Joint Detection and Estimation: Optimum Tests and Applications

George V. Moustakides, *Senior Member, IEEE*, Guido H. Jajamovich, *Student Member, IEEE*, Ali Tajer, *Member, IEEE*, and Xiaodong Wang, *Fellow, IEEE*

***Abstract*—We consider a well-defined joint detection and parameter estimation problem. By combining the Bayesian formulation of the estimation subproblem with suitable constraints on the detection subproblem, we develop optimum one- and two-step test for the joint detection/estimation setup. The proposed combined strategies have the very desirable characteristic to allow for the trade-off between detection power and estimation quality. Our theoretical developments are, then, applied to the problems of retrospective changepoint detection and multiple-input multiple-output (MIMO) radar. In the former case, we are interested in detecting a change in the statistics of a set of available data and provide an estimate for the time of change, while in the latter in detecting a target and estimating its location. Intense simulations in the MIMO radar example demonstrate that by using jointly optimum schemes, we can experience significant improvement in estimation quality, as compared to generalized the likelihood ratio test or the test that treats the two subproblems separately, with only small sacrifices in detection power.**

***Index Terms*—Joint detection/estimation, multiple-input multiple-output (MIMO) radar, retrospective change detection.**

I. Introduction

THERE are important applications in practice where one is confronted with the problem of distinguishing between different hypotheses and, depending on the decision, to proceed and estimate a set of relevant parameters. Characteristic examples are detection and estimation of objects from images [1]; retrospective changepoint detection, where one desires to detect a change in statistics but also estimate the time of the change [2], [3]; defect detection from radiographies, where in addition to detecting presence of defects one would also like to find their position and shape [4]; finally, multiple-input multiple-output (MIMO) radar where we are interested in detecting the presence of a target and also estimate several target characteristics as position, speed, etc. All these applications clearly demand for detection and estimation strategies that address the two subproblems in a *jointly* optimum manner.

Manuscript received January 16, 2011; revised July 24, 2011; accepted November 16, 2011. Date of publication January 18, 2012; date of current version June 12, 2012.

G. V. Moustakides is with the Department of Electrical and Computer Engineering, University of Patras, 26500 Rion, Greece (e-mail: moustaki@upatras.gr).

G. H. Jajamovich and X. Wang are with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: guido@ee.columbia.edu; wangx@ee.columbia.edu).

A. Tajer is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: tajer@princeton.edu).

Communicated by E. Serpedin, Associate Editor for Signal Processing.

Digital Object Identifier 10.1109/TIT.2012.2184260

One could apply two straightforward approaches to deal with combined problems. The first consists in treating the two subproblems separately and use, in each case, the corresponding optimum technique. In particular, one can use the Neyman–Pearson optimum test for detection and the optimum Bayesian estimator for parameter estimation to solve the combined problem. As it is reported in [5] and we will also have the chance to verify with our analysis, treating each part separately with the optimum scheme does not necessarily result in optimum overall performance. The second methodology consists in using the well-known generalized likelihood ratio test (GLRT) which detects and estimates at the same time with the parameter estimation part relying on the maximum likelihood estimator. Even though GLRT is very popular, it lacks solid optimality properties, especially under a finite-sample-size setup [6]. We must also mention that both approaches lack versatility since they are not capable of emphasizing each subproblem according to the needs of the corresponding application.

In [1], [7]–[10], several methodologies are reported that have as common element the reduction of the joint detection/estimation problem into an equivalent multihypothesis setup by sampling the parameter space at a *finite* number of points. It is clear that such approaches lack optimality when the parameters assume a continuum of values since they constitute approximations to the original problem. Furthermore, it is not clear how one can incorporate popular costs, as mean square error (MSE) or mean absolute error, that constitute classical figures of merit for the estimation subproblem.

Surprisingly, there is *very* limited literature that deals with *optimum* solutions of the joint detection and estimation problem. Purely, Bayesian techniques are reported in [5] and [11] under a fixed sample size setup offering optimum detection and estimation structures for the joint problem.

Bayesian approaches that do not lead to jointly optimum detection/estimation schemes are reported in [12] and [13]. In particular, the first article deals with recursive signal estimation algorithms with parallel signal detection, while the second addresses communication problems and presents receiver structures of detector/estimator type. We should also mention [14], [15] where the Bayesian approach for the joint detection/estimation problem is analyzed under a sequential setup. Again, no optimum scheme is established.

Finally, a combination of Bayesian and Neyman–Pearson-like methodology is developed in [16] and [17]. Specifically, the error probabilities under the two hypotheses, used in the classical Neyman–Pearson approach, are replaced by estimation costs. Mimicking the Neyman–Pearson formulation and con-

straining the estimation cost under the nominal hypothesis while optimizing the corresponding cost under the alternative give rise to a number of interesting combined tests that can be used as alternatives to GLRT.[1]

It is clear that in order to properly formulate the joint problem, we must define a suitable performance measure that depends on both parts, that is, the detection and the estimation. Since each subproblem is characterized by its own figure of merit, in order to produce a unique performance measure, there are basically two possible directions one can follow. We can either combine the two performance measures, which is the logic behind the Bayesian approach [5], [11], or focus on one subproblem and attempt to optimize its efficiency while assuring satisfactory performance for the second subproblem through suitable constraints. This is the direction we intend to follow in this paper. In particular, we will define a cost function for the estimation part which will depend on both, the detector and the estimator. This performance measure will, then, be optimized with respect to the detector/estimator pair assuring, at the same time, the satisfactory performance of the detection part with appropriate constraints on the decision error probabilities. This idea will lead to two novel optimum tests that have no equivalent in [5], [11], and [16].

We would like to point out that the theory in [5], [11], [16], [17] as well as the one we are going to develop in our paper makes sense only when *both* subproblems constitute desired goals in our setup, that is, when we are interested in detecting *and* estimating. These results cannot provide optimum schemes for the case where one is interested only in detection and is forced to use parameter estimation due to presence of parameters that are regarded as nuisance.

Our paper is organized as follows. In Section II, we define the joint detection and estimation problem and propose two different optimal solutions. As a quick example, our results are, then, applied to the problem of retrospective change detection. In Section III, we make a thorough presentation of the MIMO radar problem under a joint detection and estimation formulation and use the results of the previous section in order to solve this problem optimally. Specifically, we develop closed-form expressions for all quantities that are needed to apply our theory and perform simulations to evaluate the performance of the optimum scheme, addressing also computational issues. Finally, in Section IV, we have our concluding remarks.

II. Optimum Joint Detection and Parameter Estimation

Let us define the problem of interest. Motivated by most applications mentioned in Section I, we limit ourselves to the binary hypothesis case with parameters present only under the alternative hypothesis. Suppose we are given an observation vector X for which we have the following two hypotheses:[2]

$$\mathsf{H}_0 : X \sim f_0(X)$$
$$\mathsf{H}_1 : X \sim f_1(X|\Theta),\ \Theta \sim \pi(\Theta)$$

where $f_0(X), f_1(X|\Theta), \pi(\Theta)$ are known probability density functions (pdfs) and "$\sim$" means "distributed according to." For simplicity, we will assume that the pdfs are well-behaved functions not having any nonzero probability masses over lower dimensional manifolds (this is the equivalent of X not having any "point masses," or its corresponding pdf not containing any delta functions, when X is scalar). As we can see, we assume that under H_0 we know the pdf of X completely, whereas under H_1 the pdf of X contains a collection of random parameters Θ for which we have some prior available pdf $\pi(\Theta)$. The goal is to develop a mechanism that distinguishes between $\mathsf{H}_0, \mathsf{H}_1$, and, furthermore, every time it decides in favor of H_1, it provides an estimate $\hat{\Theta}$ for Θ. Our combined detection/estimation scheme is, therefore, comprised of a randomized test $\{\delta_0(X), \delta_1(X)\}$ with $\delta_i(X)$ denoting the randomization probability for deciding in favor of H_i in a random decision game (like coin flipping using an *unfair* coin), and a function $\hat{\Theta}(X)$ that provides the necessary parameter estimates. Clearly, $\delta_i(X) \geq 0$ and $\delta_0(X) + \delta_1(X) = 1$.

Let us recall, very briefly, the optimum detection and estimation theory when the two subproblems are considered separately.

Neyman–Pearson Hypothesis Testing: Fix a level $\alpha \in (0,1)$; if D denotes our decision, then we are interested in selecting a test [namely the randomization probabilities $\delta_i(X)$] so that the detection probability $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_1) = \iint \delta_1(X) f_1(X|\Theta)\pi(\Theta) d\Theta\, dX$ is maximized subject to the false alarm constraint $\mathsf{P}_0(\mathsf{D} = \mathsf{H}_1) = \int \delta_1(X) f_0(X) dX \leq \alpha$. Equivalently, the previous maximization can be replaced by the minimization of the probability of miss $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_0) = \iint \delta_0(X) f_1(X|\Theta)\pi(\Theta) d\Theta\, dX$. It is clear that P_i refers to the probability measure induced by the corresponding pdf $f_i(\cdot)$. The optimum detection scheme is the well-celebrated likelihood ratio test [18, pp. 22–25], which takes the following form for our specific setup:

$$\mathcal{L}(X) = \frac{f_1(X)}{f_0(X)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtrless}} \gamma_{\mathrm{NP}} \tag{1}$$

where

$$f_1(X) = \int f_1(X|\Theta)\pi(\Theta)\, d\Theta \tag{2}$$

is the corresponding marginal density of X. In other words, we decide H_1 whenever the likelihood ratio $\mathcal{L}(X)$ exceeds the threshold γ_{NP}, H_0 whenever it falls below the threshold, and randomize with a probability p when the likelihood ratio is equal to the threshold. The threshold γ_{NP} and the probability p are selected to satisfy the false alarm constraint with equality. The randomization probabilities $\delta_0^{\mathrm{NP}}(X), \delta_1^{\mathrm{NP}}(X)$ corresponding to the Neyman–Pearson test are given by

$$\delta_0^{\mathrm{NP}}(X) = \mathbb{1}_{\{\frac{f_1(X)}{f_0(X)} < \gamma_{\mathrm{NP}}\}} + (1-p)\mathbb{1}_{\{\frac{f_1(X)}{f_0(X)} = \gamma_{\mathrm{NP}}\}}$$
$$\delta_1^{\mathrm{NP}}(X) = \mathbb{1}_{\{\frac{f_1(X)}{f_0(X)} > \gamma_{\mathrm{NP}}\}} + p\mathbb{1}_{\{\frac{f_1(X)}{f_0(X)} = \gamma_{\mathrm{NP}}\}} \tag{3}$$

where $\mathbb{1}_\Gamma$ denotes the indicator function of the set Γ.

Bayesian Parameter Estimation: Suppose we know *with certainty* that the observation vector X comes from hypothesis H_1;

[1] In [17], for completeness, there is also a very brief presentation of the theoretical developments contained in the current article.

[2] Throughout our paper, scalar quantities are denoted with lower case, vectors with upper case, and matrices with bold-face upper case letters. We also use calligraphic letters to denote certain important quantities and a special letter style for the two hypotheses, decisions, probability measures, and their corresponding expectations.

then, we are interested in providing an estimate $\hat{\Theta}(X)$ for the parameters Θ. We measure the quality of our estimate with the help of a cost function $c(\hat{\Theta},\Theta) \geq 0$. We would like to select the optimum estimator in order to minimize the average cost $\mathsf{E}_1[c(\hat{\Theta}(X),\Theta)] = \iint c(\hat{\Theta}(X),\Theta) f_1(X|\Theta)\pi(\Theta) d\Theta\, dX$ where, as we can see, expectation is with respect to X and Θ.

From [18, p. 142], if $\mathcal{C}(U|X)$ denotes the posterior cost function

$$\mathcal{C}(U|X) = \mathsf{E}_1[c(U,\Theta)|X] = \frac{\int c(U,\Theta) f_1(X|\Theta)\pi(\Theta)\, d\Theta}{\int f_1(X|\Theta)\pi(\Theta)\, d\Theta} = \frac{\int c(U,\Theta) f_1(X|\Theta)\pi(\Theta)\, d\Theta}{f_1(X)} \quad (4)$$

where expectation is with respect to Θ for given X, then the optimum Bayesian estimator is the following minimizer:

$$\hat{\Theta}_o(X) = \arg\min_U \mathcal{C}(U|X). \quad (5)$$

In other words, for every given X, $\hat{\Theta}_o(X)$ is the value of U that minimizes $\mathcal{C}(U|X)$, with the latter considered as a function of U only. As it is the usual practice in most textbooks, e.g., [18, p. 142], [19, p. 117], in order to avoid the definition of ϵ-optimum estimators, or sequence of estimators with performance converging to the optimum, complicating unnecessarily our presentation, we will assume that $\mathcal{C}(U|X)$, for every value of X, always exhibits a minimum over U which is attained at $U = \hat{\Theta}_o(X)$, namely the Bayesian estimator. The latter is formally expressed with (5). Finally, we denote the optimum posterior cost as $\mathcal{C}_o(X)$, that is

$$\mathcal{C}_o(X) = \min_U \mathcal{C}(U|X) = \mathcal{C}(\hat{\Theta}_o(X)|X). \quad (6)$$

This quantity will play a very important role in the development of our theory as it constitutes a genuine quality index for the estimate $\hat{\Theta}_o(X)$.

Let us now consider the combined problem. We recall that the hypothesis testing part distinguishes between H_0 and H_1. As we have seen, the Neyman–Pearson approach provides the best possible detection structure for controlling and optimizing the corresponding decision error probabilities. However, with a decision mechanism that focuses on the decision errors, we cannot necessarily guarantee satisfactory performance for the estimation part. Consequently, as we argued in Section I, the detection part cannot be treated independently from the estimation part. Following this rationale, we propose two possible approaches involving one- and two-step schemes that differ in the number of decision mechanisms they incorporate and the way they combine the notion of *reliable estimate* with the detection subproblem.

A. One-Step Tests

Let us begin our analysis by introducing a proper performance measure for the estimation subproblem. Following the Bayesian approach, we assume the existence of the cost function $c(\hat{\Theta},\Theta) \geq 0$. Computing the average cost that will play the role of our performance measure is not as straightforward as in the pure estimation problem case and requires some consideration. Note that an estimate $\hat{\Theta}(X)$ is provided only when we decide in favor of H_1. On the other hand, averaging of $c(\hat{\Theta},\Theta)$ makes sense only under the alternative hypothesis H_1, since under the nominal H_0 there is no true parameter Θ. Consequently, we propose the following performance criterion:[3]

$$\mathcal{J}(\delta_1,\hat{\Theta}) = \mathsf{E}_1[c(\hat{\Theta}(X),\Theta)|\mathsf{D}=\mathsf{H}_1] = \frac{\iint \delta_1(X) c(\hat{\Theta}(X),\Theta) f_1(X|\Theta)\pi(\Theta) d\Theta\, dX}{\iint \delta_1(X) f_1(X|\Theta)\pi(\Theta) d\Theta\, dX} \quad (7)$$

where, as we can see, expectation is with respect to X and Θ. Using (2) and (4), we can immediately obtain a more convenient expression for our conditional average estimation cost, namely

$$\mathcal{J}(\delta_1,\hat{\Theta}) = \frac{\int \delta_1(X)\mathcal{C}(\hat{\Theta}(X)|X) f_1(X) dX}{\int \delta_1(X) f_1(X) dX}. \quad (8)$$

We realize that in our criterion, unlike the classical estimation problem, the estimation performance depends on *both* the estimator (through $\hat{\Theta}(X)$) *and* the detector (through $\delta_1(X)$) and we compute the average cost over the event $\{\mathsf{D}=\mathsf{H}_1\}$, which is the only case an estimate is available. Consequently, optimizing this criterion will result in the optimum detector/estimator pair.

One would immediately argue that the measure in (7), even though it takes into account both parts, its optimization does not guarantee, in any sense, the satisfactory performance of the detection subproblem. This is indeed the case but, as we mentioned in Section I, by imposing proper constraints on the two error probabilities of the detection part we can easily overcome this crucial weakness. Specifically, we can impose the familiar false alarm constraint $\mathsf{P}_0(\mathsf{D}=\mathsf{H}_1) \leq \alpha$ but *also* a constraint on the probability of miss $\mathsf{P}_1(\mathsf{D}=\mathsf{H}_0) \leq \beta$ where $\alpha,\beta \in (0,1)$. With these two constraints, we have complete control over the decision mechanism and therefore, now, it makes sense to attempt to minimize the conditional average estimation cost $\mathcal{J}(\delta_1,\hat{\Theta})$ over the decision rule *and* the estimator. Consequently, we propose the following constrained optimization problem:

$$\inf_{\delta_1,\hat{\Theta}} \mathcal{J}(\delta_1,\hat{\Theta}) = \inf_{\delta_1,\hat{\Theta}} \frac{\int \delta_1(X)\mathcal{C}(\hat{\Theta}(X)|X) f_1(X) dX}{\int \delta_1(X) f_1(X) dX} \quad (9)$$

subject to

$$\begin{aligned} \mathsf{P}_0(\mathsf{D}=\mathsf{H}_1) &= \int \delta_1(X) f_0(X) dX \leq \alpha \\ \mathsf{P}_1(\mathsf{D}=\mathsf{H}_0) &= \iint \delta_0(X) f_1(X|\Theta)\pi(\Theta) d\Theta\, dX \\ &= \int \delta_0(X) f_1(X) dX \leq \beta. \end{aligned} \quad (10)$$

Note that the two constraints guarantee satisfactory performance for the detection part and, by minimizing our criterion, we can enjoy the best possible performance in the estimation part. We would like to emphasize once more that even though we optimize the estimation performance, this is achieved by simultaneously optimizing over both parts, the detector and the estimator.

[3]Since $\delta_0(X) + \delta_1(X) = 1$, by specifying $\delta_1(X)$ we define, completely, the decision rule.

Let us carry out the desired optimization gradually based on the following equality:

$$\min_{\delta_1,\hat{\Theta}} \mathcal{J}(\delta_1,\hat{\Theta}) = \min_{\delta_1}\left\{\min_{\hat{\Theta}} \mathcal{J}(\delta_1,\hat{\Theta})\right\}. \tag{11}$$

In other words, we first fix the decision rule $\delta_1(X)$ (and hence $\delta_0(X)$ as well) and optimize $\mathcal{J}(\delta_1,\hat{\Theta})$ with respect to the estimator $\hat{\Theta}(X)$. The resulting outcome will depend on $\delta_1(X)$ and it will be further optimized with respect to this quantity over decision rules that satisfy the two constraints in (10). For the first minimization, we have the following lemma that provides the solution to this problem.

Lemma 1: Let $\varphi(X) \geq 0$ be any scalar nonnegative function; then, the following functional of $\hat{\Theta}(X)$:

$$\mathcal{D}(\hat{\Theta}) = \frac{\int \varphi(X)\mathcal{C}(\hat{\Theta}(X)|X)f_1(X)dX}{\int \varphi(X)f_1(X)dX} \tag{12}$$

is minimized when $\hat{\Theta}(X)$ is the optimum Bayesian estimator $\hat{\Theta}_o(X)$ defined in (5).

Proof: The proof is simple. We can write

$$\begin{aligned}\mathcal{D}(\hat{\Theta}) &= \frac{\int \varphi(X)\mathcal{C}(\hat{\Theta}(X)|X)f_1(X)dX}{\int \varphi(X)f_1(X)dX}\\ &\geq \frac{\int \varphi(X)\min_U \mathcal{C}(U|X)f_1(X)dX}{\int \varphi(X)f_1(X)dX}\\ &= \frac{\int \varphi(X)\mathcal{C}(\hat{\Theta}_o(X)|X)f_1(X)dX}{\int \varphi(X)f_1(X)dX}\\ &= \frac{\int \varphi(X)\mathcal{C}_o(X)f_1(X)dX}{\int \varphi(X)f_1(X)dX}\end{aligned} \tag{13}$$

where for the last two equalities we used (6). ■

Lemma 1 is a very interesting result because it demonstrates an extended optimality property of the classical Bayesian estimator. In particular, by selecting $\varphi(X) = \delta_1(X)$, we can conclude that $\hat{\Theta}_o(X)$ continues to be optimum even if estimation is dictated by a decision mechanism and not performed over all data X, as is the usual practice with Bayesian estimation. Consequently, we can now fix our estimator to the Bayesian estimator $\hat{\Theta}_o(X)$ and for the resulting performance we have

$$\begin{aligned}\tilde{\mathcal{J}}(\delta_1) &= \min_{\hat{\Theta}} \mathcal{J}(\delta_1,\hat{\Theta}) = \mathcal{J}(\delta_1,\hat{\Theta}_o)\\ &= \frac{\int \delta_1(X)\mathcal{C}_o(X)f_1(X)dX}{\int \delta_1(X)f_1(X)dX}.\end{aligned} \tag{14}$$

From (11), we can now see that

$$\min_{\delta_1,\hat{\Theta}} \mathcal{J}(\delta_1,\hat{\Theta}) = \min_{\delta_1} \tilde{\mathcal{J}}(\delta_1) = \min_{\delta_1} \frac{\int \delta_1(X)\mathcal{C}_o(X)f_1(X)dX}{\int \delta_1(X)f_1(X)dX}. \tag{15}$$

In other words, we need to minimize the resulting $\tilde{\mathcal{J}}(\delta_1)$ over $\delta_1(X)$ but with the detection structure $\delta_1(X)$ and $\delta_0(X) = 1 - \delta_1(X)$ satisfying the two constraints in (10). Before addressing this last step of our optimization problem, we need to make some remarks.

Remark 1: We recall that in our setup we have the two error probability constraints defined in (10). By fixing the false alarm probability to α, the probability of miss is minimized by the Neyman–Pearson test; call this minimum value $\beta_{\mathrm{NP}}(\alpha)$. Since no test, with false alarm probability not exceeding α, can have a probability of miss which is smaller than $\beta_{\mathrm{NP}}(\alpha)$, this suggests that in our constraint on the probability of miss, β must be selected to satisfy $\beta \geq \beta_{\mathrm{NP}}(\alpha)$. Consequently, if we select β to strictly exceed $\beta_{\mathrm{NP}}(\alpha)$, we are in fact allowing the reduction, in a completely controlled manner, of the detection power as compared to the Neyman–Pearson test, in the hope to use this extra space to improve the estimation quality. In fact, the goal is to make the best possible use of this flexibility since we are looking for the detector/estimator pair that will optimize this improvement.

Remark 2: The proposed setup, in a sense, is interested in detections *only if* they result in reliable estimates. In other words, with this formulation, the null hypothesis is not distinguished from the alternative if the latter cannot provide reliable parameter estimates. We should point out that there are applications where this setup makes sense. For example, segmentation problems in images, where one is interested in identifying objects in an image is such a possibility. Indeed, in this application, a simple object detection has no practical meaning if it is not accompanied by object boundary specification. Of course, there are other applications, as, for example, MIMO radar where, as we will see in the next section, this formulation is not suitable and where detection makes sense even if it is not followed by reliable estimation.

Let us now present our first theorem that gives the optimum solution for the joint detection/estimation problem introduced earlier.

Theorem 1: Let $0 < \alpha < 1$ and $\beta_{\mathrm{NP}}(\alpha) \leq \beta < 1$ with $\beta_{\mathrm{NP}}(\alpha)$ denoting the probability of miss of the Neyman–Pearson test when α is the false alarm probability level. Let $\lambda_o \geq 0$ be the solution of the equation

$$\mathsf{P}_1\left(\lambda_o \geq \mathcal{C}_o(X)\right) = 1 - \beta \tag{16}$$

where $\mathcal{C}_o(X)$ is defined in (6). Then, the optimum detector/estimator pair that minimizes (9) under the two constraints in (10) is comprised of the Bayesian estimator $\hat{\Theta}_o(X)$ defined in (5) for the estimation part, while the corresponding decision rule is given by

$$\mathcal{C}_o(X) \underset{H_0}{\overset{H_1}{\lessgtr}} \lambda_o, \quad \text{if } \alpha \geq \mathsf{P}_0\left(\lambda_o \geq \mathcal{C}_o(X)\right) \tag{17}$$

$$\mathcal{L}(X)[\lambda - \mathcal{C}_o(X)] \underset{H_0}{\overset{H_1}{\lessgtr}} \gamma, \quad \text{if } \alpha < \mathsf{P}_0\left(\lambda_o \geq \mathcal{C}_o(X)\right) \tag{18}$$

where in (18), λ and γ are selected so that the two error probability constraints are satisfied with equality.

Proof: The proof is presented in the Appendix. ■

From (17) and (18), we deduce that the detector component in the optimum pair takes into account the estimation part through $\mathcal{C}_o(X)$ which constitutes a quality index for the estimate $\hat{\Theta}_o(X)$. If this index is sufficiently large, then the estimate is considered unreliable and the test, in both cases, decides in favor of H_0 (since in this approach, as we argued in Remark 2, unreliable estimates are not distinguished from the null hypothesis). In particular, in (18), this decision may occur even if, in the classical Neyman–Pearson test, the likelihood ratio $\mathcal{L}(X)$ exceeds the threshold $\gamma_{\rm NP}$, that is, decides in favor of H_1.

Summarizing, our first optimum combined test consists in applying (17) or (18) to decide between the two hypotheses and every time we make a decision in favor of H_1 we use $\hat{\Theta}_o(X)$ defined in (5) to provide the optimum parameter estimate.

B. Two-Step Tests

In the previous setup, our decision was between H_0 and H_1 and we were sacrificing detection power to improve estimation. However, in many applications, giving up part of the detection capacity may be regarded as undesirable. For example, in MIMO radar, it is still helpful to detect a target even if we cannot reliably estimate its parameters.

It is possible to preserve the detection power and at the same time ameliorate the estimation performance if we follow a slightly different approach that involves *two-step mechanisms*. Specifically, we propose the use of an initial detection strategy that distinguishes between H_0 and H_1; whenever we decide in favor of H_1 then, at a second step, we compute the estimate $\hat{\Theta}(X)$ and employ a *second test* that decides whether the estimate is *reliable* or *unreliable*, denoted as $\mathsf{H}_{1\rm r}$ and $\mathsf{H}_{1\rm u}$, respectively. Consequently, we propose to make *three* different decisions $\mathsf{H}_0, \mathsf{H}_{1\rm r}$, and $\mathsf{H}_{1\rm u}$ with the union of the last two corresponding to hypothesis H_1. As we can see, estimate $\hat{\Theta}(X)$ is trustworthy only when we decide in favor of $\mathsf{H}_{1\rm r}$, but we have detection even if we discard the estimate as unreliable, that is, we decide $\mathsf{H}_{1\rm u}$.

For the first test, we use our familiar randomization probabilities $\{\delta_0(X), \delta_1(X)\}$, while for the second we employ a new pair $\{q_{1\rm r}(X), q_{1\rm u}(X)\}$. The latter functions are the randomization probabilities needed to decide between reliable/unreliable estimation *given* that the first test decided in favor of H_1. Therefore, we have $q_{1\rm r}(X), q_{1\rm u}(X) \geq 0$ and $q_{1\rm r}(X) + q_{1\rm u}(X) = 1$. For every combination of the four randomization probabilities we define, similarly to the previous section, the corresponding average conditional cost for the estimator $\hat{\Theta}(X)$, namely[4]

$$\begin{aligned}\mathcal{J}(\delta_1, q_{1\rm r}, \hat{\Theta}) &= \mathsf{E}_1[c(\hat{\Theta}(X), \Theta)|\mathsf{D} = \mathsf{H}_{1\rm r}] \\ &= \frac{\iint \delta_1(X) q_{1\rm r}(X) c(\hat{\Theta}(X), \Theta) f_1(X|\Theta)\pi(\Theta) d\Theta dX}{\iint \delta_1(X) q_{1\rm r}(X) f_1(X|\Theta)\pi(\Theta) d\Theta dX} \\ &= \frac{\int \delta_1(X) q_{1\rm r}(X) \mathcal{C}(\hat{\Theta}(X)|X) f_1(X) dX}{\int \delta_1(X) q_{1\rm r}(X) f_1(X) dX} \end{aligned} \tag{19}$$

[4]Similarly to the first decision mechanism, in the second, since $q_{1\rm r}(X) + q_{1\rm u}(X) = 1$, we only need to specify $q_{1\rm r}(X)$.

where for the last equality, again, we used (2) and (4). As we can see, we now condition on the event $\{\mathsf{D} = \mathsf{H}_{1\rm r}\}$ since this is the only case when the estimate $\hat{\Theta}(X)$ is accepted. We also note that, for given X, the probability to decide in favor of $\mathsf{H}_{1\rm r}$ is $\delta_1(X) q_{1\rm r}(X)$ because we must decide in favor of H_1 in the first step (with probability $\delta_1(X)$) and for $\mathsf{H}_{1\rm r}$ in the second (with probability $q_{1\rm r}(X)$).

In the first step, we would like to adopt the best possible detector to select between H_0 and H_1. We follow the classical Neyman–Pearson approach and impose the false alarm probability constraint $\mathsf{P}_0(\mathsf{D} = \mathsf{H}_1) \leq \alpha$, while we minimize the probability of miss $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_0)$. This leads to the Neyman–Pearson test defined in (1) with corresponding randomization probabilities $\delta_0^{\rm NP}(X)$, $\delta_1^{\rm NP}(X)$ given in (3).

Having identified the first, let us proceed to the second step of our detection/estimation mechanism that involves parameter estimation and a second test that labels the estimate as reliable/unreliable. Consider the conditional average estimation cost

$$\mathcal{J}(\delta_1^{\rm NP}, q_{1\rm r}, \hat{\Theta}) = \frac{\int \delta_1^{\rm NP}(X) q_{1\rm r}(X) \mathcal{C}(\hat{\Theta}(X)|X) f_1(X) dX}{\int \delta_1^{\rm NP}(X) q_{1\rm r}(X) f_1(X) dX} \tag{20}$$

which, of course, we would like to minimize with respect to the estimator $\hat{\Theta}$ and the second decision mechanism $q_{1\rm r}$. Note, however, that, in addition to this minimization, we are also interested in generating as many "reliable estimates" as possible when applying the second decision rule. These two goals are conflicting; therefore, we adopt a Neyman–Pearson-like approach in order to come up with an optimum scheme. In other words, we constrain one quantity and optimize the other.

To find a suitable constraint, we note that since $q_{1\rm r}(X) \leq 1$, the probability $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_{1\rm r})$ of deciding in favor of $\mathsf{H}_{1\rm r}$ (reliable estimate) satisfies

$$\begin{aligned}\mathsf{P}_1(\mathsf{D} = \mathsf{H}_{1\rm r}) &= \int \delta_1^{\rm NP}(X) q_{1\rm r}(X) f_1(X) dX \\ &\leq \int \delta_1^{\rm NP}(X) f_1(X) dX = \mathsf{P}_1(\mathsf{D} = \mathsf{H}_1) = 1 - \beta_{\rm NP}(\alpha).\end{aligned} \tag{21}$$

In other words, this probability is upper bounded by the detection probability $1 - \beta_{\rm NP}(\alpha)$ of the Neyman–Pearson test where $\beta_{\rm NP}(\alpha)$ denotes the corresponding minimum miss probability. The last inequality expresses the obvious fact that, only a portion of our initial decisions in favor of H_1 of the first step, will provide reliable estimates in the second step. Actually, it is this portion we intend to control by imposing the inequality $1 - \beta \leq \mathsf{P}_1(\mathsf{D} = \mathsf{H}_{1\rm r})$ where, because of (21), we need, again, to select $1 > \beta \geq \beta_{\rm NP}(\alpha)$. Summarizing, we are interested in solving the following constrained optimization problem:

$$\begin{aligned}&\min_{q_{1\rm r}, \hat{\Theta}} \mathcal{J}(\delta_1^{\rm NP}, q_{1\rm r}, \hat{\Theta}) \\ &\qquad = \min_{q_{1\rm r}, \hat{\Theta}} \frac{\int \delta_1^{\rm NP}(X) q_{1\rm r}(X) \mathcal{C}(\hat{\Theta}(X)|X) f_1(X) dX}{\int \delta_1^{\rm NP}(X) q_{1\rm r}(X) f_1(X) dX}\end{aligned} \tag{22}$$

subject to

$$1 - \beta \leq \mathsf{P}_1(\mathsf{D} = \mathsf{H}_{1\rm r}) = \int \delta_1^{\rm NP}(X) q_{1\rm r}(X) f_1(X) dX. \tag{23}$$

The constraint in (23) expresses our desire that at least a fraction of

$$\frac{1-\beta}{1-\beta_{\mathrm{NP}}(\alpha)} \leq \frac{\mathsf{P}_1(\mathsf{D}=\mathsf{H}_{1\mathrm{r}})}{\mathsf{P}_1(\mathsf{D}=\mathsf{H}_1)} \leq 1 \tag{24}$$

of the initial decisions in favor of H_1 must provide reliable estimates. Subject to this constraint, the goal is to obtain the best possible estimation performance by minimizing the conditional average estimation cost in (22) over the estimator and the second decision mechanism.

As in the previous test, we can perform this optimization in steps. We first fix $q_{1\mathrm{r}}(X)$; then, from Lemma 1, by selecting $\varphi(X) = \delta_1^{\mathrm{NP}}(X) q_{1\mathrm{r}}(X)$, we conclude that this criterion is minimized when $\hat{\Theta}(X) = \hat{\Theta}_o(X)$, that is, again with the optimum Bayes estimator defined in (5). Call

$$\begin{aligned}\tilde{\mathcal{J}}(q_{1\mathrm{r}}) &= \mathcal{J}(\delta_1^{\mathrm{NP}}, q_{1\mathrm{r}}, \hat{\Theta}_o) \\ &= \frac{\int \delta_1^{\mathrm{NP}}(X) q_{1\mathrm{r}}(X) \mathcal{C}_o(X) f_1(X) dX}{\int \delta_1^{\mathrm{NP}}(X) q_{1\mathrm{r}}(X) f_1(X) dX}\end{aligned} \tag{25}$$

the corresponding performance. The final step is to minimize $\tilde{\mathcal{J}}(q_{1\mathrm{r}})$ by selecting properly $q_{1\mathrm{r}}(X)$ over the class of decision mechanism that satisfy the constraint in (23) or (24). The solution to this constrained optimization problem is given in the next lemma.

Lemma 2: Let $1 > \beta \geq \beta_{\mathrm{NP}}(\alpha)$; then, the decision mechanism that minimizes $\tilde{\mathcal{J}}(q_{1\mathrm{r}})$ defined in (25) subject to the constraint in (23) is given by

$$\mathcal{C}_o(X) \underset{\mathsf{H}_{1\mathrm{u}}}{\overset{\mathsf{H}_{1\mathrm{r}}}{\gtrless}} \lambda \tag{26}$$

where λ is selected to satisfy (23) with equality and $\mathcal{C}_o(X)$ is defined in (6).

Proof: The proof goes along the same lines of the proof of Theorem 1. Since it presents no particular difficulty, it is omitted. ■

As in the previous section, $\mathcal{C}_o(X)$ constitutes a quality index for the estimate $\hat{\Theta}_o(X)$. With Lemma 2, we end up with the very plausible decision rule of accepting $\hat{\Theta}_o(X)$ as reliable whenever this index falls below some threshold λ, while the estimate is discarded as unreliable whenever the same quantity exceeds the threshold.

Summarizing our second optimum detection/estimation scheme, we first use the Neyman–Pearson test (1) to decide between H_0 and H_1. Whenever we decide in favor of H_1, we compute the estimate $\hat{\Theta}_o(X)$ from (5) and its corresponding quality index $\mathcal{C}_o(X)$ from (6); then, we use the test in (26) to characterize the estimate as reliable/unreliable.

Remark 3: In both our tests, we have the parameter β which controls the level of the probability of miss. Of course, $\beta \geq \beta_{\mathrm{NP}}(\alpha)$ since the miss probability delivered by the Neyman–Pearson test is the smallest possible (for given false alarm rate α). If we select $\beta = \beta_{\mathrm{NP}}(\alpha)$, then both our tests reduce to the case where we treat the two subproblems separately and apply the corresponding optimum scheme in each subproblem, namely, the Neyman–Pearson test to decide between the two hypotheses and the optimum Bayesian estimator to provide the necessary parameter estimate (every time the detector decides in favor of H_1). This latter test constitutes, of course, a straightforward selection for solving the joint detection/estimation problem and, as we pointed out, it is simply a limiting case of our more general schemes. The advantage offered by our methodology is the flexibility to trade detection power for estimation accuracy by just controlling this parameter β. Such flexibility does not exist when we treat the two subproblems separately or if we solve the joint problem by applying the GLRT. Furthermore, in the next section, we will have the opportunity to confirm that with the proposed detection/estimation structures, we can enjoy substantial gains in estimation quality as compared to GLRT and the test that treats the two problems separately. This will be achieved with only mild or even minor sacrifices in detection power.

C. MSE Cost and Uniform Prior

If we call $\mathcal{L}(X|\Theta) = \frac{f_1(X|\Theta)}{f_0(X)}$ the conditional likelihood ratio, then all quantities entering in the two tests can be expressed with the help of $\mathcal{L}(X|\Theta)$ and the prior probability $\pi(\Theta)$. We start with the likelihood ratio which is part of both tests and observe that it can be written as

$$\mathcal{L}(X) = \frac{f_1(X)}{f_0(X)} = \int \mathcal{L}(X|\Theta)\pi(\Theta)d\Theta. \tag{27}$$

From (4), we can see that the posterior cost $\mathcal{C}(U|X)$ can be computed as

$$\mathcal{C}(U|X) = \frac{\int c(U,\Theta)\mathcal{L}(X|\Theta)\pi(\Theta)\,d\Theta}{\int \mathcal{L}(X|\Theta)\pi(\Theta)\,d\Theta} \tag{28}$$

suggesting that the Bayes estimator $\hat{\Theta}_o(X) = \arg\min_U \mathcal{C}(U|X)$ and the corresponding optimum posterior cost $\mathcal{C}_o(X) = \min_U \mathcal{C}(U|X)$ can be expressed with the help of the conditional likelihood ratio as well.

Let us now examine the special case where for the cost function we adopt the squared error $c(U,\Theta) = \|U-\Theta\|^2$ which leads to the MSE criterion. From [18, p. 143], we know that the optimum estimator $\hat{\Theta}_o(X)$ is the conditional mean $\mathsf{E}_1[\Theta|X]$. If we also assume the prior $\pi(\Theta)$ to be uniform over some known set Ω with finite Lebesgue measure $\mu(\Omega)$, then

$$\begin{aligned}\mathcal{L}(X) &= \mu^{-1}(\Omega)\int_\Omega \mathcal{L}(X|\Theta)d\Theta \\ \hat{\Theta}_o(X) &= \frac{\int_\Omega \Theta\mathcal{L}(X|\Theta)d\Theta}{\int_\Omega \mathcal{L}(X|\Theta)d\Theta} \\ \mathcal{C}_o(X) &= \frac{\int_\Omega \|\hat{\Theta}_o(X)-\Theta\|^2\mathcal{L}(X|\Theta)d\Theta}{\int_\Omega \mathcal{L}(X|\Theta)d\Theta} \\ &= \frac{\int_\Omega \|\Theta\|^2\mathcal{L}(X|\Theta)d\Theta}{\int_\Omega \mathcal{L}(X|\Theta)d\Theta} - \|\hat{\Theta}_o(X)\|^2.\end{aligned} \tag{29}$$

We can see that $\mu(\Omega)$ does not enter in the computation of the estimate $\hat{\Theta}_o(X)$ and its quality index $\mathcal{C}_o(X)$. Although $\mu(\Omega)$ does appear in the likelihood ratio $\mathcal{L}(X)$, it is easy to verify that, in both tests, it can be transferred to the right-hand side and

absorbed by the corresponding threshold γ. We, therefore, conclude that explicit knowledge of this quantity is necessary only if we compute the thresholds and the performance analytically (and not through simulations). Finally, we note that in the MSE criterion, $\mathcal{C}_o(X)$ is the conditional variance of $\hat{\Theta}_o(X)$ which, clearly, constitutes a very meaningful quality index for the corresponding estimate.

We have now completed the development of our theory that addresses the joint detection and estimation problem. To demonstrate the power and originality of our analysis, first we apply our results to the example of retrospective change detection and then in Section III, at a much greater extent, we use them to solve the MIMO radar problem.

D. Example: Retrospective Change Detection

Retrospective change detection is the problem where within a given set of data $X = [x_1, \ldots, x_N]$, there is a possible time instant τ where the data switch statistics from some nominal pdf $f(X)$ before τ to an alternative pdf $h(X)$ after τ. We consider τ as the *last* time instant under the nominal regime. Given X, we are interested in detecting the change but also estimating the time τ the change took place. Among other applications, *offline* segmentation problems can be put under a retrospective change detection setup.

We should point out that retrospective change detection methodology is largely dominated by *sequential* techniques [3]. However, this constitutes a serious misusage of these methods since, in the retrospective formulation, the data are all available at once, whereas in the sequential setup the data become available sequentially. This means that by adopting sequential tests for the solution of the retrospective problem results in an inefficient utilization of the existing information.

Let us now apply our previous theory. Note that for $0 \leq \tau < N$, and by using the Bayes rule, the two pdfs can be decomposed as

$$\begin{aligned} f(X) &= f(x_1, \ldots, x_\tau) f(x_{\tau+1}, \ldots, x_N | x_1, \ldots, x_\tau) \\ h(X) &= h(x_1, \ldots, x_\tau) h(x_{\tau+1}, \ldots, x_N | x_1, \ldots, x_\tau). \end{aligned} \quad (30)$$

We first need to define the data pdf under the two hypotheses. Under H_0, we are under the nominal model; therefore, clearly, $f_0(X) = f(X)$. Under H_1 and with a change occurring at τ, we define the pdf $f_1(X|\tau)$ as follows:

$$f_1(X|\tau) = f(x_1, \ldots, x_\tau) h(x_{\tau+1}, \ldots, x_N | x_1, \ldots, x_\tau). \quad (31)$$

In other words, from the decompositions in (30), we combine the first part of the nominal pdf with the second part of the alternative. With this changepoint model, we assume that the data before the change affect the data after the change through the conditional pdf. In other words, the switching in the statistics is applied on the conditional data density. This is the most common model used in change detection theory but, we should mention that, it is also possible to come up with practically meaningful alternatives [20]. Note that $\tau > N-1$ means that all the data are under the nominal regime (i.e., there is no change) whereas $\tau = 0$ means that all the data are under the alternative regime. Therefore, under H_1, we have $\tau = 0, \ldots, N-1$ with some prior $\{\pi_0, \ldots, \pi_{N-1}\}$.

Let us compute the quantities that are necessary to apply our tests. Using (30), we can write for the conditional likelihood ratio

$$\mathcal{L}(X|\tau) = \frac{h(x_{\tau+1}, \ldots, x_N | x_1, \ldots, x_\tau)}{f(x_{\tau+1}, \ldots, x_N | x_1, \ldots, x_\tau)} \quad (32)$$

suggesting that the likelihood ratio, from (27), takes the form $\mathcal{L}(X) = \sum_{\tau=0}^{N-1} \pi_\tau \mathcal{L}(X|\tau)$.

Consider now the estimation problem. We propose the following cost function $c(U, \tau) = \mathbb{1}_{\{U \neq \tau\}}$, penalizing incorrect estimates by a unit cost. The average cost in this case coincides with the probability to estimate incorrectly. Observing that $\mathbb{1}_{\{U \neq \tau\}} = 1 - \mathbb{1}_{\{U = \tau\}}$, from (28) we can write

$$\mathcal{C}(U|X) = 1 - \frac{\pi_U \mathcal{L}(X|U)}{\sum_{\tau=0}^{N-1} \pi_\tau \mathcal{L}(X|\tau)} = 1 - \frac{\pi_U \mathcal{L}(X|U)}{\mathcal{L}(X)}. \quad (33)$$

Consequently, the optimum estimator that minimizes $\mathcal{C}(U|X)$ over $U \in \{0, \ldots, N-1\}$ is

$$\hat{\tau}_o(X) = \arg \max_{0 \leq U \leq N-1} \pi_U \mathcal{L}(X|U) \quad (34)$$

which is the MAP estimator [18, pp. 145–150], while the corresponding optimum posterior cost becomes

$$\mathcal{C}_o(X) = 1 - \frac{\max_{0 \leq U \leq N-1} \pi_U \mathcal{L}(X|U)}{\mathcal{L}(X)}. \quad (35)$$

The classical test that treats the two subproblems separately consists in comparing the likelihood ratio $\mathcal{L}(X)$ to the threshold γ_{NP} in order to distinguish between the two hypotheses and uses $\hat{\tau}_o(X)$ to estimate the time of change. GLRT, on the other hand, compares $\max_{0 \leq U \leq N-1} \mathcal{L}(X|U)$ to a threshold with the argument of this maximization providing the estimate for the time of change.

Applying our theory to this problem, for the single-step test, we use $\hat{\tau}_o(X)$ for the estimate of the changetime and either

$$\frac{\max_{0 \leq U \leq N-1} \pi_U \mathcal{L}(X|U)}{\mathcal{L}(X)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\lessgtr}} (1 - \lambda_o) \quad (36)$$

or

$$(\lambda - 1)\mathcal{L}(X) + \max_{0 \leq U \leq N-1} \pi_U \mathcal{L}(X|U) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\lessgtr}} \gamma \quad (37)$$

for the decision. For the two-step scheme, we compare the likelihood ratio $\mathcal{L}(X)$ to the threshold γ_{NP} to decide between the two hypotheses; use $\hat{\tau}_o(X)$ for the changepoint estimate and, finally, apply

$$\frac{\max_{0 \leq U \leq N-1} \pi_U \mathcal{L}(X|U)}{\mathcal{L}(X)} \underset{\mathsf{H}_{1_u}}{\overset{\mathsf{H}_{1_r}}{\lessgtr}} (1 - \lambda) \quad (38)$$

to label the estimate as reliable/unreliable. Both combined schemes resulting from our theory are completely original, nonsequential, and make efficient use of all available information.

III. Application to MIMO Radar

A context where performing joint detection and estimation is of particular interest is in radar systems. Radars are, often, deployed not only to detect a target but also estimate unknown parameters associated with the target, e.g., position and velocity. Recent developments in radar systems equip radars with multiple transmit and receive arrays that considerably improve their detection power and estimation accuracy compared with the conventional phased-array radars.

In this section, we examine the merits of the tests developed in the previous section for enhancing the detection and estimation quality by employing MIMO radar systems with widely separated antennas [22]. In particular, we are interested in the detection of a *single* target, and the estimation of its location every time the target is ruled present. This is, somewhat, different from the more conventional approaches in MIMO radar systems, e.g., [23] and references therein, where the probe space is broken into small subspaces and the radar detects the presence of the target in each of the subspaces separately. In this approach, as the location to be probed is given, one is only testing whether a target is present in a certain given subspace [23]. This necessitates implementing multiple detection tests in parallel, one for each subspace. In this section, we develop detectors and estimators based on the optimality theory discussed in the previous section which are used only once for the entire space.

A. System Description

We consider an MIMO radar system with M transmit and N receive antennas that are *widely* separated. Such spacing among the antennas ensures that the receivers capture uncorrelated reflections from the target. Conditions involving the distance between antennas in combination with the parameters of the transmitted waves that guarantee this behavior can be found in [23, Sec. II-A]. Both transmit and receive antennas are located at positions $\Theta_m^t \in \mathbb{R}^3$, for $m = 1, \ldots, M$, and $\Theta_n^r \in \mathbb{R}^3$, for $n = 1, \ldots, N$, respectively, known at the receiver.

The mth transmit antenna emits the waveform with baseband equivalent model given by $\sqrt{E}s_m(t)$ where E is the transmitted energy of a single transmit antenna (assuming to be the same for all transmitters); $\int_0^{T_s} |s_m(t)|^2 dt = 1$ where T_s denotes the common duration of all signals $s_m(t)$.

We aim to detect the presence of an *extended* target and when deemed to be present also estimate its position. The extended target consists of multiple scatterers exhibiting random, independent, and isotropic scintillation, each modeled with a complex random variable of zero mean and unknown distribution. This corresponds to the classical Swerling case I model [21] extended for multiple-antenna systems [22], [23]. The reflectivity factors are assumed to remain constant during a scan and are allowed to change independently from one scan to another.

We define Θ as the location of the gravity center of the target and $d_{mn}(\Theta)$ as the aggregate distance that a probing waveform $s_m(t)$ travels from the mth transmit antenna to the target and from the target to the nth receive antenna, i.e.

$$d_{mn}(\Theta) = \sqrt{\|\Theta - \Theta_m^t\|_2^2 + \|\Theta - \Theta_n^r\|_2^2}. \tag{39}$$

The time delay the waveform $s_m(t)$ is experiencing by traveling this distance $d_{mn}(\Theta)$ is equal to

$$\tau_{mn}(\Theta) = \frac{d_{mn}(\Theta)}{c} \tag{40}$$

where c is the speed of light. When the target dimensions are considerably smaller than the distance of the target from the transmit and receive antennas, the distance of the antennas to each scatterer of the target can be well approximated by their distances from the gravity center of the target. Therefore, the measured signal at the nth receive antenna is the superposition of all emitted waveforms and is given by the following equation [24]:

$$\xi_n(t) = \sqrt{E} \sum_{m=1}^{M} d_{mn}^{-\eta}(\Theta)\, g_{mn}\, s_m(t - \tau_{mn}(\Theta)) + v_n(t) \tag{41}$$

where $d_{mn}^{-\eta}$ is the path loss with η denoting the path-loss exponent; $v_n(t)$ denotes a standard complex-valued Gaussian "white noise" process distributed as[5] $\mathcal{N}_{\mathbb{C}}(0, t)$; and g_{mn} accounts for the reflectivity effects of the scatterers corresponding to the mth transmit and the nth receive antennas. It can be readily verified that $\{g_{mn}\}$ are independent and identically distributed (i.i.d.) with distribution $\mathcal{N}_{\mathbb{C}}(0, 1)$ [23], [24]. We note that we have assumed for the noises $v_n(t)$ and the coefficients g_{mn} that they are of unit variance. In fact, if we use any other values, e.g., σ_v^2 and σ_g^2, respectively, then in the final test these quantities are combined with the transmitted signal power E in the form of $E\sigma_g^2/\sigma_v^2$. Consequently, *provided* that in the general case σ_v^2 and σ_g^2 are known then, without loss of generality, we may assume $\sigma_v^2 = \sigma_g^2 = 1$ and let E express the final combination.

For $n = 1, \ldots, N$, define

$$\begin{aligned} G_n^H &= [g_{1n}, \ldots, g_{Mn}] \\ S_n'(t, \Theta) &= \sqrt{E} \left[\frac{s_1(t - \tau_{1n}(\Theta))}{d_{1n}^{\eta}(\Theta)}, \ldots, \frac{s_M(t - \tau_{Mn}(\Theta))}{d_{Mn}^{\eta}(\Theta)} \right] \end{aligned} \tag{42}$$

where we recall that $d_{mn}(\Theta)$ and $\tau_{mn}(\Theta)$ are *known* functions of Θ defined in (39) and (40) and A', A^H or $\mathbf{A}'$, $\mathbf{A}^H$ denote the transpose and Hermitian (transpose and complex conjugate), respectively, of a vector A or a matrix $\mathbf{A}$. Using (42), we can write

$$\xi_n(t) = G_n^H\, S_n(t, \Theta) + v_n(t). \tag{43}$$

Since we intend to use results from Itô Calculus, we will rewrite the previous equation under a stochastic differential equation form. We note that the Gaussian white noise process $v_n(t)$ can be considered as the derivative of a corresponding standard complex-valued Wiener process $w_n(t)$, i.e., $v_n(t) = \frac{dw_n(t)}{dt}$. Furthermore, if we consider the measured

[5] $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ denotes the distribution of a complex Gaussian random variable with mean $\mu = \mu_r + j\mu_i$ where the real and imaginary parts are uncorrelated (and therefore independent) Gaussian random variables with mean μ_r, μ_i, respectively, and of variance equal to $\sigma^2/2$.

signal $\xi_n(t)$ to be the derivative of another process $r_n(t)$, that is, $\xi_n(t) = \frac{dr_n(t)}{dt}$, then we can rewrite (43) as

$$\frac{dr_n(t)}{dt} = G_n^H\, S_n(t,\Theta) + \frac{dw_n(t)}{dt} \tag{44}$$

which in Itô Calculus is expressed under the following stochastic differential equation form:

$$dr_n(t) = G_n^H\, S_n(t,\Theta)dt + dw_n(t). \tag{45}$$

Let us, now, formulate the joint detection and estimation problem for the specific signal model we just introduced.

B. Target Detection/Localization With MIMO Radar

For $0 \le t \le T$, we distinguish the following two hypotheses satisfied by the received signals $r_n(t),\ n = 1,\ldots,N$

$\mathsf{H}_0 : dr_n(t) = dw_n(t)$

$\mathsf{H}_1 : dr_n(t) = G_n^H S_n(t,\Theta)dt + dw_n(t).$

As we can see, when there is no target present, the received signals are pure Wiener processes, whereas with the appearance of a target we have the emergence of the nonzero drifts $G_n^H S_n(t,\Theta)$.

For simplicity, let us use $\tilde{r}_n$ to denote the signal acquired by the nth receive antenna during the time interval $[0,T]$, that is, $\tilde{r}_n = \{r_n(t),\ 0 \le t \le T\}$. The collection of these N signals constitutes the complete set of observations; in other words, $\{\tilde{r}_1,\ldots,\tilde{r}_N\}$ plays the role of the observation vector X of the previous section. Of course, here, we slightly abuse our original definition since X was considered to be a random vector, while $\{\tilde{r}_1,\ldots,\tilde{r}_N\}$ is a collection of *random signals*. Nevertheless, this should not discourage us because we can always sample (arbitrarily fine) in time the N signals and end up with a random vector. This slight inconsistency was deemed necessary in order to be able to use existing results from Itô Calculus that provide, directly, closed-form expressions for the (conditional) likelihood ratio which is needed to define the tests of interest. The same results would have required a significantly more cumbersome analysis to obtain if we had followed the classical discrete-time methodology that goes through sampling of the received signals and then passes to the limit to obtain the corresponding continuous-time analog.

Clearly, our goal is to use $\{\tilde{r}_1,\ldots,\tilde{r}_N\}$ in order to decide between the presence or absence of a target and, every time a target is detected, to provide a reliable estimate of its position. To apply the theory developed in the previous section, according to Section II-D, we need to find the conditional likelihood ratio $\mathcal{L}(X|\Theta)$ which, here, becomes $\mathcal{L}(\tilde{r}_1,\ldots,\tilde{r}_N|\Theta)$. The following theorem provides the required formula.

Theorem 2: The conditional likelihood ratio $\mathcal{L}(\tilde{r}_1,\ldots,\tilde{r}_N|\Theta)$ of the received signals is given by

$$\mathcal{L}(\tilde{r}_1,\ldots,\tilde{r}_N|\Theta) = \prod_{n=1}^{N} \frac{\exp\{R_n^H(\Theta)(\mathbf{Q}_n(\Theta)+\mathbf{I}_M)^{-1}R_n(\Theta)\}}{|\mathbf{Q}_n(\Theta)+\mathbf{I}_M|} \tag{46}$$

where

$$\mathbf{Q}_n(\Theta) = \int_0^T S_n(t,\Theta)S_n^H(t,\Theta)dt$$
$$R_n^H(\Theta) = \int_0^T S_n^H(t,\Theta)dr_n(t) \tag{47}$$

$\mathbf{I}_K$ denotes the identity matrix of size K, and $|\mathbf{A}|$ the determinant of the matrix $\mathbf{A}$.

Proof: The proof is presented in the Appendix. ■

A final quantity which is of major interest for the next section is the appropriate definition of SNR. Note that, depending on the position of the target, the received signals $r_n(t)$ exhibit different SNR levels. This is due to the path-loss effect, which is, particularly, severe for distant targets. We, therefore, propose to measure the SNR by aggregating the signal and noise energies at the receivers but also *averaging* these quantities over all possible target positions $\Theta \in \Omega$. Specifically, by adopting the uniform model for Θ, we define

$$\mathrm{SNR} = \frac{\int_\Omega \left(\sum_{n=1}^N \int_0^T \mathsf{E}[|G_n^H S_n(t,\Theta)|^2]\,dt\right) d\Theta}{\int_\Omega \left(\sum_{n=1}^N \mathsf{E}[|\int_0^T dw_n(t)|^2]\right) d\Theta} \approx \frac{E}{NT}\frac{1}{\mu(\Omega)}\int_\Omega \left(\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{d_{mn}^{2\eta}(\Theta)}\right)d\Theta \tag{48}$$

where from standard Itô Calculus, the expectation in the denominator is equal to T. For the approximate equality, we overlooked the boundary effects in the numerator, that is, we assumed that $\int_0^T |s_n(t-\tau_{mn}(\Theta))|^2 dt = 1$ for all $\tau_{mn}(\Theta)$ which, of course, is not true when Θ is close to the boundary of Ω. If there is no path loss, that is $\eta = 0$, then the previous equation reduces to the simple formula $\mathrm{SNR} \approx \frac{EM}{T}$. The transmitted energy E will be tuned through these equations in order to deliver the appropriate SNR level at the receivers.

We have now developed all necessary formulas that allow us to use the results in Section II in the MIMO radar problem. In the next section, we evaluate the joint detection/estimation scheme with Monte Carlo simulations that cover various combinations of SNR values and number of transmit/receive antennas. We apply only the two-step test developed in Section II-C since, as we briefly argued earlier, it is more well suited for the MIMO radar problem.

C. Simulations

We consider the 2-D analog of the MIMO radar problem with two configurations consisting of $M = N = 2$ and $M = N = 3$ antennas, where the mth transmit and the nth receive antenna are located at $\Theta_m^t = [m,0]'$ and $\Theta_n^r = [0,n]'$ (expressed in kilometers), respectively.

The emitted waveforms are $s_m(t) = \frac{1}{\sqrt{T_s}} e^{j\frac{2\pi m}{T_s}t}$ for $t \in [0,T_s]$ where $T_s = 10^{-4}$ s is the signal duration. Moreover, we select an integration time $T = 5\,T_s = 5\times 10^{-4}$ s. This integration limit can accommodate delays $\tau_{mn}(\Theta)$ that do not exceed T (for larger delays, we simply measure noise during the interval $[0,T]$). The maximal delay defines a region Ω in space where every point $\Theta \in \Omega$ has at least one aggregate distance $d_{mn}(\Theta)$, defined in (39), from one transmit and one receive

antenna that does not exceed the value $cT \approx 150$ km. Actually, the points in space that have an aggregate distance from a pair of transmit/receive antennas not exceeding 150 km lie in the interior of a well-defined ellipse. Since we have MN pairs of transmit/receive antennas, we conclude that Ω is the union of an equal number of such ellipses. By considering that all antennas are roughly positioned at the origin, all ellipses become circles and Ω can be approximated by a disk of approximate radius of 75 km.

As is the usual practice in MIMO radar literature, we assume $\eta = 0$, namely, no path loss. This means that we are going to tune our energy parameter E through the simplified equation $\mathrm{SNR} \approx \frac{EM}{T}$. We consider SNR values $-20, -10, 0$, and 10 dB.

Assuming that the target position Θ is uniformly distributed within Ω and that for the cost function we employ the MSE criterion, we can use the formulas in (29) for the joint detection/estimation scheme. From (29) and (47), we observe the need for space and time integration. Both integrals will be evaluated numerically. For time integration, we use canonical sampling and consider L_t points $\{t_k\}$ within the time interval $[0, T]$. For integration in space, we form a canonical square grid of points for Θ. Denote with L_s the number of points $\{\Theta_l\}$ that lie in the interior of the region Ω. The two integrals are, then, approximated by sums. Specifically, the quantities in (47), for $\Theta = \Theta_l$, are approximated by

$$\mathbf{Q}_n(\Theta_l) \approx \frac{T}{L_t} \sum_{k=1}^{L_t} S_n(t_k, \Theta_l) S_n^H(t_k, \Theta_l) \tag{49}$$

and $R_n(\Theta_l)$ under H_0 (needed to compute the threshold γ_{NP}) takes the form

$$R_n^H(\Theta_l) \approx \sum_{k=1}^{L_t} S_n^H(t_k, \Theta_l) \Delta w_n(t_k) \tag{50}$$

while for the same quantity under H_1 we can write

$$R_n^H(\Theta_l) \approx \sum_{k=1}^{L_t} S_n^H(t_k, \Theta_l) \left\{ G_n^H S_n(t_k, \Theta_o) \frac{T}{L_t} + \Delta w_n(t_k) \right\}. \tag{51}$$

Parameter Θ_o denotes the "true" target position selected uniformly within Ω, and Θ_l is one of the L_s grid points in the interior of the same set. The coefficients G_n are selected randomly from a Gaussian $\mathcal{N}_{\mathbb{C}}(0, \mathbf{I}_M)$, while each $\Delta w_n(t)$ is also Gaussian $\mathcal{N}_{\mathbb{C}}(0, \frac{T}{L_t})$. For each run, the quantities G_n, Θ_o and $\Delta w_n(t_k)$ are the same for all Θ_l. For our simulations, we use $L_t = 500$ time samples $\{t_k\}$ and a grid with cells 10 km $\times$ 10 km that generates 179 points $\{\Theta_l\}$ in the interior of Ω.

For the test in Section II-C, according to (29), the likelihood ratio test is implemented as

$$\sum_{\Theta_l} \prod_{n=1}^{N} \frac{\exp\{R_n^H(\Theta_l)(\mathbf{Q}_n(\Theta_l) + \mathbf{I}_M)^{-1} R_n(\Theta_l)\}}{|\mathbf{Q}_n(\Theta_l) + \mathbf{I}_M|} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtrless}} \gamma. \tag{52}$$

Every time a decision is made in favor of H_1, we provide the following estimate of Θ_o:

$$\hat{\Theta}_o = \frac{\sum_{\Theta_l} \Theta_l \prod_{n=1}^{N} \frac{\exp\{R_n^H(\Theta_l)(\mathbf{Q}_n(\Theta_l) + \mathbf{I}_M)^{-1} R_n(\Theta_l)\}}{|\mathbf{Q}_n(\Theta_l) + \mathbf{I}_M|}}{\sum_{\Theta_l} \prod_{n=1}^{N} \frac{\exp\{R_n^H(\Theta_l)(\mathbf{Q}_n(\Theta_l) + \mathbf{I}_M)^{-1} R_n(\Theta_l)\}}{|\mathbf{Q}_n(\Theta_l) + \mathbf{I}_M|}} \tag{53}$$

with corresponding quality index $\mathcal{C}_o = \bar{\mathcal{C}}_o - \|\hat{\Theta}_o\|^2$, where

$$\bar{\mathcal{C}}_o = \frac{\sum_{\Theta_l} \|\Theta_l\|^2 \prod_{n=1}^{N} \frac{\exp\{R_n^H(\Theta_l)(\mathbf{Q}_n(\Theta_l) + \mathbf{I}_M)^{-1} R_n(\Theta_l)\}}{|\mathbf{Q}_n(\Theta_l) + \mathbf{I}_M|}}{\sum_{\Theta_l} \prod_{n=1}^{N} \frac{\exp\{R_n^H(\Theta_l)(\mathbf{Q}_n(\Theta_l) + \mathbf{I}_M)^{-1} R_n(\Theta_l)\}}{|\mathbf{Q}_n(\Theta_l) + \mathbf{I}_M|}}. \tag{54}$$

The estimate $\hat{\Theta}_o$ is characterized as reliable/unreliable depending on whether $\mathcal{C}_o$ is below/exceeds the threshold λ.

As we discussed in Remark 3, selecting $\beta = \beta_{\mathrm{NP}}(\alpha)$ reduces our test to the case where the two subproblems are treated separately with the corresponding optimum schemes. This test clearly constitutes a straightforward selection for solving the joint detection/estimation problem and becomes a special instance of our two-step scheme corresponding to a fraction value equal to 1. This is true because if we substitute in (24) $\beta = \beta_{\mathrm{NP}}(\alpha)$, the fraction value becomes 1, suggesting that all detections of the Neyman–Pearson test of the first step are considered to provide reliable estimates through the Bayesian estimator $\hat{\Theta}_o(X)$; consequently, there is no need to apply the second decision mechanism.

An alternative straightforward test is clearly the popular GLRT where we maximize the likelihood ratio $\mathcal{L}(\tilde{r}_1, \ldots, \tilde{r}_N|\Theta)$ in (46) over Θ and compare it to a threshold. We recall that GLRT decides between the two hypotheses and, at the same time, provides ML estimates for the parameters. The threshold is selected so that the corresponding false alarm probability is equal to α. Furthermore, every time the test decides in favor of H_1, we are using the corresponding estimate as reliable, consequently, for this test as well the fraction value is equal to 1.

For convenience, we will refer to GLRT and the test that treats the two subproblems separately as *conventional tests*. The goal of our simulation is to examine whether it is possible to enjoy any significant gains in performance when adopting our two-step method as compared to these two conventional tests.

Monte Carlo simulations were carried out in order to study the performance of the different tests. For each SNR value, 200 000 simulations were implemented to validate our theoretical developments. In our simulations, we fixed the false alarm probability to $\alpha = 10^{-3}$. The (conditional) MSE was computed as $\frac{1}{K} \sum \|\hat{\Theta}_o - \Theta_o\|^2$ where K is the total number of cases where the combined test decided in favor of $\mathsf{H}_{1\mathrm{r}}$ (that is, H_1 in the first step and $\mathsf{H}_{1\mathrm{r}}$ in the second).

In Fig. 1, we depict the MSE normalized by the (approximate) radius of Ω squared (75^2) as a function of the fraction

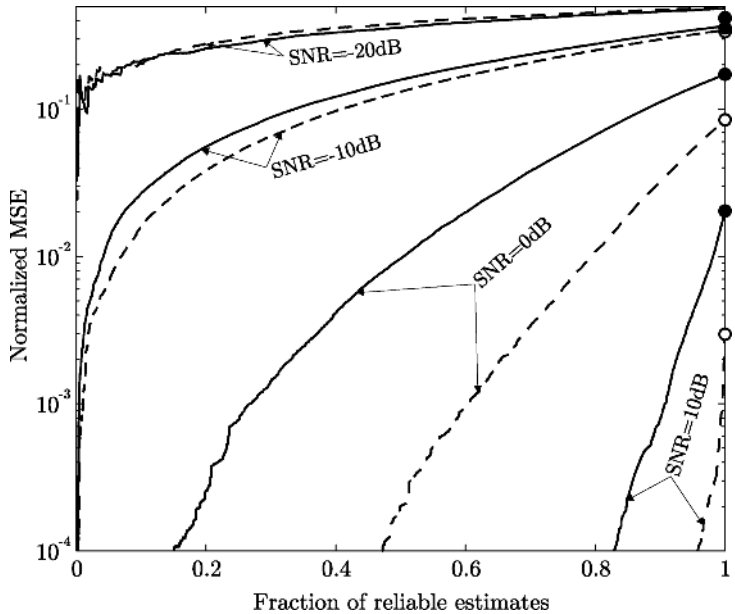

Fig. 1. Normalized MSE as a function of the fraction of reliable estimates for different values of SNR. Configuration $M = N = 2$: optimum is solid and GLRT is •; configuration $M = N = 3$: optimum is dashed and GLRT is ∘. Each curve for $\text{fraction} = 1$ yields the performance of the test that treats the two subproblems separately with the corresponding optimum scheme.

of reliable estimates, i.e., $\frac{P_1(D=H_{1r})}{P_1(D=H_1)}$. The fraction value is controlled through the threshold λ of the second decision mechanism defined in (26). We recall that fraction value equal to 1 in our test corresponds to the test that treats the two subproblems separately. For fraction value equal to 1, we also depict the performance of GLRT with little black or white circles (depending on the configuration).

The first important observation is that the two conventional tests have very similar performances. This is indeed the case since the little circles are either exactly on top of the beginning of each curve or very close to this performance value.

By moving to fraction values strictly smaller than 1, we start using our two-step test. We observe that as the fraction value becomes smaller, the corresponding estimation error performance is improving. More specifically, for $\text{SNR} = -20$ dB, we need to sacrifice more than 50% of our detections (more accurately in these cases we regard the estimates as unreliable) to reduce the MSE by a factor of 2 as compared to the performance we obtain by the two conventional tests. For larger SNR values, we can have more important (even enormous) gains. For example, for $\text{SNR} = 0$ dB by sacrificing 50% of the detections, in the 2×2 case, we gain an order of magnitude in estimation performance, while the same gain in the 3×3 configuration is achieved with only 25% reduction. Finally, in the 10 dB case, the gain can be improved even by two orders of magnitude with only a minor reduction in detection power. A last observation is that, apart the very low SNR case of -20 dB, the 3×3 antenna configuration is preferable to the 2×2 since it exhibits uniformly better performance.

Summarizing: from the simulations, we can assess that our scheme can produce non-negligible improvements in estimation accuracy, as compared to the two conventional tests, with only small sacrifices in detection power. In fact, in the case of medium SNR values, the corresponding gain can even become substantial (two orders of magnitude) with minor degradation in detection power.

We would like to complete our simulations section by pointing out that the assumption of uniform prior for the target position is clearly arbitrary. In order to overcome this practically serious modeling issue, one could use Bayesian nonparametric techniques [26], [27]. It is expected that such methods can play a significant role when we consider the problem of multiple target detection/localization and, more importantly, when the number of targets is unknown. Furthermore, techniques similar to the ones proposed in [28] could turn out to be beneficial when attempting to extend our MIMO radar results to cover the problem of target tracking.

IV. Conclusion

We have presented two possible formulations of the joint detection and estimation problem and developed the corresponding optimum solutions. Our approach consists in properly combining the Bayesian method for estimation with suitable constraints on the detection part. The resulting optimum schemes allow for the trade-off between detection power and estimation quality, thus emphasizing each subproblem according to needs of the actual application. Our theory was, then, applied to the problems of retrospective change detection and MIMO radar. In particular, in the second application, extensive simulations demonstrated the possibility to experience important improvement in estimation quality, as compared to GLRT and the approach that treats the two subproblems separately. In fact, with only small sacrifices in detection power even for low (as 0 dB) SNR values, the improvement can be non-negligible, while for medium (as 10 dB) SNR values the corresponding gain can be substantial.

APPENDIX

In order to avoid unnecessary technical complications in our proofs, we are going to make some additional simplifying assumptions. For a random variable x, let us denote with $x_{\inf}, x_{\sup}$ its essential infimum and supremum, respectively, and call $(x_{\inf}, x_{\sup})$ the interval of support of x. Observe now that the optimum posterior cost $\mathcal{C}_o(X)$, the likelihood ratio $\mathcal{L}(X) = f_1(X)/f_0(X)$ and $\mathcal{L}(X)[\lambda - \mathcal{C}_o(X)]$ with $\lambda \geq 0$, due to the randomness of X, become scalar random variables. We are going to assume that each of these three random variables has a pdf that possesses no delta functions and is strictly positive on the corresponding interval of support. This assumption guarantees that these random variables have a cumulative distribution function (cdf) which is continuous and strictly increasing from 0 to 1 in the corresponding interval of support.

1) Proof of Theorem 1: We are interested in minimizing $\tilde{\mathcal{J}}(\delta_1)$ defined in (14) subject to the two constraints $\int \delta_1(X) f_0(X) dX \leq \alpha$ and $\int \delta_0(X) f_1(X) dX \leq \beta$. We first note that if we have a pair $\{\delta_0(X), \delta_1(X)\}$ for which the second inequality is strict, then we can find another pair $\{\bar{\delta}_0(X), \bar{\delta}_1(X)\}$ which satisfies the second constraint with equality and has exactly the same estimation performance. Indeed, if we select $\bar{\delta}_1(X) = \frac{1-\beta}{\int \delta_1(X) f_1(X) dX} \delta_1(X)$, $\bar{\delta}_0(X) = 1 - \bar{\delta}_1(X)$, then we observe that since we assumed $\int \delta_0(X) f_1(X) dX < \beta$, we have $\int \delta_1(X) f_1(X) dX = 1 - \int \delta_0(X) f_1(X) dX > 1 - \beta$, suggesting that $\bar{\delta}_1(X)$ is a legitimate probability (because $\delta_1(X)$ is multiplied by a factor smaller than 1 to produce $\bar{\delta}_1(X)$); consequently, the complementary probability $\bar{\delta}_0(X)$ is legitimate as well. The fact that the alternative pair has exactly the same estimation performance, namely $\tilde{\mathcal{J}}(\delta_1) = \tilde{\mathcal{J}}(\bar{\delta}_1)$, can be verified by direct substitution.

Because of the previous observation, we can limit our search for the optimum within the class of tests that satisfy the constraint on the probability of miss with equality, that is, $\int \delta_0(X) f_1(X) dX = \beta$. Equivalently, we consider only tests that satisfy the equality constraint $\int \delta_1(X) f_1(X) dX = 1 - \beta$ on the detection probability. Under this equality, minimizing $\tilde{\mathcal{J}}(\delta_1)$ is equivalent to minimizing the numerator $\int \delta_1(X) \mathcal{C}_o(X) f_1(X) dX$ in (14).

Due to the nonnegativity of $\mathcal{C}_o(X)$, we note that, for $\lambda_o \geq 0$, the function $\mathsf{P}_1(\lambda_o \geq \mathcal{C}_o(X))$, in terms of λ_o, is the cdf of the random variable $\mathcal{C}_o(X)$. Since the cdf is continuous and strictly increasing in the interval of support of $\mathcal{C}_o(X)$, for $1-\beta \in (0,1)$, there exists unique $\lambda_o \geq 0$ that satisfies (16).

Suppose that we are in the case where $\alpha \geq \mathsf{P}_0(\mathcal{C}_o(X) \leq \lambda_o)$ and consider a test $\{\delta_0(X), \delta_1(X)\}$ that satisfies the equality $\int \delta_1(X) f_1(X) dX = 1 - \beta$. We can, then, write

$$\begin{aligned}
&\int \delta_1(X) \mathcal{C}_o(X) f_1(X) dX - \lambda_o(1-\beta) \\
&= \int \delta_1(X) \mathcal{C}_o(X) f_1(X) dX - \lambda_o \int \delta_1(X) f_1(X) dX \\
&= \int \delta_1(X) [\mathcal{C}_o(X) - \lambda_o] f_1(X) dX \\
&\geq \int \mathbb{1}_\Gamma [\mathcal{C}_o(X) - \lambda_o] f_1(X) dX \\
&= \int \mathbb{1}_\Gamma \mathcal{C}_o(X) f_1(X) dX - \lambda_o \mathsf{P}_1(\Gamma) \\
&= \int \mathbb{1}_\Gamma \mathcal{C}_o(X) f_1(X) dX - \lambda_o(1-\beta) \qquad (55)
\end{aligned}$$

where $\Gamma = \{\mathcal{C}_o(X) \leq \lambda_o\}$. Comparing the first with the last term yields $\int \delta_1(X) \mathcal{C}_o(X) f_1(X) dX \geq \int \mathbb{1}_\Gamma \mathcal{C}_o(X) f_1(X) dX$, which proves that (17) is the optimum since it minimizes the estimation criterion and satisfies both constraints. We observe in this case that, for the optimum test, the false alarm constraint can be strict.

Consider now the case $\alpha < \mathsf{P}_0(\mathcal{C}_o(X) \leq \lambda_o)$ and let us show that there is a pair λ, γ for which the test in (18) satisfies both constraints with equality. Define the functions

$$\psi_i(\lambda, \gamma) = \mathsf{P}_i(\mathcal{L}(X)[\lambda - \mathcal{C}_o(X)] \geq \gamma), \ i = 0, 1.$$

Then, $\psi_i(\lambda, \gamma)$ is clearly increasing in λ and decreasing in γ. In particular, for fixed λ, the monotonicity with respect to γ is strict within the interval of support of the random variable $\mathcal{L}(X)[\lambda - \mathcal{C}_o(X)]$ since we assumed that the corresponding pdf is positive inside this interval. We will show that $\psi_i(\lambda, \gamma)$ is also continuous with respect to each of its arguments when the other is fixed. Continuity with respect to γ is straightforward since, if we fix λ, the resulting function is the complementary cdf of $\mathcal{L}(X)[\lambda - \mathcal{C}_o(X)]$. The latter random variable was assumed to have no point masses; consequently, its (complementary) cdf is continuous. To show continuity with respect to λ requires a little more work. Consider $\epsilon > 0$, $M > 0$, and let $\lambda = \lambda_1 + \epsilon$; then, we can write

$$\begin{aligned}
\psi_i(\lambda_1, \gamma) &\leq \psi_i(\lambda_1 + \epsilon, \gamma) \\
&= \mathsf{P}_i(\mathcal{L}(X)[\lambda_1 - \mathcal{C}_o(X)] \geq \gamma - \epsilon \mathcal{L}(X)) \\
&\leq \mathsf{P}_i(\mathcal{L}(X)[\lambda_1 - \mathcal{C}_o(X)] \geq \gamma - \epsilon M) + \mathsf{P}_i(\mathcal{L}(X) \geq M) \\
&= \psi_i(\lambda_1, \gamma - \epsilon M) + \mathsf{P}_i(\mathcal{L}(X) \geq M). \qquad (56)
\end{aligned}$$

Sending $\epsilon \to 0+$ and using the continuity of the function $\psi_i(\lambda, \gamma)$ with respect to γ, we conclude that for every $M > 0$ we have

$$\psi_i(\lambda_1, \gamma) \leq \lim_{\epsilon \to 0+} \psi_i(\lambda_1 + \epsilon, \gamma) \leq \psi_i(\lambda_1, \gamma) + \mathsf{P}_i(\mathcal{L}(X) \geq M). \qquad (57)$$

Sending now $M \to \infty$, we deduce $\lim_{\epsilon \to 0+} \psi_i(\lambda_1 + \epsilon, \gamma) = \psi_i(\lambda_1, \gamma)$ suggesting right continuity with respect to λ. With similar steps, we can show left continuity.

Let us now show that for any $\lambda \geq \lambda_o$ we can find $\gamma(\lambda) \geq 0$ that satisfies the equality constraint for the detection probability, namely

$$\mathsf{P}_1(\mathcal{L}(X)[\lambda - \mathcal{C}_o(X)] \geq \gamma(\lambda)) = 1 - \beta. \qquad (58)$$

Consider $\psi_1(\lambda, \gamma) = \mathsf{P}_1(\mathcal{L}(X)[\lambda - \mathcal{C}_o(X)] \geq \gamma)$ and fix $\lambda \geq \lambda_o$; then, due to the monotonicity and continuity properties of $\psi_1(\lambda, \gamma)$, we have $\psi_1(\lambda, 0) \geq \psi_1(\lambda_o, 0) = \mathsf{P}_1(\mathcal{L}(X)[\lambda_o - \mathcal{C}(X)] \geq 0) = \mathsf{P}_1(\lambda_o - \mathcal{C}(X) \geq 0) = 1 - \beta$, with the second equality being true because $\mathcal{L}(X) \geq 0$ and $\mathsf{P}_1(\mathcal{L}(X) = 0) = 0$ [the latter being valid from the definition of λ_o in (16)]. We also observe that $\lim_{\gamma \to \infty} \psi_1(\lambda, \gamma) = 0 < 1 - \beta$. Since $\psi_1(\lambda, \gamma)$ is continuous in γ and strictly monotone, from 0 to 1, in the interval of support of $\mathcal{L}(X)[\lambda - \mathcal{C}_o(X)]$, while $\psi_1(\lambda, 0) \geq 1 - \beta$

and $\psi_1(\lambda,\gamma) < 1-\beta$ for sufficiently large γ, we can conclude that there exists unique $\gamma(\lambda)$ such that (58) is satisfied.

Regarding $\gamma(\lambda)$ as a function of λ, it is simple to show that it is increasing in λ. We will now prove that it is also continuous with respect to λ. As earlier, consider $\epsilon > 0$, $M > 0$, and let $\lambda = \lambda_1 + \epsilon$, then

$$\begin{aligned}\psi_1(\lambda_1,\gamma(\lambda_1+\epsilon)-\epsilon M) + \mathsf{P}_1(\mathcal{L}(X)\geq M) &\geq 1-\beta \\ = \psi_1(\lambda_1+\epsilon,\gamma(\lambda_1+\epsilon)) &\geq \psi_1(\lambda_1,\gamma(\lambda_1+\epsilon)). \quad (59)\end{aligned}$$

Letting first $\epsilon \to 0+$ and using the continuity of $\psi(\lambda,\gamma)$ with respect to its second argument, and, then, sending $M \to \infty$, we conclude that

$$\psi_1(\lambda_1,\gamma(\lambda_1+)) = 1-\beta \quad (60)$$

suggesting that $\gamma(\lambda_1+)$ is the solution of (58) for $\lambda = \lambda_1$; consequently, $\gamma(\lambda_1+) = \gamma(\lambda)$. In other words, $\gamma(\lambda)$ is right continuous. In the same way, we can show left continuity.

There are two pairs $\lambda,\gamma(\lambda)$ which we can describe explicitly. From the definition of λ_o, we know that when $\lambda = \lambda_o$ we have $\gamma(\lambda_o) = 0$. To find the second pair, let $\lambda \to \infty$ and assume that $\gamma(\lambda)/\lambda \to \tilde{\gamma}$; then, $\tilde{\gamma}$ is the solution to the equation

$$\mathsf{P}_1\left(\mathcal{L}(X) \geq \tilde{\gamma}\right) = 1-\beta. \quad (61)$$

This is true because the test in (18), after dividing each side by λ and letting $\lambda \to \infty$, reduces to the likelihood ratio test with threshold $\tilde{\gamma}$. Since by assumption we have $\beta > \beta_{\mathrm{NP}}(\alpha)$ where $\beta_{\mathrm{NP}}(\alpha)$ is the probability of miss of the Neyman–Pearson test, we conclude that $\tilde{\gamma} > \gamma_{\mathrm{NP}}$. This suggests that

$$\mathsf{P}_0\left(\mathcal{L}(X) \geq \tilde{\gamma}\right) < \mathsf{P}_0\left(\mathcal{L}(X) \geq \gamma_{\mathrm{NP}}\right) = \alpha. \quad (62)$$

The next step is to show that there exists a value for λ and the corresponding threshold $\gamma(\lambda)$ that satisfy the false alarm constraint with equality, namely

$$\mathsf{P}_0\left(\mathcal{L}(X)[\lambda - \mathcal{C}_o(X)] \geq \gamma(\lambda)\right) = \alpha. \quad (63)$$

Consider the function $\psi_0(\lambda,\gamma(\lambda)) = \mathsf{P}_0(\mathcal{L}(X)[\lambda - \mathcal{C}_o(X)] \geq \gamma(\lambda))$. We will show that this function is continuous in λ. Let $\epsilon > 0$; then, due to the monotonicity of the $\psi_0(\lambda,\gamma)$ with respect to each of its arguments and using the monotonicity of $\gamma(\lambda)$, we can write

$$\psi_0(\lambda_1+\epsilon,\gamma(\lambda_1)) \geq \psi_0(\lambda_1+\epsilon,\gamma(\lambda_1+\epsilon)) \geq \psi_0(\lambda_1,\gamma(\lambda_1+\epsilon)). \quad (64)$$

Using the continuity of $\psi_0(\lambda,\gamma)$ with respect to each of its arguments and the continuity of $\gamma(\lambda)$, by letting $\epsilon \to 0+$, we can conclude that

$$\lim_{\epsilon\to 0+} \psi_0(\lambda_1+\epsilon,\gamma(\lambda_1+\epsilon)) = \psi_0(\lambda_1,\gamma(\lambda_1)) \quad (65)$$

which proves right continuity. In the same way, we can show left continuity.

Consider now the function $\psi_0(\lambda,\gamma(\lambda)) - \alpha$. As we have shown, this function has opposite signs for $\lambda = \lambda_o$ and $\lambda \to \infty$ and it is continuous in λ; consequently, there exists λ and corresponding $\gamma(\lambda)$ for which the value of this function is exactly 0.

To show that the test in (18) is optimum, let $\lambda,\gamma(\lambda)$ be the previous pair and let $\{\delta_0(X),\delta_1(X)\}$ be any test that satisfies the equality constraint for the detection probability and the inequality constraint for the false alarm. Then, we can write

$$\begin{aligned}&\int \delta_1(X)\mathcal{C}_o(X)f_1(X)dX - \lambda(1-\beta) + \gamma(\lambda)\alpha \\ &\geq \int \delta_1(X)\{[\mathcal{C}_o(X)-\lambda]f_1(X) + \gamma(\lambda)f_0(X)\}dX \\ &\geq \int \mathbb{1}_\Gamma\{[\mathcal{C}_o(X)-\lambda]f_1(X) + \gamma(\lambda)f_0(X)\}dX \\ &= \int \mathbb{1}_\Gamma\, \mathcal{C}_o(X)f_1(X)dX - \lambda(1-\beta) + \gamma(\lambda)\alpha \quad (66)\end{aligned}$$

where $\Gamma = \{\frac{f_1(X)}{f_0(X)}[\lambda - \mathcal{C}_o(X)] \geq \gamma(\lambda)\}$. Again comparing the first with the last term proves optimality of the test in (18) and, therefore, concludes the proof of Theorem 1. ■

2) Proof of Theorem 2: Due to independence across receivers for the noises $\{w_n(t)\}$ and the reflection coefficients $\{g_{nm}\}$, we deduce

$$\mathcal{L}(\tilde{r}_1,\ldots,\tilde{r}_N|\Theta) = \prod_{n=1}^{N} \mathcal{L}(\tilde{r}_n|\Theta). \quad (67)$$

It is, thus, sufficient to show that

$$\mathcal{L}(\tilde{r}_n|\Theta) = \frac{\exp\{R_n^H(\Theta)(\mathbf{Q}_n(\Theta)+\mathbf{I}_M)^{-1}R_n(\Theta)\}}{|\mathbf{Q}_n(\Theta)+\mathbf{I}_M|}. \quad (68)$$

Since G_n is random, we can first compute $\mathcal{L}(\tilde{r}_n|G_n,\Theta)$ by conditioning on the coefficients G_n corresponding to the nth receiver and, then, average out G_n. For given G_n, the received signal $r_n(t)$ under the two hypotheses differs only in the drift; consequently, we can apply Girsanov's theorem [25, p. 191] to compute the corresponding likelihood ratio. We can treat the complex-valued Wiener process $\{w_n(t)\}$ as a 2-D real-valued Wiener process, with the real and imaginary parts of the complex process constituting the two independent components of the 2-D process. Since the corresponding variances, by assumption, are equal to 0.5, it is straightforward to show that

$$\begin{aligned}\mathcal{L}(\tilde{r}_n|G_n,\Theta) &= \exp\left\{-\int_0^T |G_n^H S_n(t,\Theta)|^2 dt \right. \\ &\quad \left. + 2\mathrm{Re}\left(\int_0^T [S_n^H(t,\Theta)G_n]dr_n(t)\right)\right\} \\ &= \exp\{-G_n^H\mathbf{Q}_n(\Theta)G_n + 2\mathrm{Re}(R_n^H(\Theta)G_n)\} \quad (69)\end{aligned}$$

where $\mathbf{Q}_n(\Theta), R_n(\Theta)$ are defined in (47)

In order to compute $\mathcal{L}(\tilde{r}_n|\Theta)$ from $\mathcal{L}(\tilde{r}_n|G_n,\Theta)$, we need to average out G_n. We recall that the real and imaginary parts of G_n are Gaussian uncorrelated (and thus independent) vectors, each with mean 0 and covariance matrix equal to $0.5\mathbf{I}_M$. For notational simplicity, we drop in all quantities their dependence on n and Θ. Let us also define the following decompositions into real and imaginary parts $G = G_r + jG_i$, $R = R_r + jR_i$, $\mathbf{Q} = \mathbf{Q}_r + j\mathbf{Q}_i$, and, finally, denote $\bar{G} = [G_r', G_i']'$, $\bar{R} = [R_r', R_i']'$, and $\bar{\mathbf{Q}} = [\mathbf{Q}_r, -\mathbf{Q}_i; \mathbf{Q}_i, \mathbf{Q}_r]$; then, we can write the previous likelihood ratio as follows:

$$\mathcal{L}(\tilde{r}|G,\Theta) = \exp\{-\bar{G}'\bar{\mathbf{Q}}\bar{G} + 2\bar{R}'\bar{G}\} \quad (70)$$

where we used the fact that $\mathbf{Q}$, by being Hermitian, satisfies $\mathbf{Q}_r' = \mathbf{Q}_r$ and $\mathbf{Q}_i' = -\mathbf{Q}_i$. We can now average out $\bar{G}$ by recalling that $\bar{G} \sim \mathcal{N}(0, 0.5\mathbf{I}_{2M})$. We have that

$$\begin{aligned}\mathcal{L}(\tilde{r}|\Theta) &= \int \exp\{-\bar{G}'\bar{\mathbf{Q}}\bar{G} + 2\bar{R}'\bar{G}\}\frac{1}{\pi^M}\exp\{-\bar{G}'\bar{G}\}d\bar{G} \\ &= \int \frac{1}{\pi^M}\exp\{-\bar{G}'(\bar{\mathbf{Q}}+\mathbf{I}_{2M})\bar{G} + 2\bar{R}'\bar{G}\}d\bar{G} \\ &= \frac{\exp\{\bar{R}'(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}\bar{R}\}}{\sqrt{|\bar{\mathbf{Q}}+\mathbf{I}_{2M}|}} \qquad (71)\end{aligned}$$

which can be obtained by "completing the square" in the exponential in the second line and reducing the integral, after proper normalization, into an integral of a Gaussian pdf with mean $(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}\bar{R}$ and covariance matrix $0.5(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}$ which is equal to 1.

From the nonegative definiteness of $\mathbf{Q}$, we have $Y^H\mathbf{Q}Y \geq 0$ for any complex vector Y. Using the observation that for any real vector Z, it is true that $Z'\mathbf{Q}_iZ = 0$, as a result of $\mathbf{Q}_i' = -\mathbf{Q}_i$, we can show that $[Y_r', Y_i']\bar{\mathbf{Q}}[Y_r', Y_i']' = Y^H\mathbf{Q}Y \geq 0$ where $Y = Y_r + jY_i$. Hence, $\bar{\mathbf{Q}}$ is nonegative definite as well, implying that $\bar{\mathbf{Q}}+\mathbf{I}_{2M}$ is positive definite.

Define two square matrices $\mathbf{A}, \mathbf{B}$ of size $M \times M$ as the solution to the following two equations: $(\mathbf{Q}_r+\mathbf{I}_M)\mathbf{A} - \mathbf{Q}_i\mathbf{B} = \mathbf{I}_M$ and $(\mathbf{Q}_r+\mathbf{I}_M)\mathbf{B} + \mathbf{Q}_i\mathbf{A} = \mathbf{0}_M$ (there always exists a solution due to the positive definiteness of $\bar{\mathbf{Q}}+\mathbf{I}_{2M}$); then, by direct computation, we can verify that $(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1} = [\mathbf{A}, -\mathbf{B}; \mathbf{B}, \mathbf{A}]$ and $(\mathbf{Q}+\mathbf{I}_M)^{-1} = (\mathbf{Q}_r+\mathbf{I}_M+j\mathbf{Q}_i)^{-1} = \mathbf{A}+j\mathbf{B}$. With the help of the previous equalities, we can show that $\bar{R}'(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}\bar{R} = R^H(\mathbf{Q}+\mathbf{I}_M)^{-1}R$. This proves the correctness of the exponential term in (68).

What is left to prove is that $\sqrt{|\bar{\mathbf{Q}}+\mathbf{I}_{2M}|} = |\mathbf{Q}+\mathbf{I}_M|$. Since $\bar{\mathbf{Q}}+\mathbf{I}_{2M} = [\mathbf{Q}_r+\mathbf{I}_M, -\mathbf{Q}_i; \mathbf{Q}_i, \mathbf{Q}_r+\mathbf{I}_M]$, if ρ is an eigenvalue of this matrix with corresponding eigenvector $[Y_r', Y_i']'$, then ρ is a double eigenvalue because by direct computation we can verify that $[-Y_i', Y_r']'$ is a second eigenvector (orthogonal to the first and, thus, different) for the same eigenvalue ρ. Consequently, the $2M$ eigenvalues of $\bar{\mathbf{Q}}+\mathbf{I}_{2M}$ are of the form $\rho_1, \rho_1, \ldots, \rho_M, \rho_M$ with $\rho_n > 0$ (because of the positive definiteness of $\bar{\mathbf{Q}}+\mathbf{I}_{2M}$), implying $\sqrt{|\bar{\mathbf{Q}}+\mathbf{I}_{2M}|} = \prod_{n=1}^M \rho_n$.

We can now verify that if $\rho, [Y_r', Y_i']'$ is an eigenvalue-eigenvector pair of $\bar{\mathbf{Q}}+\mathbf{I}_{2M}$, then $\rho, (Y_r + jY_i)$ is an eigenvalue-eigenvector pair of $\mathbf{Q}+\mathbf{I}_M$. This suggests that $\rho, (-Y_i + jY_r)$ must also be an eigenvalue-eigenvector pair for the same matrix. However, we observe that $(-Y_i + jY_r) = j(Y_r + jY_i)$, which means that the two eigenvectors are colinear and, therefore, coincide. Consequently, for the complex matrix $\mathbf{Q}+\mathbf{I}_M$, the eigenvalues are the $\rho_1, \ldots, \rho_M$, meaning that the corresponding determinant satisfies $|\mathbf{Q}+\mathbf{I}_M| = \prod_{n=1}^M \rho_n$. This proves that the desired equality for the two determinants demonstrates the validity of (68) and concludes the proof of Theorem 2. ■

REFERENCES

[1] B.-N. Vo, B.-T. Vo, N.-T. Pham, and D. Suter, "Joint detection and estimation of multiple objects from image observations," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5129–5141, Oct. 2010.

[2] S. Boutoille, S. Reboul, and M. Benjelloun, "A hybrid fusion system applied to off-line detection and change-points estimation," *Inf. Fusion*, vol. 11, no. 4, pp. 325–337, Oct. 2010.

[3] A. Vexler and C. Wu, "An optimal retrospective change point detection policy," *Scand. J. Stat.*, vol. 36, no. 3, pp. 542–558, Sep. 2009.

[4] L. Fillatre, I. Nikiforov, and F. Retraint, "A simple algorithm for defect detection from a few radiographies," *J. Comput.*, vol. 2, no. 6, pp. 26–34, Aug. 2007.

[5] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 3, pp. 434–444, May 1968.

[6] O. Zeitouni, J. Ziv, and N. Merhav, "When is the generalized likelihood ratio test optimal?," *IEEE Trans. Inf. Theory*, vol. 38, no. 5, pp. 1597–1602, Sep. 1992.

[7] E. J. Kelly, I. S. Reed, and W. Root, "The detection of radar echoes in noise I," *J. SIAM*, vol. 8, no. 2, pp. 309–341, Jun. 1960.

[8] E. J. Kelly, I. S. Reed, and W. Root, "The detection of radar echoes in noise II," *J. SIAM*, vol. 8, no. 3, pp. 481–505, Sep. 1960.

[9] B. Baygun and A. 0. Hero, "Optimal simultaneous detection and estimation under a false alarm constraint," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 688–703, May 1995.

[10] M. Rezaeian and B.-N. Vo, "Error bounds for joint detection and estimation of a single object with random finite set observation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1493–1505, Mar. 2010.

[11] A. Fredriksen, D. Middleton, and D. Vandelinde, "Simultaneous signal detection and estimation under multiple hypotheses," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 5, pp. 607–614, Sep. 1972.

[12] N. E. Nahi, "Optimal recursive estimation with uncertain observation," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 4, pp. 457–462, Jul. 1969.

[13] R. L. Stratonovitch, "Detection and estimation of signals in noise when one or both are non-Gaussian," *Proc. IEEE*, vol. 58, no. 5, pp. 670–679, May 1970.

[14] T. G. Birdsall and J. O. Gobien, "Sufficient statistics and reproducing densities in simultaneous sequential detection and estimation," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 6, pp. 760–768, Nov. 1973.

[15] E. Grossi, M. Lops, and V. Maroulas, "A sequential procedure for simultaneous detection and state estimation of Markov signals," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, Jun. 2009, pp. 654–658.

[16] G. V. Moustakides, Finite sample size optimality of GLR tests Nov. 2009 [Online]. Available: arXiv:0903.3795

[17] G. V. Moustakides, "Optimum joint detection and estimation," in *Proc. Int. Symp. Inf. Theory*, St. Petersburg, Russia, 2011, pp. 2984–2988.

[18] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. New York: Springer-Verlag, 1998.

[19] B. Levy, *Principles of Signal Detection and Parameter Estimation*. New York: Springer-Verlag, 2008.

[20] A. G. Tartakovsky and G. V. Moustakides, "State-of-the-art in Bayesian changepoint detection," *Seq. Anal.*, vol. 29, no. 2, pp. 125–145, Apr. 2010.

[21] M. Skolnik, *Introduction to Radar Systems*, 3rd ed. New York: McGraw-Hill, 2002.

[22] A. M. Haimovich, R. S. Blum, and L. J. Cimini, "MIMO radar with widely separated antennas," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 116–129, Jan. 2008.

[23] E. Fishler, A. Haimovich, R. S. Blum, L. J. Cimini, D. Chizhik, and R. A. Valenzuela, "Spatial diversity in radar-models and detection performance," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 823–837, Mar. 2006.

[24] A. Tajer, G. Jajamovich, X. Wang, and G. V. Moustakides, "Optimal joint target detection and parameter estimation by MIMO radar," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 1, pp. 127–145, Feb. 2010.

[25] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag, 1991.

[26] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.

[27] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Stat. Assoc.*, vol. 101, no. 476, pp. 1566–1581, 2006.

[28] E. B. Fox, D. S. Choi, and A. S. Willsky, "Nonparametric Bayesian methods for large-scale multi-target tracking," in *Proc. 40th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2006, pp. 2009–2013.

George V. Moustakides (M'82–SM'97) received the diploma in Electrical and Mechanical Engineering from the National Technical University of Athens, Greece, in 1979, the M.Sc. in Systems Engineering from the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, in 1980, and the Ph.D. in Electrical Engineering and Computer Science from Princeton University, Princeton, in 1983. From 1983 to 1986 he was a junior researcher with INRIA, France and during 1987 to 1990 he held a research position with the Computer Technology Institute of Patras, Greece. In 1991 he joined the University of Patras, Greece, as an Associate Professor and became Professor in 1996. Prof. Moustakides also held several appointments as Visiting Scholar, Senior Researcher and Adjunct Professor with Princeton University, University of Pennsylvania, Columbia University, Institute for Systems Research – University of Maryland, Georgia Institute of Technology, University of Southern California and INRIA. His research interests include sequential detection, statistical signal processing and signal processing for hearing aids. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY.

Guido H. Jajamovich (S'09) received the Ingeniero Electrónico degree (with honors) from the University of Buenos Aires, Buenos Aires, Argentina, in June 2006 and the M.S. degree from L'École Normale Supérieure de Cachan (ENS de Cachan), Paris, France, majoring in applied mathematics. He is currently working towards the Ph.D. degree from the Electrical Engineering Department, Columbia University, New York. During Fall 2007, he was a visitor at the Center for Imaging Science, Applied Mathematics and Statistics Department, The Johns Hopkins University, Baltimore, MD. His research interests include statistical signal processing and genomic signal processing.

Ali Tajer (S'05–M'10) is a Postdoctoral Research Associate at Princeton University. He received the Ph.D. degree from Columbia University and the M.Sc. and B.Sc. degrees from Sharif University of Technology, all in Electrical Engineering. His research interests are in the areas of applied statistics, information theory, and signal processing.

Xiaodong Wang (S'98–M'98–SM'04–F'08) received the Ph.D. degree in electrical engineering from Princeton University, Princeton, NJ. He is a Professor in the Department of Electrical Engineering, Columbia University, New York. His research interests fall in the general areas of computing, signal processing, and communications, and has published extensively in these areas. Among his publications is a book entitled Wireless Communication Systems: Advanced Techniques for Signal Reception, (Prentice-Hall, 2003). His current research interests include wireless communications, statistical signal processing, and genomic signal processing. Dr. Wang received the 1999 NSF CAREER Award and the 2001 IEEE Communications Society and Information Theory Society Joint Paper Award. He has served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and the IEEE TRANSACTIONS ON INFORMATION THEORY.