,\bar{m}}). \tag{12}$$

The above[6] rule selects H_0 if the distance between $\bar{f}$ and f_0 is lower than that between $\bar{f}$ and $f_{1,\bar{m}}$. In the opposite case, it decides for H_1. Now, let us exploit (10) and (11) to recast (12) as

$$-h(\bar{f}) + h(\bar{f}, f_0) \underset{H_0}{\overset{H_1}{\gtrless}} -h(\bar{f}) + h(\bar{f}, f_{1,\bar{m}}), \tag{13}$$

or, equivalently,

$$h(\bar{f}, f_0) - h(\bar{f}, f_{1,\bar{m}}) \underset{H_0}{\overset{H_1}{\gtrless}} 0. \tag{14}$$

From the information theory point of view, the above forms highlight that the considered decision rule minimizes the loss of information which occurs when $\bar{f}$ is approximated with $f_{1,\bar{m}}$ or f_0 [22]. Moreover, using (10), we can rewrite (14) as

$$E_{\bar{f}}[\log f_{1,\bar{m}}(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_s, p_{r,\bar{m}})] - E_{\bar{f}}[\log f_0(\boldsymbol{Z}; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s)] \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{15}$$

$$\Rightarrow \sum_{k=1}^{K} E_{\bar{g}}[\log g_{1,\bar{m}}(\boldsymbol{z}_k; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_s, p_{r,\bar{m}})] - \sum_{k=1}^{K} E_{\bar{g}}[\log g_0(\boldsymbol{z}_k; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s)] \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{16}$$

$$\Rightarrow E_{\bar{g}}[\log g_{1,\bar{m}}(\boldsymbol{z}; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_s, p_{r,\bar{m}})] - E_{\bar{g}}[\log g_0(\boldsymbol{z}; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s)] \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{17}$$

$$\Rightarrow \Lambda(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s, \bar{m}) = E_{\bar{g}}\left[\log \frac{g_{1,\bar{m}}(\boldsymbol{z}; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_s, p_{r,\bar{m}})}{g_0(\boldsymbol{z}; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s)}\right] \underset{H_0}{\overset{H_1}{\gtrless}} 0. \tag{18}$$

Note that starting from (17), we have dropped the subscript of $\boldsymbol{z}_k$ since $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K$ are IID random vectors. Moreover, equation (16) is obtained by replacing (4) into (15), whereas equation (17) comes from the fact that the $\boldsymbol{z}_k$s are identically distributed. Test (18) cannot be applied in practice because it involves the computation of the expected log-likelihood ratio and requires the knowledge of $\bar{g}(\cdot)$ (or, equivalently, of $\bar{f}(\cdot)$). For this reason, we replace $\Lambda(\cdot;\cdot)$ with a suitable estimate which is function of the observed data. To this end, we resort to the following unbiased estimator [42]

$$\widehat{\Lambda}(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s, \bar{m}) = \frac{1}{K}\sum_{k=1}^{K} \log \frac{g_{1,\bar{m}}(\boldsymbol{z}_k; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_s, p_{r,\bar{m}})}{g_0(\boldsymbol{z}_k; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s)} = \frac{1}{K} \log \frac{f_{1,\bar{m}}(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_s, p_{r,\bar{m}})}{f_0(\boldsymbol{Z}; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s)}. \tag{19}$$

In fact, since the technical assumptions of the Kintchine's Strong Law of Large Numbers hold true, due to (6) and (7), in the limit for $K \to +\infty$, we have that $\widehat{\Lambda}(\cdot;\cdot)$ converges almost surely to $\Lambda(\cdot;\cdot)$, namely

$$\widehat{\Lambda}(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s, \bar{m}) \xrightarrow{a.s.} \Lambda(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s, \bar{m}). \tag{20}$$

Summarizing, test (18) is replaced by

$$\widehat{\Lambda}(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}^{\bar{m}}, \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s, \bar{m}) \underset{H_0}{\overset{H_1}{\gtrless}} 0, \tag{21}$$

where $\widehat{\Lambda}(\cdot;\cdot)$ is a random variable whose value depends on the observed data.

An alternative derivation for (21) consists in replacing $f_{1,\bar{m}}(\cdot)$ and $f_0(\cdot)$ with $g_{1,\bar{m}}(\cdot)$ and $g_0(\cdot)$, respectively, in (12), while the EDPD [32], whose expression is

$$\Xi(\boldsymbol{z}) = \frac{1}{K}\sum_{k=1}^{K} \delta(\boldsymbol{z} - \boldsymbol{z}_k), \tag{22}$$

is used in place of $\bar{f}(\cdot)$. Thus, the KLD between the EDPD and the distribution of a generic $\boldsymbol{z}_k$ is measured to decide which hypothesis is in force. As a matter of fact, criterion (12) becomes

$$D(\Xi\|g_0) \underset{H_0}{\overset{H_1}{\gtrless}} D(\Xi\|g_{1,\bar{m}}) \tag{23}$$

[6]As shown below, using $D(\bar{f}\|f_0)$ and $D(\bar{f}\|f_{1,\bar{m}})$ instead of $D(f_0\|\bar{f})$ and $D(f_{1,\bar{m}}\|\bar{f})$, respectively, allows us to obtain a reasonable estimate of the statistical expectation with respect to $\bar{f}$ that is unknown.

and (as shown in Appendix A) leads to[7]

$$\log \frac{\prod_{k=1}^{K} g_{1,\bar{m}}(\boldsymbol{z}_k;\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s,p_{r,\bar{m}})}{\prod_{k=1}^{K} g_0(\boldsymbol{z}_k;\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s)} \underset{H_0}{\overset{H_1}{\gtrless}} 0. \tag{24}$$

Finally, note that the detection threshold of test (21) is set to zero and, as a consequence, it does not allow for a control on the probability of type I error also known as P_{fa} in Detection Theory [37]. This issue is of primary importance in radar, since the Probability of Detection P_d and the P_{fa} are interconnected. Thus, in order to introduce a tuning parameter for the P_{fa}, instead of selecting the PDF corresponding to the minimum KLD, we can consider the following modification of (12)

$$D(\bar{f}\|f_0) \underset{H_0}{\overset{H_1}{\gtrless}} D(\bar{f}\|f_{1,\bar{m}}) + \eta, \tag{25}$$

where η represents a degree of freedom in deciding for H_0 or H_1. Therefore, repeating the above line of reasoning, the decision rule (24) becomes

$$\log \frac{f_{1,\bar{m}}(\boldsymbol{Z};\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s,p_{r,\bar{m}})}{f_0(\boldsymbol{Z};\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{26}$$

where the detection threshold, η, is suitably tuned in order to guarantee the desired[8] P_{fa}. Remarkably, the decision rule (26) is statistically equivalent to the LRT or the Neyman-Pearson test [34].

B. KLIC-Based Detectors for Known Model and Unknown Parameters

Let us now consider (15) again and assume that only the number of parameters of interest, $p_{r,\bar{m}}$ say, is known, whereas $\boldsymbol{\theta}_{r,1}$, $\boldsymbol{\theta}_{r,0}$, and $\boldsymbol{\theta}_s$ are unknown. In this case the family of candidate models becomes

$$\begin{aligned}\mathcal{F}_p = \mathcal{F}_0 \cup \mathcal{F}_1 = \{f_0(\cdot;\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s) : \boldsymbol{\theta}_{r,0} \in \Theta_r^0,\ \boldsymbol{\theta}_s \in \Theta_s\} \\ \cup \{f_{1,\bar{m}}(\cdot;\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s,p_{r,\bar{m}}) :\ \boldsymbol{\theta}_{r,1}^{\bar{m}} \in \Theta_r^{\bar{m}},\ \boldsymbol{\theta}_s \in \Theta_s\},\end{aligned} \tag{27}$$

where Θ_r^0 and $\Theta_r^{\bar{m}}$ form a partition of the parameter space of interest while Θ_s is the nuisance parameter space. Note that in this case, the hypotheses of (1) become composite implying that, in order to build up a decision rule based upon (12), the unknown parameters $\boldsymbol{\theta}_{r,1}^{\bar{m}}$, $\boldsymbol{\theta}_{r,0}$, and $\boldsymbol{\theta}_s$ must be estimated from data. Among different alternatives, we resort to the ML approach, which enjoys "good" asymptotic properties[9] [48]. In fact, given a model, the consistency (when it holds) of the MLE ensures that it converges in probability to the true parameter value, which is also the minimizer of the KLD (see [22], [49] and references therein). Thus, in (15), we replace $\boldsymbol{\theta}_{r,1}^{\bar{m}}$, $\boldsymbol{\theta}_{r,0}$, and $\boldsymbol{\theta}_s$ with their respective MLEs under each hypothesis. Specifically, denoting by $\widehat{\boldsymbol{\theta}}_{r,0}$ and $\widehat{\boldsymbol{\theta}}_{r,1}^{\bar{m}}$ the MLE of $\boldsymbol{\theta}_{r,0}$ and $\boldsymbol{\theta}_{r,1}^{\bar{m}}$, respectively, as well as by $\widehat{\boldsymbol{\theta}}_{s,i}$ the MLE of $\boldsymbol{\theta}_s$ under the H_i hypothesis, $i = 0, 1$, (15) can be recast as

$$\begin{aligned}&E_{\bar{f}}[\log f_{1,\bar{m}}(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}_{r,1}^{\bar{m}},\widehat{\boldsymbol{\theta}}_{s,1},p_{r,\bar{m}})] \\ &- E_{\bar{f}}[\log f_0(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}_{r,0},\widehat{\boldsymbol{\theta}}_{s,0})] \underset{H_0}{\overset{H_1}{\gtrless}} 0 \\ &\Rightarrow \Lambda_1(\boldsymbol{Z};\bar{m}) = E_{\bar{f}}\left[\log \frac{f_{1,\bar{m}}(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}_{r,1}^{\bar{m}},\widehat{\boldsymbol{\theta}}_{s,1},p_{r,\bar{m}})}{f_0(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}_{r,0},\widehat{\boldsymbol{\theta}}_{s,0})}\right] \underset{H_0}{\overset{H_1}{\gtrless}} 0.\end{aligned} \tag{28}$$

Now, in place of the expectation with respect to the unknown $\bar{f}$, we use an unbiased estimator of $\Lambda_1(\boldsymbol{Z};\bar{m})$ (see also equation (48) of [23]), namely

$$\widehat{\Lambda}_1(\boldsymbol{Z};\bar{m}) = \log \frac{f_{1,\bar{m}}(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}_{r,1}^{\bar{m}},\widehat{\boldsymbol{\theta}}_{s,1},p_{r,\bar{m}})}{f_0(\boldsymbol{Z};\widehat{\boldsymbol{\theta}}_{r,0},\widehat{\boldsymbol{\theta}}_{s,0})}, \tag{29}$$

and introduce a threshold to control the P_{fa} yielding

$$\widehat{\Lambda}_1(\boldsymbol{Z};\bar{m}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{30}$$

It is important to underline that the above test is statistically equivalent to the GLRT for known p_r.

The same result can be derived showing that the MLEs minimize the KLD between the EDPD and the candidate PDF with respect to the unknown parameters. To this end, let us start from (23) and minimize both sides with respect to the unknown parameters

$$\min_{\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s} D(\Xi\|g_0) \underset{H_0}{\overset{H_1}{\gtrless}} \min_{\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s} D(\Xi\|g_{1,\bar{m}}). \tag{31}$$

As shown in Appendix A, the above problem[10] yields

$$\log \frac{\prod_{k=1}^{K} g_{1,\bar{m}}(\boldsymbol{z}_k;\widehat{\boldsymbol{\theta}}_{r,1}^{\bar{m}},\widehat{\boldsymbol{\theta}}_{s,1},p_{r,\bar{m}})}{\prod_{k=1}^{K} g_0(\boldsymbol{z}_k;\widehat{\boldsymbol{\theta}}_{r,0},\widehat{\boldsymbol{\theta}}_{s,0})} \underset{H_0}{\overset{H_1}{\gtrless}} 0, \tag{32}$$

whose left-hand side coincides with (29), namely, it is statistically equivalent to the GLRT.

[7]It is important to stress here that (23) differs from (12) in the fact that $\bar{f}$ is replaced by the EDPD.

[8]Hereafter, symbol η will be used to denote the generic detection threshold.

[9]However, in situations of practical interest, the amount of data makes the ML estimates far from the asymptotic regime (see, for instance, Section III.A of Part II) and might severely degrade the estimation fidelity. In such situations, alternative strategies can be pursued such as the Bayesian estimation (if the prior can be suitably modeled), specific covariance regularizations [43]–[45], or the exploitation of a priori information about system/interference symmetries [8], [46], [47].

[10]Notice that criterion (31) first estimates the unknown parameters to minimize the KLD between the candidate PDF and the EDPD of $\boldsymbol{z}_k$ and, then, selects the hypothesis corresponding to the minimum KLD.

C. KLIC-Based Detectors for Unknown Model and Parameters

In this last subsection, we deal with the most general case where $p_r \in \Omega_r$, $\boldsymbol{\theta}_{r,1}^m$, $\boldsymbol{\theta}_{r,0}$, and $\boldsymbol{\theta}_s$ are unknown. As stated in Section II, under this assumption, there exist multiple alternative hypotheses depending on the model order $p_m = p_{r,m} + p_s$.

As possible strategy to select the most likely hypothesis, we might follow the same line of reasoning as in the previous subsection replacing p_r with its MLE. However, if on one hand, this approach (given p_r) makes sense for $\boldsymbol{\theta}_{r,1}$, $\boldsymbol{\theta}_{r,0}$, and $\boldsymbol{\theta}_s$, on the other, when the considered models are nested, it fails in the estimation of p_r. In fact, the log-likelihood function monotonically increases with p_r and, as a consequence, the ML approach will always select the maximum order leading to an overfitting. Therefore, an alternative approach is required in order to find "good" approximations of the negative cross entropy which moderate the overfitting inclination of the ML approach. To this end, we follow the same line of reasoning used to derive the MOS rules in [23], where suitable Taylor series expansions of the cross entropy (used in (15)) are exploited. Then, the dependence on p_m is removed by optimizing these expansions over the latter, which is tantamount to minimizing the approximations of the KLD between $\bar{f}$ and $f_{1,m}$ with respect to the unknown parameters.

As for the null hypothesis, it is independent of $p_{r,m}$ and, hence, we can exploit previously devised estimators. Specifically, we replace $E_{\bar{f}}[\log f_0(\boldsymbol{Z}; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s)]$ with the same estimator as in the previous subsection, namely $\log f_0(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_{r,0}, \widehat{\boldsymbol{\theta}}_{s,0})$.

Now, let us define $\mathfrak{I}_m(\boldsymbol{\theta}_{r,1}^m, \boldsymbol{\theta}_s) = -h(\bar{f}, f_{1,m})$ and denote by $\widehat{\mathfrak{I}}_m$ an estimator of the former. Following the lead of [23], several alternatives are possible for $\widehat{\mathfrak{I}}_m$, namely

- through equations (57)-(62) of [23], we obtain

$$\widehat{\mathfrak{I}}_m = \log f_{1,m}(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_{r,1}^m, \widehat{\boldsymbol{\theta}}_{s,1}, p_{r,m}) - (p_{r,m} + p_s); \quad (33)$$

- through equations (59)-(60) and (73) of [23], we have that

$$\widehat{\mathfrak{I}}_m = \log f_{1,m}(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_{r,1}^m, \widehat{\boldsymbol{\theta}}_{s,1}, p_{r,m}) - \frac{1+\rho}{2}(p_{r,m} + p_s),\ \rho > 1; \quad (34)$$

- through equations (79)-(86) of [23], we get

$$\widehat{\mathfrak{I}}_m = \log f_{1,m}(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_{r,1}^m, \widehat{\boldsymbol{\theta}}_{s,1}, p_{r,m}) - \frac{p_{r,m} + p_s}{2} \log(\mathcal{T}) + C, \quad (35)$$

where C is a constant. Note that (35) results from an asymptotic approximation for sufficiently large values of $\mathcal{T}$ of the more general form [23]

$$\widehat{\mathfrak{I}}_m = \log f_{1,m}(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_{r,1}^m, \widehat{\boldsymbol{\theta}}_{s,1}, p_{r,m}) - \frac{1}{2} \log\det \widehat{\boldsymbol{J}} + C, \quad (36)$$

where

$$\widehat{\boldsymbol{J}} = \left[-\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \log f_{1,m}(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}^m, \boldsymbol{\theta}_{s,1}, p_{r,m}) \right]\Bigg|_{\substack{\boldsymbol{\theta}_{r,1}^m = \widehat{\boldsymbol{\theta}}_{r,1}^m \\ \boldsymbol{\theta}_{s,1} = \widehat{\boldsymbol{\theta}}_{s,1}}} \quad (37)$$

It is clear that other approximations are possible by considering the asymptotic behavior with respect to suitable subsets of the parameters.

Finally, an estimate of m can be obtained as

$$\widehat{m} = \arg\max_{m \in \{1,\ldots,M\}} \widehat{\mathfrak{I}}_m \quad (38)$$

with $p_{r,\widehat{m}}$ the corresponding estimate of $p_r \in \Omega_r$, and we can replace each addendum of (15) with the respective approximation to come up with the following adaptive rule

$$\widehat{\mathfrak{I}}_{\widehat{m}} - \log f_0(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_{r,0}, \widehat{\boldsymbol{\theta}}_{s,0}) \underset{H_0}{\overset{H_{1,\widehat{m}}}{\gtrless}} 0, \quad (39)$$

which, introducing the threshold to control the P_{fa}, can be recast as

$$\max_{m \in \{1,\ldots,M\}} \left\{ \log f_{1,m}(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_{r,1}^m, \widehat{\boldsymbol{\theta}}_{s,1}, p_{r,m}) - h(m) \right\} - \log f_0(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_{r,0}, \widehat{\boldsymbol{\theta}}_{s,0}) \underset{H_0}{\overset{H_{1,\widehat{m}}}{\gtrless}} \eta$$

$$\Rightarrow \max_{m \in \{1,\ldots,M\}} \left\{ \widehat{\Lambda}_1(\boldsymbol{Z}; m) - h(m) \right\} \underset{H_0}{\overset{H_{1,\widehat{m}}}{\gtrless}} \eta, \quad (40)$$

where $\widehat{\Lambda}_1(\boldsymbol{Z}; m)$ is given by (29) with m in place of $\bar{m}$ and

$$h(m) = \begin{cases} (p_{r,m} + p_s), & \text{(a)} \\ \dfrac{1+\rho}{2}(p_{r,m} + p_s),\ \rho > 1, & \text{(b)} \\ \dfrac{(p_{r,m} + p_s)}{2} \log(\mathcal{T}), & \text{(c)} \end{cases} \quad (41)$$

is a penalty term inherited from MOS rules [23]. It is important to highlight that (a), (b), and (c) are related to the Akaike Information Criterion, Generalized Information Criterion, and Bayesian Information Criterion, respectively.

Before concluding this section, two important remarks are in order. Firstly, notice that test (40) solves problem (3), which contains only one null hypothesis and several alternative hypotheses; as shown in the companion paper (Part II), in a radar scenario, this kind of hypothesis testing problems naturally arises. Therefore, we can adopt the conventional definition for the P_{fa}, i.e., the probability of rejecting H_0 when the latter is true. Now, in order to set the detection threshold, it is important to highlight that the decision statistic (left-hand side of (40)) does not depend on a specific value of m (recall also that data distribution under H_0 does not depend on m) but on the variation range of m, namely, the set $\{1, \ldots, M\}$ (that defines the multiple hypothesis testing problem). In fact, the last maximization of (40) is over $m \in \{1, \ldots, M\}$ and allows for the estimation of the model order that, under H_0, is meaningless. As a consequence, threshold setting can be accomplished as for conventional binary

hypothesis tests. Finally, observe that $\widehat{\Lambda}_1(\boldsymbol{Z};m)$ is the Compressed Log-likelihood Ratio (CLR) assuming that the model order is $p_{r,m}+p_s$. Thus, for sufficiently large $\mathcal{T}$, it would be reasonable to consider (when possible) alternative decision statistics which share the same asymptotic behavior as the CLR. For instance, the CLR can be replaced by the decision statistics of the Rao or Wald test [37]. Another modification of (40) may consist in replacing $\widehat{\Lambda}_1(\boldsymbol{Z};m)$ with the decision statistics derived applying *ad hoc*[11] GLRT-based design procedures [18], [50]–[53]. Therefore, decision rule (40) represents a starting point for the design of detection architectures dealing with one null hypothesis and multiple alternative hypotheses.

In Appendix B, we show that (40) can also be considered as the result of the joint application of the Bayesian and ML framework after assigning a suitable prior to the number of parameters.

IV. Invariance Issues and CFAR Property

The design of architectures capable of guaranteeing the CFAR property is an issue of primary concern in radar (as well as in other application fields) since it allows for reliable target detection also in those situations where the interference (or unwanted components) spectral properties are unknown or highly variable. As a matter of fact, controlling and keeping low the number of false alarms is a precaution aimed at avoiding the disastrous effects that the latter may ensue. Thus, system engineers set detection thresholds in order to guarantee very small values for the P_{fa} [54]–[58]. Unfortunately, the CFAR property is not granted by a generic detection scheme and, hence, before claiming the CFARness for a given receiver, it must be proved that its decision statistic does not depend on the interference parameters under the null hypothesis. However, there exist some cases where the decision statistic is functionally invariant to those transformations that modify the nuisance parameters, which do not enter the decision process, and, at the same time, preserve the hypothesis testing problem. As a consequence, under the null hypothesis, such statistics follow a distribution that is independent of the nuisance parameters ensuring the CFAR property [50], [53], [59].

The above remarks suggest that it may be reasonable to look for decision rules that are invariant to nuisance parameters in the sense described above. To this end, we can invoke the *Principle of Invariance* [34], [60], whose key idea consists in finding a specific group of transformations, formed by a set $\mathcal{G}$ equipped with a binary composition operation $\circ$, that leaves unaltered the formal structure of the hypothesis testing problem (also inducing a group of transformations in the parameter space) and the family of distribution under each hypothesis. Then, data compression can be accomplished by resorting to the so-called *maximal invariant statistics* which organize data into distinguishable classes of equivalence with respect to the group of transformations wherein such statistics are constant. Now, given a group of transformations, every invariant test can be written in terms of the maximal invariant statistic whose distribution may depend only on a function of the parameters (induced maximal invariant). If the latter exists and is constant over Θ_0, then any invariant test guarantees the CFAR property with respect to the unknown nuisance parameters.

In what follows, we provide two propositions which allow us to state when (40) is invariant with respect to a given group of transformations and, possibly, enjoys the CFAR property.

Proposition 1: Let us assume that there exists a group of transformations $\mathcal{L}=(\mathcal{G},\circ)$, which acts on data through $l(\cdot)$ and leaves both the following binary hypothesis testing problems

$$\mathcal{P}_m:\begin{cases} H_0: & \boldsymbol{Z}\sim f_0(\boldsymbol{Z};\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s),\\ H_{1,m}: & \boldsymbol{Z}\sim f_{1,m}(\boldsymbol{Z};\boldsymbol{\theta}^m_{r,1},\boldsymbol{\theta}_s,p_{r,m}),\end{cases} \tag{42}$$

for all $m\in\{1,\ldots,M\}$ and the data distribution family unaltered, then the decision statistic of (40) is invariant with respect to $\mathcal{L}$.

Proof: Since each problem $\mathcal{P}_m$ is invariant with respect to $\mathcal{L}$ by definition, the CLR for the mth testing problem, namely $\widehat{\Lambda}_1(\boldsymbol{Z};m)$, is invariant to the same transformation group as shown in [34, Problem 6.27] and [61, Proposition 7.13], namely

$$\widehat{\Lambda}_1(l(\boldsymbol{Z});m)=\widehat{\Lambda}_1(\boldsymbol{Z};m),\quad \forall m\in\{1,\ldots,M\}. \tag{43}$$

As a consequence, we obtain that

$$\begin{aligned}&\max_{m\in\{1,\ldots,M\}}\left\{\widehat{\Lambda}_1(l(\boldsymbol{Z});m)-h(m)\right\}\\ &=\max_{m\in\{1,\ldots,M\}}\left\{\widehat{\Lambda}_1(\boldsymbol{Z};m)-h(m)\right\}.\end{aligned} \tag{44}$$

The last equality establishes the invariance of the left-hand side of (40) with respect to $\mathcal{L}$ and concludes the proof. ■

From a practical point of view, if, for all $m\in\{1,\ldots,M\}$, the generic $\mathcal{P}_m$ is invariant with respect to a subgroup $\mathcal{L}_m$ of a more general group, then it is possible to obtain $\mathcal{L}$ as the intersection (from the invariance standpoint) of the subgroups $\mathcal{L}_1,\ldots,\mathcal{L}_M$ [62]. As an example, consider a detection problem where the signal is assumed to belong to a subspace whose size, m say, is unknown and takes on values in $\{1,\ldots,M\}$. Thus, we can formulate this problem as in (3) and, given m, the linear group of transformations leaving unaltered the subproblem $\mathcal{P}_m$ can be readily obtained from [63]. It turns out that the group of transformations that does not alter (the canonical form [64] of) (3) is given by the group preserving the structure of the subspace for each size $m\in\{1,\ldots,M\}$, that, for the considered case, is the group of nonsingular upper triangular matrices. Moreover, observe that if (40) is invariant with respect to $\mathcal{L}$ then, by *Theorem 6.2.1* of [34], its decision statistic can be expressed as a function of the previously described maximal invariant statistics. Next, under the hypothesis of *Theorem 6.3.2* of [34], the distribution of a maximal invariant statistic depends on the induced maximal invariant which is denoted by $\xi(\boldsymbol{\theta})$ and the following proposition holds true.

Proposition 2: Let us assume that **Proposition 1** is valid and that

$$\forall\boldsymbol{\theta}\in\Theta_0:\ \xi(\boldsymbol{\theta})=c, \tag{45}$$

with $c\in\mathbb{R}$, then (40) ensures the CFAR property.

Proof: Since the left-hand side of (40) is invariant, it is a function of the maximal invariant statistic, whose distribution

[11] In the sense that they break the ML principle.

does not depend on the specific value of the parameter vector under H_0 but only on c. ■

V. Conclusions

In this paper, we have developed a new design framework based upon information theoretic criteria to address multiple hypothesis testing problems. Specifically, we have considered multiple (possibly nested) alternative hypotheses and only one null hypothesis. Such problems frequently arise in radar scenarios of practical interest. The proposed design procedure exploits the KLIC to come up with decision statistics that incorporate a penalized CLR, where the penalty term depends on the number of unknown parameters, which, in turn, is related to the specific alternative hypothesis under consideration. Interestingly, such a framework provides an information theoretic derivation for the LRT and GLRT and lays the foundation for the design of detection architectures capable of operating in the presence of multiple alternative hypotheses solving the limitations of the ML approach which exhibits an overestimation inclination when the hypotheses are nested. Finally, we have shown that under some specific conditions, decision schemes devised within this framework ensure the CFAR property.

In the second part of this work [35], we apply the proposed framework to several radar detection problems highlighting the practical value of the former. Future research tracks may encompass the analysis where the CLR is replaced by asymptotically equivalent statistics or the derivation of other decision rules based upon the joint ML and maximum a posteriori estimation procedure testing several priors for the number of parameters.

Appendix A
KLIC Based on EDPD

In this appendix, we derive (24) and (32) from (23) and (31), respectively. Let us start from (23), which implies that

$$E_{\Xi}[\log g_{1,\bar{m}}(\boldsymbol{z};\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s,p_{r,\bar{m}})] - E_{\Xi}[\log g_0(\boldsymbol{z};\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s)] \underset{H_0}{\overset{H_1}{\gtrless}} 0 \tag{46}$$

$$\Rightarrow \sum_{k=1}^{K}\log g_{1,\bar{m}}(\boldsymbol{z}_k;\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s,p_{r,\bar{m}}) - \sum_{k=1}^{K}\log g_0(\boldsymbol{z}_k;\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s) \underset{H_0}{\overset{H_1}{\gtrless}} 0. \tag{47}$$

The last equation can be rewritten to obtain (24).

As for (31), notice that this problem is equivalent to

$$\min_{\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s}\{-E_{\Xi}[\log g_0(\boldsymbol{z};\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s)]\} \underset{H_0}{\overset{H_1}{\gtrless}} \min_{\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s}\{-E_{\Xi}[\log g_{1,\bar{m}}(\boldsymbol{z};\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s,p_{r,\bar{m}})]\} \tag{48}$$

$$\Rightarrow \underbrace{\min_{\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s}\left[-\sum_{k=1}^{K}\log g_0(\boldsymbol{z}_k;\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s)\right]}_{P_1} \underset{H_0}{\overset{H_1}{\gtrless}} \underbrace{\min_{\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s}\left[-\sum_{k=1}^{K}\log g_{1,\bar{m}}(\boldsymbol{z}_k;\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s,p_{r,\bar{m}})\right]}_{P_2}. \tag{49}$$

Now, P_1 and P_2 are equal to

$$-\max_{\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s}\sum_{k=1}^{K}\log g_0(\boldsymbol{z}_k;\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s) \tag{50}$$

and

$$-\max_{\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s}\sum_{k=1}^{K}\log g_{1,\bar{m}}(\boldsymbol{z}_k;\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s,p_{r,\bar{m}}), \tag{51}$$

respectively, and, hence, the sought minimizers coincide with the MLEs of $\boldsymbol{\theta}_{r,1}^{\bar{m}}$, $\boldsymbol{\theta}_{r,0}$, and $\boldsymbol{\theta}_s$ (the latter under each hypothesis), namely

$$\left(\widehat{\boldsymbol{\theta}}_{r,1}^{\bar{m}},\widehat{\boldsymbol{\theta}}_{s,1}\right) = \arg\min_{\boldsymbol{\theta}_{r,1}^{\bar{m}},\boldsymbol{\theta}_s} D(\Xi\|g_{1,\bar{m}}), \tag{52}$$

$$\left(\widehat{\boldsymbol{\theta}}_{r,0},\widehat{\boldsymbol{\theta}}_{s,0}\right) = \arg\min_{\boldsymbol{\theta}_{r,0},\boldsymbol{\theta}_s} D(\Xi\|g_0). \tag{53}$$

With the above result in mind, (49) can be recast as

$$\sum_{k=1}^{K}\log g_{1,\bar{m}}(\boldsymbol{z}_k;\widehat{\boldsymbol{\theta}}_{r,1}^{\bar{m}},\widehat{\boldsymbol{\theta}}_{s,1},p_{r,\bar{m}}) - \sum_{k=1}^{K}\log g_0(\boldsymbol{z}_k;\widehat{\boldsymbol{\theta}}_{r,0},\widehat{\boldsymbol{\theta}}_{s,0}) \underset{H_0}{\overset{H_1}{\gtrless}} 0$$

which is equivalent to (32).

Appendix B
Alternative Derivation of (40)

Decision rule (40) can be viewed as the result of a suitable log-likelihood regularization, which aims at overcoming the previously described limitations of the ML approach in the case of nested models. To this end, we assume that the number of parameters of interest under H_1 is[12] a discrete random variable with a proper prior promoting low-dimensional models in order to mitigate the natural inclination of the ML approach to overestimate the model size.

With the above remarks in mind, a possible probability mass function for p_r is chosen as

$$\pi(p_r) = \frac{1}{A}e^{-g(p_r)}, \quad p_r \in \Omega_r, \tag{54}$$

[12]Notice that we are considering here a conventional binary hypothesis test.

where $g(\cdot)$ is a positive and increasing function[13] of p_r and A is a normalization constant such that

$$A = \sum_{p_r \in \Omega_r} e^{-g(p_r)} < +\infty. \tag{55}$$

For the ensuing development, it is important to note that if p_r is a continuous random variable, the joint PDF of $\boldsymbol{Z}$ and p_r (under H_1) can be recast as

$$f_1(\boldsymbol{Z}, p_r; \boldsymbol{\theta}_{p_r}, \boldsymbol{\theta}_{p_s}) = f_1(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}, \boldsymbol{\theta}_s | p_r) f(p_r), \tag{56}$$

where $f(p_r)$ is the PDF of p_r. However, since p_r is a discrete random variable, the joint PDF of $\boldsymbol{Z}$ and p_r is available only in the generalized sense exhibiting Dirac delta functions in correspondence of the values assumed by p_r. Thus, if we consider the following decision rule

$$\frac{\max\limits_{p_r \in \Omega_r} \max\limits_{\boldsymbol{\theta}_{r,1}, \boldsymbol{\theta}_s} f_1(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}, \boldsymbol{\theta}_s | p_r) \pi(p_r)}{\max\limits_{\boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s} f_0(\boldsymbol{Z}; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{57}$$

at the numerator, we attempt to maximize the multipliers of the Dirac delta functions, i.e., it only focuses on the lines where the probability masses are located. In this sense, the optimization at the numerator can be interpreted in terms of the joint ML and maximum a posteriori estimation procedure [23], [33]. The above architecture can be written as

$$\begin{aligned} &\max_{p_r \in \Omega_r} \max_{\boldsymbol{\theta}_{r,1}, \boldsymbol{\theta}_s} \log f_1(\boldsymbol{Z}; \boldsymbol{\theta}_{r,1}, \boldsymbol{\theta}_s | p_r) \pi(p_r) \\ &\quad - \max_{\boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s} \log f_0(\boldsymbol{Z}; \boldsymbol{\theta}_{r,0}, \boldsymbol{\theta}_s) \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \end{aligned} \tag{58}$$

After maximizing with respect to $\boldsymbol{\theta}_{r,1}$, $\boldsymbol{\theta}_{r,0}$, and $\boldsymbol{\theta}_s$, given p_r, (58) can be recast as

$$\max_{p_r \in \Omega_r} \left\{ \log \frac{f_1(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_{r,1}, \widehat{\boldsymbol{\theta}}_{s,1} | p_r)}{f_0(\boldsymbol{Z}; \widehat{\boldsymbol{\theta}}_{r,0}, \widehat{\boldsymbol{\theta}}_{s,0})} + \log \pi(p_r) \right\} \underset{H_0}{\overset{H_1}{\gtrless}} \eta. \tag{59}$$

The above decision scheme has the same structure as (40) and, when $g(p_r) = h(p_r)$, they are equivalent. Nevertheless, several alternatives can be used for $g(p_r)$ to come up with different penalization terms.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the anonymous Reviewers for their precious remarks as well as Prof. G. Ricci and Prof. A. De Maio for the interesting discussions.

[13]Notice that the expression of $g(\cdot)$ is assumed known. However, in practice, it can be chosen according to the application requirements and/or suitable performance criteria.

REFERENCES

[1] J. S. Bergin, P. M. Techau, W. L. Melvin, and J. R. Guerci, "GMTI STAP in target-rich environments: Site-specific analysis," in *Proc. IEEE Radar Conf.*, 2002, pp. 391–396.

[2] M. C. Wicks, W. L. Melvin, and P. Chen, "An efficient architecture for nonhomogeneity detection in space-time adaptive processing airborne early warning radar," in *Proc. Radar (Conf. Publ. No 449)*, Oct. 1997, pp. 295–299.

[3] R. S. Adve, T. B. Hale, and M. C. Wicks, "Transform domain localized processing using measured steering vectors and non-homogeneity detection," in *Proc. IEEE Radar Conf. Radar Next Millennium (Cat. No.99CH36249)*, Apr. 1999, pp. 285–290.

[4] L. Jiang and T. Wang, "Robust non-homogeneity detector based on reweighted adaptive power residue," *IET Radar, Sonar Navigation*, vol. 10, no. 8, pp. 1367–1374, 2016.

[5] M. Rangaswamy, "Non-homogeneity detector for Gaussian and non-Gaussian interference scenarios," in *Proc. Sensor Array Multichannel Signal Process. Workshop Proc.*, 2002, Aug. 2002, pp. 528–532.

[6] W. L. Melvin and J. A. Scheer, *Principles of Modern Radar: Advanced Techniques*, Edison, NJ, USA: S. Publishing, 2013.

[7] W. Liu, W. Xie, and Y. Wang, "Rao and wald tests for distributed targets detection with unknown signal steering," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1086–1089, Nov. 2013.

[8] J. Liu, H. Li, and B. Himed, "Persymmetric adaptive target detection with distributed MIMO radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 51, no. 1, pp. 372–382, Jan. 2015.

[9] K. Gerlach and M. J. Steiner, "Adaptive detection of range distributed targets," *IEEE Trans. Signal Process.*, vol. 47, no. 7, pp. 1844–1851, Jul. 1999.

[10] F. Bandiera, O. Besson, D. Orlando, G. Ricci, and L. L. Scharf, "GLRT-based direction detectors in homogeneous noise and subspace interference," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2386–2394, Jun. 2007.

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

[12] F. Bandiera and G. Ricci, "Adaptive detection and interference rejection of multiple point-like radar targets," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4510–4518, Dec. 2006.

[13] F. Bandiera, D. Orlando, and G. Ricci, "CFAR detection of extended and multiple point-like targets without assignment of secondary data," *IEEE Signal Process. Lett.*, vol. 13, no. 4, pp. 240–243, Apr. 2006.

[14] D. Adamy, *EW101: A. First Course in Electronic Warfare*, Norwood, MA, USA: A. House, 2001.

[15] A. Farina, "ECCM techniques," in *Radar Handbook*, M. I. Skolnik, Ed. New York: McGraw-Hill, 2008, ch. 24.

[16] F. Bandiera, A. Farina, D. Orlando, and G. Ricci, "Detection algorithms to discriminate between radar targets and ECM signals," *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 5984–5993, Dec. 2010.

[17] V. Carotenuto, C. Hao, D. Orlando, A. De Maio, and S. Iommelli, "Detection of multiple noise-like jammers for radar applications," in *Proc. 5th IEEE Int. Workshop Metrology AeroSpace*, Jun. 2018, pp. 328–333.

[18] L. Yan, P. Addabbo, C. Hao, D. Orlando, and A. Farina, "New ECCM techniques against noise-like and/or coherent interferers," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 2, pp. 1172–1188, Apr. 2020.

[19] J. Stuller, "Generalized likelihood signal resolution," *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 276–282, May 1975.

[20] J. Thomas and J. Wolf, "On the statistical detection problem for multiple signals," *IRE Trans. Inf. Theory*, vol. 8, no. 4, pp. 274–280, 1962.

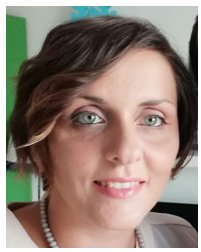
[21] S. Buzzi, M. Lops, and L. Venturino, "Track-before-Detect procedures for early detection of moving target from airborne radars," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 3, pp. 937–954, Jul. 2005.

[22] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference, A Practical Information-Theoretic Approach*, 2nd ed. New York, USA: Springer-Verlag, 2002.

[23] P. Stoica and Y. Selen, "Model-order selection: A. review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

[24] A. A. Neath and J. E. Cavanaugh, "The bayesian information criterion: Background, derivation, and applications," *WIREs Comput. Statist.*, vol. 4, no. 2, pp. 199–203, Mar. 2012.

[25] R. J. Bhansali and D. Y. Downham, "Some properties of the order of an autoregressive model selected by a generalization of akaike's FPE criterion," *Biometrika*, vol. 64, pp. 547–551, 1977.
[26] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.
[27] E. Fishler, M. Grosmann, and H. Messer, "Detection of signals by information theoretic criteria: General asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1027–1036, May 2002.
[28] H. L. Van Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*. New York: Wiley, 2002.
[29] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, Mar. 1951.
[30] E. Faro, G. Giunta, S. Han, D. Orlando, and L. Pallotta, "Subspace-based target detection in the presence of multiple alternative hypotheses," in *Proc. IEEE Radar Conf.*, 2020, pp. 1–6.
[31] S. M. Kay, "The multifamily likelihood ratio test for multiple signal model detection," *IEEE Signal Process. Lett.*, vol. 12, no. 5, pp. 369–371, May 2005.
[32] A. Van derVaart, *Asymptotic Statistics, Ser. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[33] A. Yeredor, "The joint MAP-ML criterion and its relation to ML and to extended least-squares," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3484–3492, Dec. 2000.
[34] E. Lehmann and J. Romano, *Testing Statistical Hypotheses, Ser. Springer Texts in Statistics*. New York, NY, USA: Springer, 2008.
[35] P. Addabbo, S. Han, F. Biondi, G. Giunta, and D. Orlando, "Adaptive radar detection in the presence of multiple alternative hypotheses using Kullback-Leibler information criterion-part II: Applications," *IEEE Trans. Signal Process.*, early access, doi: 10.1109/TSP.2021.3089440.
[36] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Basic Principles*. Raleigh, NC, USA: S. Publishing, 2010.
[37] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, vol. 2. Prentice Hall, 1998.
[38] L. Yan, C. Hao, P. Addabbo, D. Orlando, and A. Farina, "An improved adaptive radar detector based on two sets of training data," in *Proc. IEEE Radar Conf.*, Apr. 2019, pp. 1–6.
[39] P. K. Sen and J. M. Singer, *Large Sample Methods in Statistics: An Introduction With Applications*. New York: Springer US, 1993.
[40] Y. Selén, *Model Selection and Sparse Modeling*. Dept. Inf. Technol., Uppsala Univ., 2007.
[41] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 2012.
[42] S. Eguchia and J. B. Copas, "Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma," *J. Multivariate Anal.*, vol. 97, no. 9, pp. 2034–2040, 2006.
[43] Y. Chen, A. Wiesel, and A. O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4097–4107, Sep. 2011.
[44] E. Ollila and D. E. Tyler, "Regularized M-Estimators of scatter matrix," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 6059–6070, Nov. 2014.
[45] M. Steiner and K. Gerlach, "Fast converging adaptive processor or a structured covariance matrix," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 4, pp. 1115–1126, Oct. 2000.
[46] R. Nitzberg, "Application of maximum likelihood estimation of persymmetric covariance matrices to adaptive processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-16, no. 1, pp. 124–127, Jan. 1980.
[47] G. Foglia, C. Hao, G. Giunta, and D. Orlando, "Knowledge-aided adaptive detection in partially homogeneous clutter: Joint exploitation of persymmetry and symmetric spectrum," *Digit. Signal Process.*, vol. 67, no. *Supplement C.*, pp. 131–138, Aug. 2017.
[48] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1, upper saddle river, New Jersey: Prentice Hall, 1993.
[49] C. J. Geyer, "Stat 8112 lecture notes: The wald consistency theorem," *Univ. Minnesota, Sch. Statist.*, pp. 1–13, 2012.
[50] 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, 1992.
[51] Y. I. Abramovich and B. A. Johnson, "GLRT-based detection-estimation for undersampled training conditions," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3600–3612, Aug. 2008.
[52] C. Hao, D. Orlando, G. Foglia, and G. Giunta, "Knowledge-based adaptive detection: Joint exploitation of clutter and system symmetry properties," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1489–1493, Oct. 2016.
[53] F. Bandiera, D. Orlando, and G. Ricci, "Advanced radar detection schemes under mismatched signal models, synthesis lectures on signal processing No. 8," San Rafael, US: Morgan Claypool Publishers, 2009.
[54] A. Farina and F. A. Studer, "A review of CFAR detection techniques in radar systems," *Microw. J.*, vol. 29, no. 9, pp. 115–128, 1986.
[55] H. Rohling, "Radar CFAR thresholding in clutter and multiple target situations," *IEEE Trans. Aerosp. Electron. Syst.*, no. 4, pp. 608–621, Jul. 1983.
[56] M. Barkat, S. D. Himonas, and P. K. Varshney, "CFAR detection for multiple target situations," in *Proc. IEE Proc. F. (Radar Signal Process.)*, 1989, pp. 193–209.
[57] J. R. Roman, M. Rangaswamy, D. W. Davis, Q. Zhang, B. Himed, and J. H. Michels, "Parametric adaptive matched filter for airborne radar applications," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 2, pp. 677–692, Apr. 2000.
[58] F. Gini and M. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Process.*, vol. 82, no. 12, pp. 1847–1859, 2002.
[59] E. J. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 2, pp. 115–127, Mar. 1986.
[60] L. L. Scharf and C. Demeure, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis, Ser. Addison-Wesley Series in Electrical and Computer Engineering*. New Jersey: Addison-Wesley Publishing Company, 1991.
[61] M. Eaton, *Multivariate Statistics: A Vector Space Approach, Ser. Lecture Notes-Monograph Series*. New York: Institute of Mathematical Statistics, 1983.
[62] D. Robinson, S. Axler, F. Gehring, and P. Halmos, *A Course in the Theory of Groups, Ser. Graduate Texts in Mathematics*. New York, NY, USA: Springer, 1996.
[63] S. Bose and A. O. Steinhardt, "Adaptive array detection of uncertain rank one waveforms," *IEEE Trans. Signal Process.*, vol. 44, no. 11, pp. 2801–2809, Nov. 1996.
[64] D. Ciuonzo, A. De Maio, and D. Orlando, "A unifying framework for adaptive radar detection in homogeneous plus structured interference–Part I: On the maximal invariant statistic," *IEEE Trans. Signal Process.*, vol. 64, no. 11, pp. 2894–2906, Jun. 2016.



Pia Addabbo (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunication engineering, and the Ph.D. degree in information engineering from the Università degli Studi del Sannio, Benevento, Italy, in 2005, 2008, and 2012, respectively. She is currently a Researcher with the Giustino Fortunato University, Benevento, Italy. She is the coauthor of scientific publications in international journals and conferences. Her research interests include statistical signal processing applied to radar target recognition, global navigation satellite system reflectometry, and hyperspectral unmixing. She is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE ACCESS.

Sudan Han received the B.S. degree in information engineering and the M.S. and Ph.D. degrees in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2012, 2015, and 2019, respectively. She is currently an Assistant Researcher with the Defense Innovation Institute, Beijing, China. Her main research interests include clutter suppression, target detection, and radar signal processing.

Filippo Biondi (Member, IEEE) was born in L'Aquila, Italy, in 1974. He received the B.S. degree in information engineering and the M.S. degree in telecommunication engineering from Salento University, Lecce, Italy, in 2009 and 2011, respectively, and the Ph.D. degree in telecommunication engineering from the University of L'Aquila, L'Aquila, Italy, in 2015. His research interests include the development of signal processing and statistical signal processing techniques on radar equipment. In particular, he is working on synthetic aperture radar (SAR), interferometry, SAR-polarimetry, multichromatic analysis, and tomography, also using a convex optimization approach. He is also currently developing scientific research on orbital angular momentum for SAR 3D reconstraction and quantum radar, in particular quantum electrodynamics and quantum entanglement. From April 2010 to April 2011, he was the Guest Scientist with the Nation Research Council (ISSIA-CNR), Bari, Italy and the Italian Space Agency, Geodesy Space Center, Matera Italy. He is currently an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING JOURNAL.

Gaetano Giunta (Senior Member, IEEE) received the M.S. degree electronic engineering from the University of Pisa, Pisa, Italy and the Ph.D. degree in information and communication engineering from the University of Rome La Sapienza, Rome, Italy, in 1985 and 1990, respectively. During 1989–1990, he was also a Research Fellow with Signal Processing Laboratory, the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. In 1992, he became an Assistant Professor with the INFO-COM Department, University of Rome La Sapienza. In 1998, he taught signal processing with the University of Roma Tre, Rome, Italy. From 2001 to 2005, he was with the University of Roma Tre as an Associate Professor of telecommunications. Since 2005, he has been a Full Professor of telecommunications with the same University. His main research interests include signal and image processing for mobile communications and wireless networks. He is a Member of the IEEE Societies of Communications, Signal Processing, and Vehicular Technology. He was also a Reviewer for various IEEE transactions, IET (formerly IEE) proceedings, and EURASIP journals, and a TPC Member for various international conferences and symposia in the same fields.

Danilo Orlando (Senior Member, IEEE) was born in Gagliano del Capo, Italy, on August 9, 1978. He received the Dr. Eng. degree (with Hons.) in computer engineering and the Ph.D. degree (with maximum score) in information engineering from the University of Salento, Lecce, Italy, in 2004 and 2008, respectively. From 2007 to 2010, he was with the University of Cassino, 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 was a Visiting Scientist with the NATO Under sea Research Centre, La Spezia, Italy. From September 2011 to April 2015, he was with Elettronica SpA, engaged as a System Analyst in the field of electronic warfare. In May 2015, he joined Universitè degli Studi "Niccolò Cusano," Rome, Italy, where he is currently an Associate Professor. He has held visiting positions with the Department of Avionics and Systems of ENSICA (currently Institut Supérieur de l'Aéronautique et de l'Espace, ISAE), Toulouse, France, in 2007, and with Chinese Academy of Science, Beijing, China, from 2017 to 2020. He has authored or coauthored papers in about 130 scientific publications in international journals, conferences, and books. His main research focuses on statistical signal processing with more emphasis on adaptive detection and tracking of multiple targets in multisensor scenarios. He was the Senior Area Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING and is currently an Associate Editor for the IEEE OPEN JOURNAL ON SIGNAL PROCESSING, *EURASIP Journal on Advances in Signal Processing*, and *MDPI Remote Sensing*.