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 Baysian 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 case. The proposed combined strategies have the very desirable characteristic to allow for the trade-off between detection power and estimation efficiency. Our theoretical developments are then applied to the problems of retrospective changepoint detection and 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 demonstrate that by using the jointly optimum schemes, we can experience significant improvement in estimation quality with small sacrifice in detection power.

Index Terms

Joint detection-estimation, Retrospective change detection, MIMO radar.

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 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.

In the literature, there are basically two (mainly ad-hoc) approaches that deal with combined problems. The first consists in treating the two subproblems separately and applying in each case the corresponding optimum technique. For instance one can use the Neyman-Pearson optimum test for detection and the optimum Bayesian estimator for parameter estimation to solve the combined problem. As we will see in our analysis, and it is usually the case in combined problems, treating each part separately with the optimum scheme, does not necessarily result in optimum overall performance. The second methodology consists in using the Generalized Likelihood Ratio Test (GLRT) which detects and estimates at the same time with the parameter estimation part relying on the maximum likelihood estimator. Both approaches lack versatility and are not capable of emphasizing each subproblem according to the needs of the corresponding application.

Surprisingly, one can find *very* limited literature that deals with optimum solutions of the joint detection and estimation problem. A purely Bayesian technique is reported in [5], whereas a combination of Bayesian and Neyman-Pearson-like methodology is developed in [6]. Specifically in [6] 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 constraining the estimation cost under the nominal hypothesis while optimizing the corresponding cost under the alternative, gives rise to a number of interesting combined tests that can be used in place of GLRT.

Here we will build upon the methodology of [6] but we are going to formulate the combined problem in a more natural way. In particular we will define a performance measure for the estimation part which we are going to optimize assuring, in parallel, the satisfactory performance of the detection part by imposing suitable constraints on the decision error probabilities. This idea will lead to two novel combined tests that have no equivalent in [5],[6].

We would like to point out that the theory in [5],[6] as well as the one we are going to develop in our work, 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 nuisance parameters.

Our article 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 schemes, 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 the Introduction, we limit ourselves to the binary hypothesis case with parameters present only under the alternative hypothesis. Suppose we are given an observation signal X for which we have the following two hypotheses

$\mathsf{H}_0 : \ X \sim f_0(X)$

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

where $f_0(X), f_1(X|\theta), \pi(\theta)$ are known pdfs. Specifically, 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 available some prior 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; 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 $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_1)$ is maximized subject to the false alarm constraint $\mathsf{P}_0(\mathsf{D} = \mathsf{H}_1) \leq \alpha$. Equivalently, the previous maximization can be replaced by the minimization of the probability of miss $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_0)$. The optimum detection scheme is the well celebrated likelihood ratio test, which takes the following form

for our specific setup

$$\mathcal{L}(X) = \frac{f_1(X)}{f_0(X)} = \frac{\int f_1(X|\theta)\pi(\theta)\,d\theta}{f_0(X)} \underset{\mathsf{H_0}}{\overset{\mathsf{H_1}}{\gtrless}} \gamma_{\mathrm{NP}}. \tag{1}$$

In other words we decide $\mathsf{H_1}$ whenever the likelihood ratio $\mathcal{L}(X)$ exceeds the threshold γ_{NP}; $\mathsf{H_0}$ whenever it falls below 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

$$\delta_0^{\mathrm{NP}}(X) = \mathbb{1}_{\left\{\frac{f_1(X)}{f_0(X)} < \gamma_{\mathrm{NP}}\right\}} + (1-p)\mathbb{1}_{\left\{\frac{f_1(X)}{f_0(X)} = \gamma_{\mathrm{NP}}\right\}}$$
$$\delta_1^{\mathrm{NP}}(X) = \mathbb{1}_{\left\{\frac{f_1(X)}{f_0(X)} > \gamma_{\mathrm{NP}}\right\}} + p\mathbb{1}_{\left\{\frac{f_1(X)}{f_0(X)} = \gamma_{\mathrm{NP}}\right\}}, \tag{2}$$

where $\mathbb{1}_{\mathcal{A}}$ denotes the index function of the set $\mathcal{A}$.

Bayesian parameter estimation: Suppose that we know *with certainty* that the observations X come from hypothesis $\mathsf{H_1}$, 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 $\mathsf{E_1}[C(\hat{\theta}(X), \theta)]$, where expectation is with respect to X and θ.

From [7, Page 142] we have that the optimum Bayesian estimator is the following minimizer (provided it exists)

$$\hat{\theta}_o(X) = \arg\inf_U \mathcal{C}(U|X), \tag{3}$$

where $\mathcal{C}(U|X)$ is the posterior cost function

$$\mathcal{C}(U|X) = \mathsf{E_1}[C(U, \theta)|X] = \frac{\int C(U, \theta)f_1(X|\theta)\pi(\theta)\,d\theta}{\int f_1(X|\theta)\pi(\theta)\,d\theta} = \frac{\int C(U, \theta)f_1(X|\theta)\pi(\theta)\,d\theta}{f_1(X)}, \tag{4}$$

and expectation, as we can see from the last equality, is with respect to θ for given X. Finally we denote the optimum posterior cost as $\mathcal{C}_o(X)$, that is,

$$\mathcal{C}_o(X) = \inf_U \mathcal{C}(U|X) = \mathcal{C}(\hat{\theta}_o(X)|X). \tag{5}$$

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 $\mathsf{H_0}$ and $\mathsf{H_1}$. 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 efficiency for

the estimation part. Consequently, we understand, that the detection part cannot be treated independently from the estimation part. Following this rationale, we propose two possible approaches involving single 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. Single-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 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

$$\mathcal{J}(\delta_0, \delta_1, \hat{\theta}) = \mathsf{E}_1[C(\hat{\theta}(X), \theta)|\mathsf{D} = \mathsf{H}_1] = \frac{\mathsf{E}_1[C(\hat{\theta}(X), \theta)\mathbb{1}_{\{\mathsf{D}=\mathsf{H}_1\}}]}{\mathsf{P}_1(\mathsf{D} = \mathsf{H}_1)}, \tag{6}$$

where expectation is with respect to X and θ. We realize that with our criterion, the estimation performance depends not only on the estimator but also on the detection mechanism. As we can see, we compute the average cost over the event $\{\mathsf{D} = \mathsf{H}_1\}$, which is the only case an estimate is available.

One would immediately argue that the measure in (6) does not consider in any sense the decision errors, that is, the quality of the detector. However, these errors can be efficiently controlled through suitable constraints. 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 $\mathcal{J}(\delta_0, \delta_1, \hat{\theta})$ over the decision rule $\{\delta_0(X), \delta_1(X)\}$ and the estimator $\hat{\theta}(X)$. Note that the two constraints guarantee satisfactory performance for the detection part and, by minimizing the criterion, we can enjoy optimum performance in the estimation part.

Let us carry out the desired optimization gradually. We first fix the decision rule $\{\delta_0(X), \delta_1(X)\}$ and optimize $\mathcal{J}(\delta_0, \delta_1, \hat{\theta})$ with respect to the estimator $\hat{\theta}(X)$. We have the following lemma that provides the solution to this problem.

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

$$\mathcal{D}(\hat{\theta}) = \frac{\iint \varphi(X) C(\hat{\theta}(X), \theta) f_1(X|\theta) \pi(\theta) d\theta dX}{\iint \varphi(X) f_1(X|\theta) \pi(\theta) d\theta dX} \tag{7}$$

is minimized when $\hat{\theta}(X)$ is the optimum Bayesian estimator $\hat{\theta}_o(X)$ defined in (3) and (4).

Proof: The proof is simple. We can write

$$
\begin{aligned}
\mathcal{D}(\hat{\theta}) &= \frac{\iint \varphi(X) C(\hat{\theta}(X), \theta) f_1(X|\theta) \pi(\theta) d\theta dX}{\iint \varphi(X) f_1(X|\theta) \pi(\theta) d\theta dX} \\
&= \frac{\int \varphi(X) \left(\int C(\hat{\theta}(X), \theta) f_1(X|\theta) \pi(\theta) d\theta \right) dX}{\int \varphi(X) \left(\int f_1(X|\theta) \pi(\theta) d\theta \right) dX} \\
&= \frac{\int \varphi(X) \mathcal{C}(\hat{\theta}(X)|X) f_1(X) dX}{\int \varphi(X) f_1(X) dX} \\
&\geq \frac{\int \varphi(X) \inf_U \mathcal{C}(U|X) f_1(X) dX}{\int \varphi(X) f_1(X) dX} \\
&= \frac{\int \varphi(X) \mathcal{C}(\hat{\theta}_o(X)|X) f_1(X) dX}{\int \varphi(X) f_1(X) dX} = \frac{\int \varphi(X) \mathcal{C}_o(X) f_1(X) dX}{\int \varphi(X) f_1(X) dX}, \quad (8)
\end{aligned}
$$

where for the last two equalities we used (5). ∎

Lemma 1 is a very interesting result because it demonstrates an extended optimality property for the classical Bayesian estimator. In particular by selecting $\varphi(X) = \delta_1(X)$ we 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 $\hat{\theta}_o(X)$ with corresponding optimized performance measure equal to

$$
\bar{\mathcal{J}}(\delta_0, \delta_1) = \mathcal{J}(\delta_0, \delta_1, \hat{\theta}_o) = \frac{\int \delta_1(X) \mathcal{C}_o(X) f_1(X) dX}{\int \delta_1(X) f_1(X) dX}. \quad (9)
$$

It is clear that our intention is to further minimize $\bar{\mathcal{J}}(\delta_0, \delta_1)$ over the class of detectors that satisfy the two error constraints. Before addressing this problem however, we need to make some remarks.

Remark 1: One can argue that by constraining the false alarm probability to α and by using the Neyman-Pearson optimum test for detection and then the Bayesian estimator for estimation (in other words, treating the two subproblems separately) has definite optimality properties, since this combination optimizes both the detection and the estimation part. This is indeed true, however with such a scheme the main emphasis is on the detection part. For estimation, after optimizing the corresponding performance (by using $\hat{\theta}_o(X)$), we have no further control. In fact if the resulting estimation performance is not satisfactory, there is no room for further improvement. This weakness is clearly circumvented by the proposed formulation which offers, as we discuss next, the additional flexibility to trade detection power for estimation efficiency, according to the needs of the designer.

Remark 2: We recall that in our setup we have the two constraints $P_0(D = H_1) \leq \alpha$ and $P_1(D = H_0) \leq \beta$. By fixing the false alarm probability to α, the probability of miss is minimized by the Neyman-Pearson

test; call this minimum value $\beta(\alpha)$. Since no test, with false alarm probability not exceeding α, can have a probability of miss that goes below $\beta(\alpha)$, this suggests that in our constraint on the probability of miss, β must be selected to satisfy $\beta \geq \beta(\alpha)$. We are thus reducing, in a controlled manner, the detection power as compared to the Neyman-Pearson test (since we allow more misses) aiming in improving the effectiveness of our estimation. We have the following theorem that provides the optimum scheme.

Theorem 1: Consider the two constraints $\mathsf{P}_0(\mathsf{D} = \mathsf{H}_1) \leq \alpha$ and $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_0) \leq \beta$, where $0 < \alpha < 1$ and $\beta(\alpha) \leq \beta < 1$ with $\beta(\alpha)$ denoting the probability of miss of the Neyman-Pearson test. Let $\lambda_o > 0$ be the solution of the equation[1]

$$\mathsf{P}_1\left(\lambda_o \geq \mathcal{C}_o(X)\right) = 1 - \beta, \tag{10}$$

where $\mathcal{C}_o(X)$ is defined in (5). Then the optimum combined scheme is comprised of the Bayesian estimator $\hat{\theta}_o(X)$ defined in (3), (4), for the estimation part while the decision rule that optimizes the average conditional cost $\bar{\mathcal{J}}(\delta_0, \delta_1)$ in (9) under the two error constraints is given by

$$\mathcal{C}_o(X) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\lessgtr}} \lambda_o, \quad \text{if } \alpha \geq \mathsf{P}_0\left(\lambda_o \geq \mathcal{C}_o(X)\right) \tag{11}$$

$$\frac{f_1(X)}{f_0(X)}[\lambda - \mathcal{C}_o(X)] \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtrless}} \gamma, \quad \text{if } \alpha < \mathsf{P}_0\left(\lambda_o \geq \mathcal{C}_o(X)\right), \tag{12}$$

where in (12) λ, γ are selected so that the two error probability constraints are satisfied with equality.

Proof: The proof is presented in the Appendix. ■

From (11) and (12) we deduce that the optimum detector 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, in both cases, the test decides in favor of H_0. In particular, in (12), this decision may occur even if the classical likelihood ratio exceeds the threshold γ_{NP}, suggesting decision in favor of H_1.

Sumarizing, our first optimum combined test consist in applying (11) or (12) 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 (3) 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 most applications, giving up part of the detection capacity may be

---

[1]For simplicity we assume that $\mathcal{C}_o(X)$ and $f_1(X)/f_0(X)$, when considered as random variables, have no atoms under both hypotheses (the corresponding pdfs have no delta functions). This avoids the need for randomization every time a test statistic hits a threshold.

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 H_{1r} and H_{1u} respectively. Consequently we propose to make *three* different decisions $\mathsf{H}_0, \mathsf{H}_{1r}$ and H_{1u} with the union of the last two corresponding to hypothesis H_1. As we can see, we "trust" the estimate $\hat{\theta}(X)$ only when we decide in favor of H_{1r}, but we have detection even if we discard the estimate as unreliable, that is, we decide H_{1u}.

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 $\{q_{1r}(X), q_{1u}(X)\}$. 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 $q_{1r}(X), q_{1u}(X) \geq 0$ and $q_{1r}(X) + q_{1u}(X) = 1$. For every combination of the four randomization probabilities we define, similarly to the previous subsection, the corresponding average conditional cost for the estimator $\hat{\theta}(X)$, namely

$$\mathcal{J}(\delta_0, \delta_1, q_{1r}, q_{1u}, \hat{\theta}) = \mathsf{E}_1[C(\hat{\theta}(X), \theta)|\mathsf{D} = \mathsf{H}_{1r}] = \frac{\int \delta_1(X) q_{1r}(X) \mathcal{C}(\hat{\theta}(X)|X) f_1(X) dX}{\int \delta_1(X) q_{1r}(X) f_1(X) dX}. \tag{13}$$

As we can see, we now condition on the event $\{\mathsf{D} = \mathsf{H}_{1r}\}$ 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 H_{1r} is $\delta_1(X) q_{1r}(X)$ because we must decide in favor of H_1 in the first step (with probability $\delta_1(X)$) and for H_{1r} in the second (with probability $q_{1r}(X)$).

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 $\delta_0^{\mathrm{NP}}(X), \delta_1^{\mathrm{NP}}(X)$ given in (2).

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 average conditional cost $\mathcal{J}(\delta_0^{\mathrm{NP}}, \delta_1^{\mathrm{NP}}, q_{1r}, q_{1u}, \hat{\theta})$, assume $q_{1r}(X), q_{1u}(X)$ fixed, then from Lemma 1 and by selecting $\varphi(X) = \delta_1^{\mathrm{NP}}(X) q_{1r}(X)$, 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 (3) and (4). Call

$$\bar{\mathcal{J}}(q_{1r}, q_{1u}) = \mathcal{J}(\delta_0^{\mathrm{NP}}, \delta_1^{\mathrm{NP}}, q_{1r}, q_{1u}, \hat{\theta}_o) = \frac{\int \delta_1^{\mathrm{NP}}(X) q_{1r}(X) \mathcal{C}_o(X) f_1(X) dX}{\int \delta_1^{\mathrm{NP}}(X) q_{1r}(X) f_1(X) dX}, \tag{14}$$

the corresponding performance. It is then clear that we would like to minimize even further this criterion by selecting properly our second decision mechanism which is expressed with the help of the randomization probabilities $\{q_{1r}(X), q_{1u}(X)\}$. Note however that, in addition to this minimization, we are also interested in generating as many "reliable estimates" as possible when applying the second test. These two goals are clearly 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, because $q_{1r}(X) \leq 1$, the probability $\mathsf{P}_1(\mathsf{D} = \mathsf{H}_{1r})$ of deciding in favor of H_{1r} (reliable estimate) satisfies

$$\mathsf{P}_1(\mathsf{D} = \mathsf{H}_{1r}) = \int \delta_1^{\mathrm{NP}}(X) q_{1r}(X) f_1(X) dX \leq \int \delta_1^{\mathrm{NP}}(X) f_1(X) dX = \mathsf{P}_1(\mathsf{D} = \mathsf{H}_1) = 1 - \beta(\alpha). \tag{15}$$

In other words this probability is upper bounded by the detection probability $1 - \beta(\alpha)$ of the Neyman-Pearson test where, we recall, $\beta(\alpha)$ denotes the corresponding probability of miss. This inequality reveals the obvious fact that, only a portion of our initial decisions in favor of H_1 provide reliable estimates in the second step. Actually it is this part we intend to control by imposing the following inequality

$$1 - \beta \leq \mathsf{P}_1(\mathsf{D} = \mathsf{H}_{1r}) = \int \delta_1^{\mathrm{NP}}(X) q_{1r}(X) f_1(X) dX \tag{16}$$

with $1 > \beta \geq \beta(\alpha)$. The constraint in (16) expresses our desire that *at least a fraction of* $\frac{1-\beta}{1-\beta(\alpha)} \leq \frac{\mathsf{P}_1(\mathsf{D}=\mathsf{H}_{1r})}{\mathsf{P}_1(\mathsf{D}=\mathsf{H}_1)} \leq 1$ *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, that is, minimize the performance measure $\bar{\mathcal{J}}(q_{1r}, q_{1u})$. The solution to this optimization problem is given in the next lemma.

Lemma 2: Let $1 > \beta \geq \beta(\alpha)$, then the test that minimizes the average conditional cost $\bar{\mathcal{J}}(q_{1r}, q_{1u})$ defined in (14) subject to the constraint in (16), is given by

$$\mathcal{C}_o(X) \underset{\mathsf{H}_{1u}}{\overset{\mathsf{H}_{1r}}{\lessgtr}} \lambda, \tag{17}$$

where λ is selected to satisfy (16) with equality and $\mathcal{C}_o(X)$ is defined in (5).

Proof: The proof follows a methodology which is very similar to the one used in the proof of Theorem 1. Since it presents no particular difficulties, it is omitted. ∎

As in the previous subsection, $\mathcal{C}_o(X)$ 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 is

below some threshold λ while the estimate is discarded as unreliable whenever the same quantity exceeds the threshold.

Summarizing our second detection/estimation scheme: We first use the Neyman-Pearson test (1) to decide between $\mathsf{H}_0, \mathsf{H}_1$. Whenever we decide in favor of H_1 we compute the estimate $\hat{\theta}_o(X)$ from (3) and its corresponding quality index $\mathcal{C}_o(X)$ from (5); then we use the test in (17) to characterize the estimate as reliable/unreliable.

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 we can write it as

$$\mathcal{L}(X) = \frac{f_1(X)}{f_0(X)} = \int \mathcal{L}(X|\theta)\pi(\theta)d\theta. \tag{18}$$

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

$$\mathcal{C}(U|X) = \frac{\int C(U,\theta)\mathcal{L}(X|\theta)\pi(\theta)\,d\theta}{\int \mathcal{L}(X|\theta)\pi(\theta)\,d\theta} \tag{19}$$

suggesting that the Bayes estimator $\hat{\theta}_o(X) = \arg\inf_U \mathcal{C}(U|X)$ and the corresponding optimum posterior cost $\mathcal{C}_o(X) = \inf_U \mathcal{C}(U|X)$ 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 [7, Page 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

$$\begin{aligned}
\mathcal{L}(X) &= \frac{f_1(X)}{f_0(X)} = \mu^{-1}(\Omega) \int_\Omega \mathcal{L}(X|\theta)d\theta \\
\hat{\theta}_o(X) &= \frac{\int_\Omega \theta\mathcal{L}(X|\theta)d\theta}{\int_\Omega \mathcal{L}(X|\theta)d\theta} \\
\mathcal{C}_o(X) &= \frac{\int_\Omega \|\hat{\theta}_o(X) - \theta\|^2\mathcal{L}(X|\theta)d\theta}{\int_\Omega \mathcal{L}(X|\theta)d\theta} \\
&= \frac{\int_\Omega \|\theta\|^2\mathcal{L}(X|\theta)d\theta}{\int_\Omega \mathcal{L}(X|\theta)d\theta} - \|\hat{\theta}_o(X)\|^2.
\end{aligned} \tag{20}$$

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 no explicit knowledge of this quantity is necessary. 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 reasonable 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.

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 $0 \leq \tau < N$, the two pdfs can be decomposed as

$$f(X) = f(x_1, \ldots, x_\tau) \times f(x_{\tau+1}, \ldots, x_N | x_1, \ldots, x_\tau)$$
$$h(X) = h(x_1, \ldots, x_\tau) \times h(x_{\tau+1}, \ldots, x_N | x_1, \ldots, x_\tau). \tag{21}$$

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

$$f_1(X|\tau) = f(x_1, \ldots, x_\tau) \times h(x_{\tau+1}, \ldots, x_N | x_1, \ldots, x_\tau). \tag{22}$$

In other words, from the decompositions in (21), we combine the first part of the nominal pdf with the second part of the alternative. With this changepoint model, the data before the change affect the data after the change through the conditional pdf. This is the most common model used in change detection theory [8]. Note that $\tau > N - 1$ means that all the data are under the nominal regime (i.e. there is no change) whereas $\tau = 0$ that all the data are under the alternative regime. Therefore, under H_1 we have $\tau \in \{0, \ldots, N - 1\}$ with some prior $\{\pi_0, \ldots, \pi_{N-1}\}$.

Let us compute the quantities that are necessary to apply our tests. Using (21) we can write for the conditional likelihood ratio

$$\mathcal{L}(X|\tau) = \frac{h(x_{\tau+1}, \ldots, x_N | x_1, \ldots, x_\tau)}{f(x_{\tau+1}, \ldots, x_N | x_1, \ldots, x_\tau)}, \tag{23}$$

suggesting that the likelihood ratio, from (18), 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 is clearly the probability to estimate incorrectly. Observing that $\mathbb{1}_{\{U \neq \tau\}} = 1 - \mathbb{1}_{\{U=\tau\}}$, from (19) we can write

$$\mathcal{C}(U|X) = 1 - \frac{\mathcal{L}(X|U)\pi_U}{\sum_{\tau=0}^{N-1} \pi_\tau \mathcal{L}(X|\tau)} = 1 - \frac{\mathcal{L}(X|U)\pi_U}{\mathcal{L}(X)}. \tag{24}$$

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

$$\hat{\tau}_o(X) = \arg \max_{0 \leq U \leq N-1} \mathcal{L}(X|U)\pi_U, \tag{25}$$

which is the MAP estimator [7, Pages 145-150]; while the corresponding optimum posterior cost becomes

$$\mathcal{C}_o(X) = 1 - \frac{\max_{0 \leq U \leq N-1} \mathcal{L}(X|U)\pi_U}{\mathcal{L}(X)}. \tag{26}$$

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 use $\hat{\tau}_o(X)$ 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

$$\frac{\max_{0 \leq U \leq N-1} \mathcal{L}(X|U)\pi_U}{\mathcal{L}(X)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtrless}} (1 - \lambda_o), \tag{27}$$

or

$$(\lambda - 1)\mathcal{L}(X) + \max_{0 \leq U \leq N-1} \mathcal{L}(X|U)\pi_U \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtrless}} \gamma, \tag{28}$$

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

$$\frac{\max_{0 \leq U \leq N-1} \mathcal{L}(X|U)\pi_U}{\mathcal{L}(X)} \underset{\mathsf{H}_{1u}}{\overset{\mathsf{H}_{1r}}{\gtrless}} (1 - \lambda) \tag{29}$$

to label the estimate as reliable/unreliable. Both combined schemes resulting from our theory, are completely original 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 multiple-input multiple-output (MIMO) radar systems with widely-separated antennas [9]. In particular we are interested in the detection of a target, and the estimation of its location every time a target is ruled present. This is somewhat different from the more conventional approaches in MIMO radar systems, e.g., [10] 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 [10]. 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 a MIMO radar system with M transmit and N receive antennas that are *widely* separated (satisfy the conditions in [10, Sec. II.A]). Such spacing among the antennas ensures that the receivers capture uncorrelated reflections from the target. Both transmit and receive antennas are located at positions $\theta_m^t \in \mathbb{R}^3$, for $m \in \{1, \ldots, M\}$, and $\theta_n^r \in \mathbb{R}^3$, for $n \in \{1, \ldots, N\}$, 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); $\int_0^{T_s} |s_m(t)|^2 dt = 1$ and 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 [11] extended for multiple-antenna systems [9], [10]. 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.,

$$d_{mn}(\theta) = \sqrt{\|\theta - \theta_m^t\|_2^2 + \|\theta - \theta_n^r\|_2^2}. \tag{30}$$

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

$$\tau_{mn}(\theta) = \frac{d_{mn}(\theta)}{c}, \tag{31}$$

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 received signal at the nth receive antenna is the superposition of all emitted waveforms and is given by [12]

$$r_n(t) = \sqrt{E} \sum_{m=1}^{M} d_{mn}^{-\eta}(\theta)\, g_{mn}\, s_m(t - \tau_{mn}(\theta)) + w_n(t), \tag{32}$$

where $d_{mn}^{-\eta}$ is the path-loss with η denoting the path-loss exponent; $w_n(t)$ the additive white Gaussian complex valued noise distributed as[2] $\mathcal{N}_{\mathbb{C}}(0,1)$; 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 $\mathcal{N}_{\mathbb{C}}(0,1)$ [10], [12]. We note that we have assumed for the noises $w_n(t)$ and the coefficients g_{mn} that they have variance equal to 1. In fact if we use any other values e.g. σ_w^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 $E\sigma_g^2/\sigma_w^2$. Consequently, *provided* that in the general case σ_w^2 and σ_g^2 are known then, without loss of generality, we may assume $\sigma_w^2 = \sigma_g^2 = 1$ and let E express the final combination.

For $n \in \{1, \ldots, N\}$ define

$$\begin{aligned} G_n^H &= [g_{1n}, \ldots, g_{Mn}] \\ S_n'(t,\theta) &= \sqrt{E} \left[\frac{s_1(t - \tau_{1n}(\theta))}{d_{1n}^\eta(\theta)}, \ldots, \frac{s_M(t - \tau_{Mn}(\theta))}{d_{Mn}^\eta(\theta)} \right], \end{aligned} \tag{33}$$

---

[2]$\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ 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 $\sigma^2/2$.

where we recall that $d_{mn}(\theta)$ and $\tau_{mn}(\theta)$ are *known* functions of θ defined in (30), (31) and $\mathbf{A}', \mathbf{A}^H$ denote the transpose and Hermitian (transpose and complex conjugate) respectively of the matrix $\mathbf{A}$. Under these definitions we can write

$$r_n(t) = G_n^H \cdot S_n(t, \theta) + w_n(t). \tag{34}$$

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

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

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

We have written the received signals in a stochastic differential equation form, since the $\{w_n(t)\}$ are Wiener (white Gaussian noise) processes. As we can see, when there is no target present the measured 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 $\bar{r}_n$ to denote the signal acquired by the nth receive antenna during the time-interval $[0, T]$, that is, $\bar{r}_n = \{r_n(t), \ 0 \leq t \leq T\}$. The collection of these N signals constitutes the complete set of observations, in other words, $\{\bar{r}_1, \ldots, \bar{r}_N\}$ plays the role of the observation signal X of the previous section. Clearly, our goal is to use $\{\bar{r}_1, \ldots, \bar{r}_N\}$ 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-C, we need to find the conditional likelihood ratio $\mathcal{L}(\bar{r}_1, \ldots, \bar{r}_N | \theta)$. The following theorem provides the required formula.

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

$$\mathcal{L}(\bar{r}_1, \ldots, \bar{r}_N | \theta) = \prod_{n=1}^{N} \frac{e^{R_n^H(\theta)(\mathbf{Q}_n(\theta) + \mathbf{I}_M)^{-1} R_n(\theta)}}{|\mathbf{Q}_n(\theta) + \mathbf{I}_M|}, \tag{35}$$

where

$$\begin{aligned} \mathbf{Q}_n(\theta) &= \int_0^T S_n(t, \theta) S_n^H(t, \theta) dt, \\ R_n^H(\theta) &= \int_0^T S_n^H(t, \theta) dr_n(t), \end{aligned} \tag{36}$$

$\mathbf{I}_K$ 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 that 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

$$\text{SNR} = \frac{\int_\Omega \left(\sum_{n=1}^N \int_0^T \mathsf{E}[|G_n^H S_n(t,\theta)|^2]\, dt \right) d\theta}{\int_\Omega \left(\sum_{n=1}^N \mathsf{E}[|\int_0^T dw_n(t)|^2] \right) d\theta} \approx \frac{E}{NT} \frac{1}{\mu(\Omega)} \int_\Omega \left(\sum_{n=1}^N \sum_{m=1}^M \frac{1}{d_{mn}^{2\eta}(\theta)} \right) d\theta, \qquad (37)$$

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 $\text{SNR} \approx \frac{E \times M}{T}$. 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 enable us to use the results of Section II in the MIMO radar problem. In the next subsection 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-B since, as we briefly argued earlier, it is more well suited for the MIMO radar problem.

C. Simulations

We consider the two-dimensional 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 Km), 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 $T_s = 10^{-4}$ sec is the signal duration. Moreover, we select an integration time $T = 5 \times T_s = 5 \times 10^{-4}$ sec. 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 (30), from one transmit and one receive antenna that does not exceed the value $c \times T \approx 150$ 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 $M \times N$ 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 disc 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 SNR $\approx \frac{E \times M}{T}$. 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 (20) for the joint detection/estimation scheme. From (20) and (36) 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 (36), for $\theta = \theta_l$, are approximated by

$$\mathbf{Q}_n(\theta_l) \approx \frac{T}{L_t} \times \sum_{k=1}^{L_t} S_n(t_k, \theta_l) S_n^H(t_k, \theta_l) \tag{38}$$

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

$$R_n^H(\theta_l) \approx \sum_{k=1}^{L_t} S_n^H(t_k, \theta_l) \Delta w_n(t_k), \tag{39}$$

while for the same quantity under H_1 we can write

$$R_n^H(\theta_l) \approx \sum_{k=1}^{L_t} S_n^H(t_k, \theta_l) \{G_n^H S_n(t_k, \theta_o) \times \frac{T}{L_t} + \Delta w_n(t_k)\}. \tag{40}$$

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 of Section II-B, according to (20), the likelihood ratio test is implemented as

$$\sum_{\theta_l} \prod_{n=1}^{N} \frac{e^{R_n^H(\theta_l)(\mathbf{Q}_n(\theta_l) + \mathbf{I}_M)^{-1} R_n(\theta_l)}}{|\mathbf{Q}_n(\theta_l) + \mathbf{I}_M|} \mathop{\gtrless}_{\mathsf{H}_0}^{\mathsf{H}_1} \gamma. \tag{41}$$

Every time a decision is made in favor of H_1 we provide the following estimate of θ_o

$$\hat{\theta}_o = \frac{\displaystyle\sum_{\theta_l} \theta_l \prod_{n=1}^{N} \frac{e^{R_n^H(\theta_l)(\mathbf{Q}_n(\theta_l)+\mathbf{I}_M)^{-1}R_n(\theta_l)}}{|\mathbf{Q}_n(\theta_l)+\mathbf{I}_M|}}{\displaystyle\sum_{\theta_l} \prod_{n=1}^{N} \frac{e^{R_n^H(\theta_l)(\mathbf{Q}_n(\theta_l)+\mathbf{I}_M)^{-1}R_n(\theta_l)}}{|\mathbf{Q}_n(\theta_l)+\mathbf{I}_M|}}, \tag{42}$$

with corresponding quality index

$$\mathcal{C}_o = \frac{\displaystyle\sum_{\theta_l} \|\theta_l\|^2 \prod_{n=1}^{N} \frac{e^{R_n^H(\theta_l)(\mathbf{Q}_n(\theta_l)+\mathbf{I}_M)^{-1}R_n(\theta_l)}}{|\mathbf{Q}_n(\theta_l)+\mathbf{I}_M|}}{\displaystyle\sum_{\theta_l} \prod_{n=1}^{N} \frac{e^{R_n^H(\theta_l)(\mathbf{Q}_n(\theta_l)+\mathbf{I}_M)^{-1}R_n(\theta_l)}}{|\mathbf{Q}_n(\theta_l)+\mathbf{I}_M|}} - \|\hat{\theta}_o\|^2. \tag{43}$$

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

We also consider the GLRT where we maximize the likelihood ratio $\mathcal{L}(\bar{r}_1, \ldots, \bar{r}_N | \theta)$ in (35) over θ and compare it to a threshold. The threshold is selected so that the corresponding false alarm probability is equal to α. We recall that GLRT provides ML estimates for θ and, as we mentioned, cannot trade detection power for estimation.

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 H_{1r} (that is, H_1 in the first step and H_{1r} 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 of reliable estimates, i.e. $\frac{\mathsf{P}_1(\mathsf{D}=\mathsf{H}_{1r})}{\mathsf{P}_1(\mathsf{D}=\mathsf{H}_1)}$. The fraction value is controlled through the threshold λ. Fraction value equal to 1 in our test corresponds to the performance of the classical approach where detection and estimation are treated separately. For the same value we also present the performance of the GLRT. We observe that for $\mathrm{SNR} = -20\,\mathrm{dB}$ 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. For larger SNR values we can have significant (even enormous) gains. For example for $\mathrm{SNR} = 0\,\mathrm{dB}$ 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. We conclude from our simulations that apart the very low SNR case of $-20\,\mathrm{dB}$, the 3×3 antenna configuration
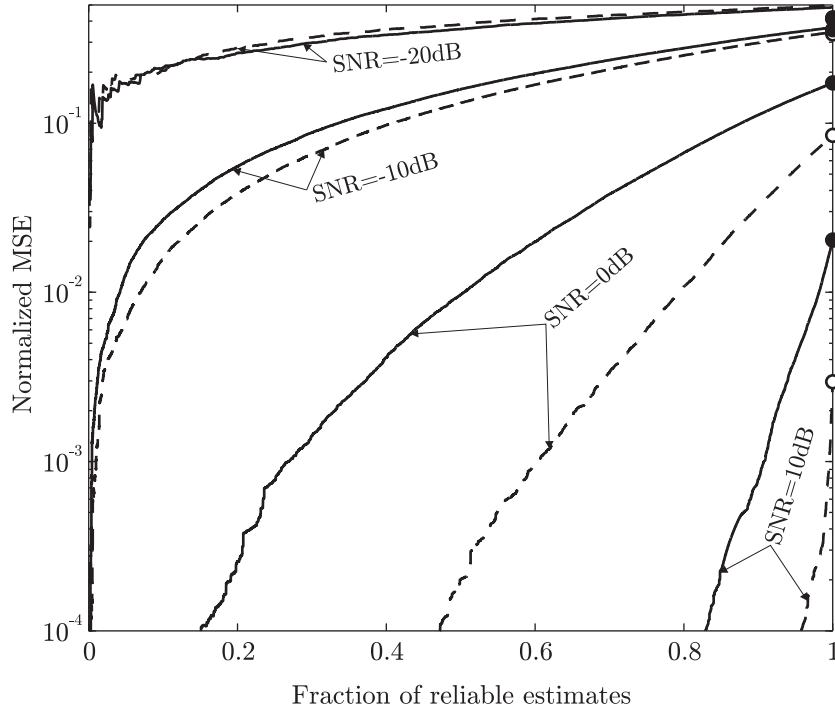
Fig. 1.   Normalized MSE as a function of the fraction of reliable estimates for different values of SNR. Configuration $M = N = 2$: optimum is solid and GLRT is $\bullet$; configuration $M = N = 3$: optimum is dashed and GLRT is $\circ$.

is preferable to the 2×2 since it can return significant performance gains. Finally, we observe that GLRT and the classical approach that treats the two subproblems separately have very comparable performance.

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 efficiency, thus emphasizing each subproblem according to needs of the original application. Our theory was then applied to the problems of retrospective change detection and MIMO radar. In particular in the second application, intense simulations demonstrated the possibility to experience significant gains in estimation quality with small sacrifices in detection power.

V. APPENDIX

Proof of Theorem 1: We are interested in minimizing $\bar{\mathcal{J}}(\delta_0, \delta_1)$ defined in (9) 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 same estimation performance, namely $\bar{\mathcal{J}}(\delta_0, \delta_1) = \bar{\mathcal{J}}(\bar{\delta}_0, \bar{\delta}_1)$, can be verified by direct substitution.

With 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 $\bar{\mathcal{J}}(\delta_0, \delta_1)$ is equivalent to minimizing the numerator $\int \delta_1(X) \mathcal{C}_o(X) f_1(X) dX$ in (9).

Due to the nonnegativity of $\mathcal{C}_o(X)$ and our assumption that $\mathcal{C}_o(X)$ does not contain any atoms we have that (10) has a unique solution $\lambda_o > 0$. 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

$$\int \delta_1(X) \mathcal{C}_o(X) f_1(X) dX - \lambda_o(1 - \beta)$$
$$= \int \delta_1(X) \mathcal{C}_o(X) f_1(X) dX - \lambda_o \int \delta_1(X) f_1(X) dX$$
$$= \int \delta_1(X) [\mathcal{C}_o(X) - \lambda_o] f_1(X) dX$$
$$\geq \int \mathbb{1}_{\mathcal{A}} [\mathcal{C}_o(X) - \lambda_o] f_1(X) dX$$
$$= \int \mathbb{1}_{\mathcal{A}} \mathcal{C}_o(X) f_1(X) dX - \lambda_o \mathsf{P}_1(\mathcal{A})$$
$$= \int \mathbb{1}_{\mathcal{A}} \mathcal{C}_o(X) f_1(X) dX - \lambda_o(1 - \beta), \quad (44)$$

where $\mathcal{A} = \{\mathcal{C}_o(X) \leq \lambda_o\}$. Comparing the first and the last term yields $\int \delta_1(X) \mathcal{C}_o(X) f_1(X) dX \geq \int \mathbb{1}_{\mathcal{A}} \mathcal{C}_o(X) f_1(X) dX$, which proves that (11) 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 (12) satisfies both constraints with equality. We are first going to prove that for any $\lambda \geq \lambda_o$ we can find $\gamma(\lambda) \geq 0$ to satisfy the equality constraint for the detection probability, namely

$$\mathsf{P}_1 \left(\frac{f_1(X)}{f_0(X)}[\lambda - \mathcal{C}_o(X)] \geq \gamma(\lambda) \right) = 1 - \beta. \tag{45}$$

Call $\psi(\lambda, \gamma) = \mathsf{P}_1([f_1(X)/f_0(X)][\lambda - \mathcal{C}_o(X)] \geq \gamma) - (1 - \beta)$, fix $\lambda > \lambda_o$, then we observe that $\psi(\lambda, 0) > \psi(\lambda_o, 0) = 0$. Furthermore $\lim_{\gamma \to \infty} \psi(\lambda, \gamma) = -(1 - \beta) < 0$. Consequently there exists $\gamma(\lambda)$ such that (45) is true. 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 $\gamma(\lambda_o) = 0$. Consider now $\lambda \to \infty$ and assume that $\gamma(\lambda)/\lambda \to \bar{\gamma}$, then $\bar{\gamma}$ is the solution to the equation

$$\mathsf{P}_1 \left(\frac{f_1(X)}{f_0(X)} \geq \bar{\gamma} \right) = 1 - \beta. \tag{46}$$

This is true because the test in (12), after dividing each side by λ and letting $\lambda \to \infty$ reduces to the likelihood ratio test with threshold $\bar{\gamma}$. Since by assumption we have $\beta > \beta(\alpha)$ where $\beta(\alpha)$ is the probability of miss of the Neyman-Pearson test, we conclude that $\bar{\gamma} > \gamma_{\mathrm{NP}}$. This suggests that

$$\mathsf{P}_0 \left(\frac{f_1(X)}{f_0(X)} \geq \bar{\gamma} \right) < \mathsf{P}_0 \left(\frac{f_1(X)}{f_0(X)} \geq \gamma_{\mathrm{NP}} \right) = \alpha \tag{47}$$

Now we need to show that there exists a value for λ and the corresponding threshold $\gamma(\lambda)$ that satisfy the false alarm constraint with equality, namely

$$\mathsf{P}_0 \left(\frac{f_1(X)}{f_0(X)}[\lambda - \mathcal{C}_o(X)] \geq \gamma(\lambda) \right) = \alpha. \tag{48}$$

Call $\phi(\lambda) = \mathsf{P}_0([f_1(X)/f_0(X)][\lambda - \mathcal{C}_o(X)] \geq \gamma(\lambda)) - \alpha$. Then, because of our previous analysis, it is easy to verify that $\phi(\lambda)$ has opposite signs for $\lambda = \lambda_o$ and $\lambda \to \infty$, meaning that there exists a $\lambda > \lambda_o$ such that $\phi(\lambda) = 0$, or that the false alarm constraint is satisfied with equality.

To show that the test in (12) is optimum, let $\lambda, \gamma(\lambda)$ be the previous pair and consider any test $\{\delta_0(X), \delta_1(X)\}$ that satisfies the equality constraint for the detection probability and the inequality

constraint for the false alarm. Then we can write

$$
\int \delta_1(X)\mathcal{C}_o(X)f_1(X)dX - \lambda(1-\beta) + \gamma(\lambda)\alpha
$$

$$
\geq \int \delta_1(X)\{[\mathcal{C}_o(X) - \lambda]f_1(X) + \gamma(\lambda)f_0(X)\}dX
$$

$$
\geq \int \mathbb{1}_\mathcal{A}\{[\mathcal{C}_o(X) - \lambda]f_1(X) + \gamma(\lambda)f_0(X)\}dX
$$

$$
= \int \mathbb{1}_\mathcal{A}\mathcal{C}_o(X)f_1(X)dX - \lambda(1-\beta) + \gamma(\lambda)\alpha, \quad (49)
$$

where $\mathcal{A} = \{\frac{f_1(X)}{f_0(X)}[\lambda - \mathcal{C}_o(X)] \geq \gamma(\lambda)\}$. Again comparing the first and the last term, proves optimality of the test in (12) and therefore concludes the proof of Theorem 1. ∎

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

$$
\mathcal{L}(\bar{r}_1,\ldots,\bar{r}_N|\theta) = \prod_{n=1}^{N} \mathcal{L}(\bar{r}_n|\theta). \tag{50}
$$

It is thus sufficient to show that

$$
\mathcal{L}(\bar{r}_n|\theta) = \frac{e^{R_n^H(\theta)(\mathbf{Q}_n(\theta)+\mathbf{I}_M)^{-1}R_n(\theta)}}{|\mathbf{Q}_n(\theta)+\mathbf{I}_M|}. \tag{51}
$$

Since G_n is random, we can first compute $\mathcal{L}(\bar{r}_n|G_n,\theta)$ 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 [13, Page 191] to compute the corresponding likelihood ratio. We can treat the complex valued Wiener process $\{w_n(t)\}$ as a two dimensional real valued Wiener process, with the real and imaginary part of the complex process constituting the two independent components of the two dimensional process. Since the corresponding variances, by assumption, are equal to 0.5, it is straightforward to show that

$$
\mathcal{L}(\bar{r}_n|G_n,\theta) = e^{-\int_0^T |G_n^H S_n(t,\theta)|^2 dt + 2\mathrm{Re}\left(\int_0^T [S_n^H(t,\theta)G_n]dr_n(t)\right)} = e^{-G_n^H \mathbf{Q}_n(\theta)G_n + 2\mathrm{Re}(R_n^H(\theta)G_n)}, \tag{52}
$$

where $\mathbf{Q}_n(\theta), R_n(\theta)$ are defined in (36)

In order to compute $\mathcal{L}(\bar{r}_n|\theta)$ from $\mathcal{L}(\bar{r}_n|G_n,\theta)$ 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 $\mathcal{G} = [G_r', G_i']'$, $\mathcal{R} = [R_r', R_i']'$, $\bar{\mathbf{Q}} = [\mathbf{Q}_r, -\mathbf{Q}_i; \mathbf{Q}_i, \mathbf{Q}_r]$;

then we can write the previous likelihood ratio as follows

$$\mathcal{L}(\bar{r}|G,\theta) = e^{-\mathcal{G}'\bar{\mathbf{Q}}\mathcal{G}+2\mathcal{R}'\mathcal{G}}, \tag{53}$$

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 $\mathcal{G}$ by recalling that $\mathcal{G} \sim \mathcal{N}(0, 0.5\mathbf{I}_{2M})$. By "completing the square" we have

$$\mathcal{L}(\bar{r}|\theta) = \int e^{-\mathcal{G}'\bar{\mathbf{Q}}\mathcal{G}+2\mathcal{R}'\mathcal{G}} \frac{1}{\pi^M} e^{-\mathcal{G}'\mathcal{G}} d\mathcal{G}$$

$$= \frac{e^{\mathcal{R}'(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}\mathcal{R}}}{\sqrt{|\bar{\mathbf{Q}}+\mathbf{I}_{2M}|}} \times \int \frac{e^{-(\mathcal{G}-(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}\mathcal{R})'(\bar{\mathbf{Q}}+\mathbf{I}_{2M})(\mathcal{G}-(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}\mathcal{R})}}{\sqrt{\frac{(2\pi)^{2M}}{|2(\bar{\mathbf{Q}}+\mathbf{I}_{2M})|}}} d\mathcal{G}$$

$$= \frac{e^{\mathcal{R}'(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}\mathcal{R}}}{\sqrt{|\bar{\mathbf{Q}}+\mathbf{I}_{2M}|}}, \tag{54}$$

where the last integral is equal to 1 since it is the integral of a Gaussian pdf with mean $(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}\mathcal{R}$ and covariance matrix $0.5(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}$.

From the nonnegative 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 $Z'\mathbf{Q}_i Z = 0$, 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 nonnegative 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 have $\mathcal{R}'(\bar{\mathbf{Q}}+\mathbf{I}_{2M})^{-1}\mathcal{R} = R^H(\mathbf{Q}+\mathbf{I}_M)^{-1}R$. This proves the correctness of the exponential term in (51).

What is left to show 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 co-linear 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 the desired equality for the two determinants, demonstrates the validity of (51) and concludes the proof of Theorem 2. ∎

REFERENCES

[1] B.-N. Vo et al., "Joint detection and estimation of multiple objects from image observations," *IEEE Trans. Sig. Proc.*, 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, C. Wu, "An optimal retrospective change point detection policy," *Scand. J. Stat.*, vol. 36, no. 3, pp. 542-558, Sept. 2009.

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

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

[6] G. V. Moustakides, "Finite sample size optimality of GLR tests," *arXiv:0903.3795*, Nov. 2009.

[7] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed., Springer, 1998.

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

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

[10] E. Fishler, A. Haimovich, and R. S. Blum, "Spatial diversity in radar-models and detection performance," *IEEE Trans. Sig. Proc.*, vol. 54, no. 3, pp. 823-837, March 2006.

[11] M. Skolnik, *Introduction to Radar Systems*, 3rd ed., McGraw-Hill, 2002.

[12] A. Tajer, G. Jajamovich, X. Wang, and G. V. Moustakides, "Optimal joint target detection and parameter estimation by mimo radar," *IEEE J. Sel. Top. Sig. Proc.*, vol. 4, no. 1, pp. 127-145, Feb. 2010.

[13] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, New York, 1991.