OPTIMAL SIMULTANEOUS DETECTION AND ESTIMATION UNDER A FALSE ALARM CONSTRAINT

Bülent Baygün* and Alfred O. Hero III

Keywords: simultaneous decisions, fundamental tradeoffs, min-max criterion, order selection, signal classification, signal detection, likelihood ratio.

ABSTRACT

This paper addresses the problem of finite sample simultaneous detection and estimation which arises when estimation of signal parameters is desired but signal presence is uncertain. In general, a joint detection and estimation algorithm cannot simultaneously achieve optimal detection and optimal estimation performance. In this paper we develop a multihypothesis testing framework for studying the tradeoffs between detection and parameter estimation (classification) for a finite discrete parameter set. Our multihypothesis testing problem is based on the worst case detection and worst case classification error probabilities of the class of joint detection and classification algorithms which are subject to a false alarm constraint. This framework leads to the evaluation of greatest lower bounds on the worst case decision error probabilities and a construction of decision rules which achieve these lower bounds. For illustration, we apply these methods to signal detection, order selection, and signal classification for a multi-component signal in noise model. For two or fewer signals, a SNR of 3dB and signal space dimension of $N = 10$ numerical results are obtained which establish the existence of fundamental tradeoffs between three performance criteria: probability of signal detection, probability of correct order selection, probability of correct classification. Furthermore, based on numerical performance comparisons between our optimal decision rule and other suboptimal penalty function methods, we observe that Rissanen's order selection penalty method is nearly min-max optimal.

Associate Editorial areas: Detection; Estimation

*Bülent Baygün (Corresponding Author)
Schlumberger-Doll Research,
Old Quarry Road,
Ridgefield, CT 06877-4108
(203)-431-5524
(203)-438-3819 (FAX)
baygun@sdr.slb.com

Alfred O. Hero III
Dept. of EECS
The University of Michigan
Ann Arbor, MI 48109-2122
(313)-763-0564
(313)-763-8041 (FAX)
hero@eecs.umich.edu

This work was presented in part at ICASSP92, San Francisco, CA, March 23-26, 1992.

List of Tables

List of Figures

1 Introduction

Many statistical decision problems in engineering applications fall into one of two categories: detection and point estimation. In the *detection* problem an observed random quantity may consist of "noise alone" or "signal masked by noise"; the objective is to decide if there is a signal in the observation subject to a constraint on false alarm. In the *point estimation* problem a signal which is known to be present in the observations has an unknown feature represented by a parameter; the objective is to decide on the parameter value. However, one frequently encounters applications where estimation has to be performed under uncertainty of signal presence. These include applications such as fault detection and diagnosis in dynamical system control [23], target detection and direction finding with an array of sensors [26], image and speech segmentation [13], and digital communications [17]. The associated decision problem is called *simultaneous* or *joint detection and estimation.*

If we constrain the probability of false alarm to be equal to α, one can consider two approaches to the design of decision rules for joint detection and estimation. The first is the simple *uncoupled design strategy* where detection performance is optimized under the false alarm constraint and the estimator is gated by this optimal detector. In this case, one can implement a conditionally optimal estimator which produces an estimate only if the optimal detector decides that the signal is present. While this uncoupled strategy guarantees optimal detection performance, in general there is no guarantee that the gated estimation performance will be acceptable. The second approach is the *coupled* design strategy where estimation performance is directly optimized under the false alarm constraint. As in the uncoupled design, the false alarm constraint prescribes a gated estimator. However, while this gating is optimal for estimation, unlike the uncoupled design it is generally not optimal for detection. Note that under both the coupled and uncoupled strategies the false alarm probabilities are identical. However, while in the uncoupled case the false alarms are generated in such a way as to minimize their impact on detection performance, in the coupled case these false alarms are generated to minimize their impact on estimation performance. The uncoupled strategy provides an upper bound on the detection performance while the coupled strategy provides an upper bound on estimation performance. By comparing the detection/estimation performance of the uncoupled detection-optimal strategy to the detection/estimation performance of the coupled estimation-optimal strategy we can study the fundamental tradeoff between optimal detection and optimal estimation subject to a false alarm constraint.

This paper provides a framework for studying the tradeoffs between detection and estimation based on the worst case detection and worst case estimation error probabilities of the class of simultaneous detection and estimation rules for a finite discrete parameter space. We then formulate and solve a constrained min-max multihypothesis testing problem with non-standard cost structure. This gives the form for the optimal estimator and optimal detector and gives tight lower bounds on the worst case estimation and detection error probabilities which can be used to study tradeoffs.

To illustrate our results, we focus on the following multi-component signal in noise model. A measured waveform Y consists of either a compound signal in additive noise, or noise alone. If present, the signal is the sum of p randomly scaled waveforms (components), out of a possible N equal-power orthogonal waveforms $\{S_1, \ldots, S_N\}$ which are known a priori. That is, the signal is known to lie in a N-dimensional subspace, called the *signal space*, whose basis is $\{S_1, \ldots, S_N\}$.

Hence the observation model has the form:

$$Y = \sum_{l=1}^{p} a_l S_{i_l} + W. \tag{1}$$

Here both the number p and the identity (indices) of the p signal components $\{S_{i_1}, \ldots, S_{i_p}\}$ are unknown. Assume that it is known *a priori* that p is upper bounded by some given constant P, $P \leq N$. We define three related objectives: i) *signal detection* which is to decide if $p > 0$; ii) *signal power estimation* (order selection) which, if $p > 0$, is to specify the actual number $p \in \{1, \ldots, P\}$ of signal components; and iii) *signal component estimation* (classification) which, if $p = p_o > 0$, is to identify the p_o signal components present. These objectives arise in a number of applications including telecommunications, harmonic retrieval, surveillance, and air traffic control.

In the context of the multi-component signal model (1), our results yield the following structure for the optimal constrained rules. The optimal constrained classifier uses a set of $M = \sum_{p=1}^{P} \binom{N}{p}$ likelihood ratios (one for each hypothesized set $\{S_{i_1}, \ldots, S_{i_p}\}$ of signal components, $i_1, \ldots, i_p \in \{1, \ldots, N\}, p = 1, \ldots, P$) to implement a weighted generalized likelihood ratio test, with randomized threshold, followed by a weighted maximum likelihood estimator. The optimal constrained order selector uses a set of P weighted averages of $\binom{N}{p}$ likelihood ratios, $p = 1, \ldots, P$, each average corresponding to a fixed number p of signal components. The optimal constrained detector compares a weighted average of all M likelihood ratios to a threshold. In each of the above three cases the weights and the detection threshold are determined by: 1) the solution to a related non-linear optimization problem; and 2) the false alarm constraint α.

We show that the optimal constrained classifier in the multiple-component signal example (1) has an equivalent form: compare the maximum of the sum of the log-likelihood function and an optimal penalty function of p to a threshold and if the threshold is exceeded use this penalized log-likelihood to perform maximum likelihood estimation. This penalized likelihood structure is closely related to Akaike's AIC [26], and Rissanen's MDL [18] order selection criteria. The common feature is that the optimal constrained classifier, AIC, and MDL all penalize the log-likelihood for overestimation of p. Unlike the AIC and MDL penalties, the penalty associated with the optimal constrained classifier ensures optimal worst case estimation performance in the finite sample regime. Furthermore, this "optimal penalty" takes specific account of a false alarm constraint. We perform a numerical study in which we construct the optimal weight functions for optimal detection. order selection, and classification, implement the optimal likelihood ratio tests, and analyze the relative performances for the case of $p = 2$ or fewer signal components. In this manner we establish the existence of significant tradeoffs between optimal detection, optimal estimation, and optimal order selection. This study also establishes the remarkable result that the MDL order selection penalty is nearly optimal, in the sense of achieving the finite sample min-max constrained classification performance attained with our optimal penalty function, when SNR is 3dB, signal space dimension is $N = 10$, and the number of independent snapshots is between 18 and 26.

Relation to Previous Work

Optimal coupled design strategies for detection and estimation have been studied by only a few authors. Pioneering works along the lines of coupled design in simultaneous detection and estimation include the papers by Middleton and Esposito [14], Fredriksen *et.al* [7] and Birdsall and Gobien [3]. The common ground in each of these studies is the Bayesian viewpoint; that is, the parameters are assigned prior probabilities so that average performance can be optimized. Kelly *et.al* [10, 11] studied the problem of simultaneous detection and estimation using a combination of a generalized likelihood ratio test and a maximum likelihood classifier. They noted that this strategy is optimal only for certain cases; our work reinforces this point by specifying conditions for optimality of their strategy. Stuller [22] extended the generalized likelihood ratio test approach to multiple composite hypothesis testing, by breaking the problem into a sequence of binary composite hypothesis tests. He provided rather stringent sufficient conditions for min-max optimality of this strategy, pointing out that the question of min-max optimality in the general case is yet to be investigated. The min-max multiple hypothesis testing strategy presented in our paper can also be interpreted as a sequence of binary composite hypothesis tests, thereby providing a link to Stuller's paper and establishing the structure of optimal sequential binary tests.

An outline of the paper is as follows. Section 2 introduces the statistical framework that will be used in this paper. Section 3 provides theoretical results whose proofs are contained in the Appendix. In Section 5 we specialize the theory to three different problems: outlier detection and identification, detection and classification of a step change, and detection and parameter estimation of a multi-component signal in noise.

2 Problem Statement

A parametric statistical experiment [9] is defined as the indexed probability space $(\Omega, \sigma, P_\theta)$ where θ is a parameter lying in a parameter space Θ, Ω is the set of possible outcomes of the experiment, σ is a sigma algebra consisting of subsets of Ω, and P_θ is a probability measure defined on σ. The parameter space Θ summarizes all of the uncertainty in the probability model P_θ for the experiment. It is important to emphasize that θ is a fixed non-random parameter.

Define the finite partition, called a $(J+1)$*-ary partition*, $\{\Theta_0, \ldots, \Theta_J\}$ of Θ. For fixed $\theta = \theta_{true}$, denoted the "true θ", let X be a random variable defined on Ω and taking values in a set $\mathcal{X}$ called the observation space. We assume that X has a probability density function $f_\theta(x)$ with respect to some dominating measure μ. Let θ_{true} be contained in partition element Θ_j for a particular $j \in \{0, \ldots, J\}$. The objective is to correctly decide on the partition element Θ_j containing θ_{true} based on a realization $X(\omega) = x$ of $\mathbf{X}$. We can express this classification problem in terms of testing between the $J+1$ exhaustive and mutually exclusive hypotheses [24]:

$$\begin{array}{lll} \underline{H}_0: & X \sim f_\theta, & \theta \in \Theta_0 \\ \vdots & & \vdots \\ \underline{H}_J: & X \sim f_\theta, & \theta \in \Theta_J \end{array} . \qquad (2)$$

When θ_{true} is contained in Θ_j the hypothesis $\underline{H}_j$ is said to be true and the other hypotheses are said to be false. In this case $\underline{H}_j$ is said to be the "true state of nature." If the partition elements $\Theta_0, \ldots, \Theta_J$ are single-point sets, then the hypotheses (2) are called simple hypotheses. Otherwise, if a partition element Θ_l consists of more than one point θ then specification of $\underline{H}_l$ does not specify a unique distribution P_θ and $\underline{H}_l$ is called a composite hypothesis. A simple hypothesis will be identified by the absence of an underscore, e.g. H_l.

We specialize our treatment to the case of a discrete parameter space Θ with $K+1$ elements denoted by indices $\{0, \ldots, K\}$. We will assume that Θ_0 corresponds to the set of $K - M + 1$ parameters $\Theta_0 = \{0, \ldots, K - M\}$ where M is a positive integer less than or equal to K. We identify two special partitions which will play an important role in the sequel. The binary partition,

$\{\Theta_0, \Theta_1\}$ where $\Theta_1 = \{K-M+1, \ldots, K\}$, which specifies a composite *detection* problem:

$$\begin{array}{lll} \underline{H}_0: & X \sim f_\theta, & \theta \in \{0, \ldots, K-M\} \\ \underline{H}_1: & X \sim f_\theta, & \theta \in \{K-M+1, \ldots, K\} \end{array}, \tag{3}$$

where $\underline{H}_0$ is called the *null hypothesis* and $\underline{H}_1$ is called the *alternative hypothesis*. The $(M+1)$-ary partition, $\{\Theta_0, \Theta_1, \ldots, \Theta_M\}$, where $\Theta_1, \ldots, \Theta_M$ are the single-point sets $\{K-M+1\}, \ldots, \{K\}$, respectively, specifies a *joint detection-classification* problem with simple alternatives:

$$\begin{array}{lll} \underline{H}_0: & X \sim f_\theta, & \theta \in \{0, \ldots, K-M\} \\ H_1: & X \sim f_\theta, & \theta = K-M+1 \\ \vdots & \vdots & \\ H_M: & X \sim f_\theta, & \theta = K \end{array}. \tag{4}$$

The primary difference between detection (3) and joint detection-classification (4) is that decision strategies for detection can only be penalized for erroneously deciding on the composite alternative $\underline{H}_1$ while decision strategies for joint detection-classification can bear an additional penalty for erroneous classification among the alternatives $H_1, \ldots, H_M$.

The set of decision strategies for the general $(J+1)$-ary hypothesis testing problem (2) is specified by the set of test functions [24].

Definition 1 *A test function $\underline{\phi} = [\phi_0, \ldots, \phi_J]^T$ for the multiple hypotheses $\underline{H}_0, \ldots, \underline{H}_J$ is a $(J+1)$-dimensional vector function on $\mathcal{X}$ such that $\underline{\phi}(x) \in [0,1]^{(J+1)}$ and $\sum_{j=0}^J \phi_j(x) = 1$, $\forall x \in \mathcal{X}$.*

For a given realization $X = x$, $\phi_j(x)$ is the conditional probability of deciding $\underline{H}_j$. Consequently, $1 - \phi_j(x)$ is the conditional probability of not deciding $\underline{H}_j$ and $\phi_j(x) + \phi_k(x)$ is the conditional probability of deciding either $\underline{H}_j$ or $\underline{H}_k$. The summation condition $\sum_{j=0}^J \phi_j(x) = 1$ ensures that exactly one of $\underline{H}_0, \ldots, \underline{H}_J$ must be decided.

Let $\underline{\phi} = [\phi_0, \phi_1, \ldots, \phi_M]^T$ be an arbitrary test function for testing among the hypotheses $\underline{H}_0, H_1, \ldots, H_M$. This test function defines a simultaneous detection-classification rule. Specifically, since detection is a binary decision between $H_0 : \theta \in \Theta_0$ and $\underline{H}_1 : \theta \in \overline{\Theta}_0$, where $\overline{\Theta}_0 = \Theta - \Theta_0 \bigcup_{k=1}^M \Theta_k$, the first element ϕ_0 of $\underline{\phi}$ specifies a binary test function $\underline{\phi}^D$ for detection:

$$\begin{aligned} \underline{\phi}^D &= [\phi_0, \sum_{k=1}^M \phi_k]^T \\ &= [\phi_0, 1 - \phi_0]^T. \end{aligned} \tag{5}$$

On the other hand, define $\mathcal{X}^{\underline{H}_1}$ as the set of x for which $\phi_0(x) \neq 1$, that is for $X = x \in \mathcal{X}^{\underline{H}_1}$ the decision $\underline{H}_1$ occurs with non-zero probability. Then $\underline{\phi}$ specifies an M-ary test function $\underline{\phi}^C : \mathcal{X}^{\underline{H}_1} \to \mathcal{C}_M$ for classification:

$$\underline{\phi}^C(x) = [\frac{1}{1-\phi_0(x)}\phi_1(x), \ldots, \frac{1}{1-\phi_0(x)}\phi_M(x)]^T, \quad x \in \mathcal{X}^{\underline{H}_1}, \tag{6}$$

where $\frac{\phi_j(x)}{1-\phi_0(x)}$ is the conditional probability of classifying θ into $\Theta_j = \{\theta_j\}$ given that $X = x \in \mathcal{X}^{\underline{H}_1}$.

Conversely, if a test function $\underline{\phi}^D = [\phi_0^D, 1-\phi_0^D]^T$ for detection and a test function $\underline{\phi}^C = [\phi_1^C, \ldots, \phi_M^C]^T$ for classification are available, a simultaneous detection-classification rule $\underline{\phi} = [\phi_0, \ldots, \phi_M]^T$ is easily constructed via the identification:

$$\underline{\phi} = [\phi_0^D, (1-\phi_0^D)\phi_1^C, \ldots, (1-\phi_0^D)\phi_M^C]^T. \tag{7}$$

We call $\underline{\phi}$ a "gated" classification rule since the classification rule $\underline{\phi}^C$ is enabled by the detection rule ϕ_0^D when $1-\phi_0^D \neq 0$, i.e. signal detection can occur with non-zero probability.

The average performance of a particular test function $\underline{\phi}$ is determined by: i) probability of *false alarm* $P_\theta(FA)$ ii) probability of *miss* $P_\theta(M)$; and iii) probability of *erroneous classification* $P_\theta(EC)$:

$$\begin{array}{rcll} P_\theta(FA) & = & E_\theta[1-\phi_0], & \theta \in \Theta_0, \\ P_\theta(M) & = & E_\theta[\phi_0], & \theta \notin \Theta_0 \quad , \\ P_\theta(EC) & = & E_\theta[1-\phi_{r_J(\theta)}], & \theta \notin \Theta_0 \end{array} \tag{8}$$

where $r_J(\theta) \in \{0, \ldots, J\}$ is the *set partition function* which takes the value j if $\theta \in \Theta_j$.

We will be interested in those test functions whose false alarm probability $P_\theta(FA)$ is less than or equal to a prespecified constant $\alpha \in [0,1]$ [24].

Definition 2 *A test function $\underline{\phi}$ is* of level α *if*

$$\max_{\theta\in\Theta_0} E_\theta[1-\phi_0] \leq \alpha, \tag{9}$$

for a specified $\alpha \in [0,1]$.

The classical Neyman-Pearson criterion of signal detection [12] states that it is desirable to minimize the miss probability $P_\theta(M)$, $\theta \notin \Theta_0$, subject to the constraint (9). On the other hand, in terms of signal classification, minimizing $P_\theta(EC)$, $\theta \notin \Theta_0$ is desirable. However, since $P_\theta(M)$ and $P_\theta(EC)$ generally vary as a function of θ, θ-uniform minimization of these probabilities is in general impossible and a different approach must be taken.

3 Constrained Min-Max Tests

For the purposes of establishing θ-uniform lower bounds on $P_\theta(M)$ and $P_\theta(EC)$ it makes sense to consider the form and performance of constrained min-max test functions of level α. Define the set $\mathcal{D}_\alpha$ of all test functions $\underline{\phi} = [\phi_0, \ldots, \phi_J]^T$ of level α by:

$$\mathcal{D}_\alpha \stackrel{def}{=} \{\underline{\phi} : \mathcal{X} \mapsto [0,1]^{(J+1)}, \sum_{j=0}^{J} \phi_j = 1, \text{ and } \max_{\theta\in\Theta_0} E_\theta[1-\phi_0] \leq \alpha\}. \tag{10}$$

Definition 3 *A test function $\underline{\phi}^* = [\phi_0^*, \ldots, \phi_J^*]^T$ is a constrained min-max test of level α between the hypotheses $\underline{H}_0, \ldots, \underline{H}_J$ if $\underline{\phi}^* \in D_\alpha$, i.e.*

$$\max_{\theta\in\Theta_0} E_\theta[1-\phi_0^*] \leq \alpha, \tag{11}$$

and if for any other test function $\underline{\phi} = [\phi_0, \ldots, \phi_J]^T \in D_\alpha$

$$\max_{\theta\notin\Theta_0} E_\theta[1-\phi_{r_J(\theta)}^*] \leq \max_{\theta\notin\Theta_0} E_\theta[1-\phi_{r_J(\theta)}]. \tag{12}$$

Observe that, if a constrained min-max test $\underline{\phi}^*$ of level α can be found, the left hand side of (12) provides an achievable lower bound on the maximum error probability $\max_{\theta\notin\Theta_0} E_\theta[1-\phi_{r_J(\theta)}]$ of any level α test.

The first step in deriving the form of constrained min-max tests $\underline{\phi}^*$ for the hypotheses in (2) is to show that a composite null hypothesis $\underline{H}_0$ can be reduced to an equivalent simple null hypothesis. Define the K-dimensional unit simplex $\mathcal{C}_K$:

$$\mathcal{C}_K = \{\underline{p} \in [0,1]^K : \sum_{j=0}^{K} p_j = 1\}.$$

For any weight vectors $\underline{b} = [b_0, \ldots, b_{K-M}]^T \in \mathcal{C}_{K-M+1}$ and $\underline{c} = [c_1, \ldots, c_M]^T \in \mathcal{C}_M$, define

$$f_0^{(\underline{b})} \stackrel{def}{=} \sum_{\theta\in\Theta_0} b_\theta f_\theta. \tag{13}$$

$$q_j \stackrel{def}{=} \sum_{\theta\in\Theta_j} c_\theta, \; j = 1, \ldots, J, \tag{14}$$

$$f_j^{(\underline{c})} \stackrel{def}{=} \sum_{\theta\in\Theta_j} (c_\theta/q_j) f_\theta, \; j = 1, \ldots, J. \tag{15}$$

The weights $\{b_\theta\}_{\theta\in\Theta_0}$ can be regarded as unit normalized weights on the null states of nature $\theta \in \Theta_0$; the weights $\{q_j\}_{j=1}^J$ can be regarded as unit normalized weights on the composite states of nature $\{\Theta_j\}_{j=1}^J$; and the weights $\{c_\theta/q_j\}_{\theta\in\Theta_j}$ can be regarded as unit normalized weights on the states of nature $\theta \in \Theta_j$.

Consider the following reduced hypotheses:

$$\begin{array}{rlll} H_0^{(\underline{b})}: & X \sim f_0^{(\underline{b})} & & \\ \underline{H}_1: & X \sim f_\theta, & \theta \in \Theta_1 & \\ \vdots & \vdots & & . \\ \underline{H}_J: & X \sim f_\theta, & \theta \in \Theta_J & \end{array} \tag{16}$$

Note that relative to (2) the null hypothesis in (16) has been reduced to a simple null hypothesis. Define the expectation $E_0^{(\underline{b})}[g(X)]$ of $g(X)$ under the simple hypothesis $H_0^{(\underline{b})}$:

$$E_0^{(\underline{b})}[g(X)] \stackrel{def}{=} \int_{\mathcal{X}} g(x) f_0^{(\underline{b})}(x) d\mu(x) = \int_{\mathcal{X}} g(x) \sum_{\theta\in\Theta_0} b_\theta f_\theta(x) d\mu(x) \tag{17}$$

The following theorem is proven in the Appendix.

Theorem 1 *For arbitrary $\underline{b} \in \mathcal{C}_{K-M+1}$, let $\underline{\phi}^{(\underline{b})} = [\phi_0^{(\underline{b})}, \ldots, \phi_J^{(\underline{b})}]^T$ be a constrained min-max test of level α for testing among the hypotheses (16) with simple null hypothesis $H_0^{(\underline{b})}$. If there exists a weight vector $\underline{b} = \underline{b}^*$ such that:*

$$\max_{\theta\in\Theta_0} E_\theta[1 - \phi_0^{(\underline{b}^*)}] = \alpha, \tag{18}$$

then $\underline{\phi}^* \stackrel{def}{=} \underline{\phi}^{(\underline{b}^*)}$ *is a constrained min-max test of level* α *for testing among the hypotheses (2) with composite null hypothesis* $\underline{H}_0$*. Furthermore such a* $\underline{b}^*$ *exists if*

$$E_0^{(\underline{b}^*)}[1 - \phi_0^{(\underline{b}^*)}] = \alpha \tag{19}$$

and $\underline{b}^*$ *is a "least favorable prior distribution" in the sense that for any other* $\underline{b} \in \mathcal{C}_{K-M+1}$*:*

$$\int_{\mathcal{X}} [1 - \phi_0^{(\underline{b}^*)}(x)] f_0^{(\underline{b})}(x) d\mu(x) \leq \int_{\mathcal{X}} [1 - \phi_0^{(\underline{b}^*)}(x)] f_0^{(\underline{b}^*)}(x) d\mu(x). \tag{20}$$

The condition (18) says that for a specific $\underline{b}^*$ the level α constrained min-max test $\underline{\phi}^{(\underline{b}^*)}$ for the reduced hypotheses (16) must also be of level α for the original hypotheses (2). Under this condition Theorem 1 states that the composite null hypothesis $\underline{H}_0$ can be reduced to simple null hypothesis $H_0^{(\underline{b})}$ by a $\underline{b}^*$ weighting of the f_θ over $\theta \in \Theta_0$. Once such a reduction is achieved we need merely consider constrained min-max tests for the hypotheses (16) with simple null hypothesis $H_0^{(\underline{b})}$ and then select appropriate $\underline{b}^*$ to satisfy condition (18). The existence of a weight vector $\underline{b}^*$ which satisfies the sufficient conditions (19)-(20) is related to the existence of a detector having *constant false alarm rate* (CFAR) [20].

The following theorem, proven in Appendix, specifies the form of constrained min-max tests for the set of hypotheses:

$$\begin{array}{lll} H_0: & X \sim f_0 & \\ \underline{H}_1: & X \sim f_\theta, & \theta \in \Theta_1 \\ \vdots & \vdots & \\ \underline{H}_J: & X \sim f_\theta, & \theta \in \Theta_J \end{array}, \tag{21}$$

where f_0 is an arbitrary pdf, e.g. $f_0 = f_0^{(\underline{b}^*)}$.

Theorem 2 *Fix the level* $\alpha \in [0,1]$*. For arbitrary* $\underline{c} \in \mathcal{C}_M$*, define* q_j *and* $f_j^{(\underline{c})}$ *as in (14)-(15). Let* $j_{max} \stackrel{def}{=} \arg\max_{j>0} q_j f_j^{(\underline{c})}$ *and define the test function* $\underline{\phi}^{(\underline{c})} = [\phi_0^{(\underline{c})}, \ldots, \phi_J^{(\underline{c})}]^T$ *by the following assignments:*

$$\phi_0^{(\underline{c})}(x) = \left\{ \begin{array}{ll} 1 & , \textit{if } \max_{j>0} \left\{ q_j f_j^{(\underline{c})}(x) \right\} < \lambda f_0(x) \\ \xi & , \textit{if } \max_{j>0} \left\{ q_j f_j^{(\underline{c})}(x) \right\} = \lambda f_0(x) \\ 0 & , \textit{if } \max_{j>0} \left\{ q_j f_j^{(\underline{c})}(x) \right\} > \lambda f_0(x) \end{array} \right. , \tag{22}$$

and for $j = 1, \ldots, J$*:*

$$\phi_j^{(\underline{c})}(x) = \left\{ \begin{array}{ll} 1 & , \textit{if } \max_{j>0} \left\{ q_j f_j^{(\underline{c})}(x) \right\} > \lambda f_0(x) \textit{ and } j = j_{max} \\ 1-\xi & , \textit{if } \max_{j>0} \left\{ q_j f_j^{(\underline{c})}(x) \right\} = \lambda f_0(x) \textit{ and } j = j_{max} \\ 0 & , \textit{else} \end{array} \right. , \tag{23}$$

where $\lambda \geq 0$ *and* $\xi \in [0,1]$ *are functions of* $\underline{c}$ *selected to satisfy the constraint on the false alarm probability:*

$$E_0[1 - \phi_0^{(\underline{c})}] = \alpha. \tag{24}$$

Then there exists a weight vector $\underline{c} = \underline{c}^*$, called the "optimal weight vector", for which:

$$\sum_{\theta \notin \Theta_0} c^*_\theta E_\theta[1 - \phi^{(\underline{c}^*)}_{r_J(\theta)}] = \max_{\underline{c} \in \mathcal{C}_M} \sum_{\theta \notin \Theta_0} c_\theta E_\theta[1 - \phi^{(\underline{c})}_{r_J(\theta)}]. \tag{25}$$

and $\underline{\phi}^* \stackrel{def}{=} \underline{\phi}^{(\underline{c}^*)}$ defined by the relations (22)-(25) is a constrained min-max test of level α for testing among the hypotheses $H_0, \underline{H}_1, \ldots, \underline{H}_J$.

Next we give a corollary which specifies the form of the constrained min-max tests for composite hypotheses $\underline{H}_0, \ldots, \underline{H}_J$ by combining the results of Theorems 1 and 2.

Corollary 1 *Fix the level $\alpha \in [0,1]$. For arbitrary $\underline{c} \in \mathcal{C}_M$ and $\underline{b} \in \mathcal{C}_{K-M+1}$, let $f_0^{(\underline{b})}$, q_j, and $f_j^{(\underline{c})}$, $j = 1, \ldots, J$, be as defined in (13)-(15). Let $j_{max} = \arg\max_{j>0} q_j f_j^{(\underline{c})}(x)$ and define the test function $\underline{\phi}^{(\underline{b},\underline{c})} = [\phi_0^{(\underline{b},\underline{c})}, \ldots, \phi_J^{(\underline{b},\underline{c})}]^T$ by the following assignments:*

$$\phi_0^{(\underline{b},\underline{c})}(x) = \begin{cases} 1 & , \text{ if } \max_{j>0}\left\{q_j f_j^{(\underline{c})}(x)\right\} < \lambda f_0^{(\underline{b})}(x) \\ \xi & , \text{ if } \max_{j>0}\left\{q_j f_j^{(\underline{c})}(x)\right\} = \lambda f_0^{(\underline{b})}(x) \\ 0 & , \text{ if } \max_{j>0}\left\{q_j f_j^{(\underline{c})}(x)\right\} > \lambda f_0^{(\underline{b})}(x) \end{cases}, \tag{26}$$

and for $j = 1, \ldots, J$:

$$\phi_j^{(\underline{b},\underline{c})}(x) = \begin{cases} 1 & , \text{ if } j = j_{max} \text{ and } \max_{j>0}\left\{q_j f_j^{(\underline{c})}(x)\right\} > \lambda f_0^{(\underline{b})}(x) \\ 1-\xi & , \text{ if } j = j_{max} \text{ and } \max_{j>0}\left\{q_j f_j^{(\underline{c})}(x)\right\} = \lambda f_0^{(\underline{b})}(x) \\ 0 & , \text{ else} \end{cases}, \tag{27}$$

where $\lambda \geq 0$ and $\xi \in [0,1]$ are functions of $\underline{b}$ and $\underline{c}$ determined by the false alarm constraint:

$$E_0^{(\underline{b})}[1 - \phi_0^{(\underline{b},\underline{c})}] = \int_{\mathcal{X}} (1 - \phi_0^{(\underline{b},\underline{c})}(x)) f_0^{(\underline{b})}(x) dx = \alpha \tag{28}$$

Then for each weight vector $\underline{b}$, there exists a weight vector $\underline{c} = \underline{c}^*$ for which

$$\sum_{\theta \notin \Theta_0} c^*_\theta E_\theta[1 - \phi^{(\underline{b},\underline{c}^*)}_{r_J(\theta)}] = \max_{\underline{c} \in \mathcal{C}_M} \sum_{\theta \notin \Theta_0} c_\theta E_\theta[1 - \phi^{(\underline{b},\underline{c})}_{r_J(\theta)}], \tag{29}$$

and $\underline{\phi}^{(\underline{b},\underline{c}^*)}$ is a constrained min-max test of level α for testing among the hypotheses (16) with simple null hypothesis $H_0^{(\underline{b})}$. Furthermore, if there exists a weight vector $\underline{b} = \underline{b}^*$ for which

$$\max_{\theta \in \Theta_0} E_\theta[1 - \phi_0^{(\underline{b}^*,\underline{c}^*)}] = \alpha \tag{30}$$

then $\underline{\phi}^* \stackrel{def}{=} \underline{\phi}^{(\underline{b}^*,\underline{c}^*)}$ defined by the relations (26)-(30) is a constrained min-max test of level α for testing among the hypotheses (2) with composite null $\underline{H}_0$.

A simpler shorthand notation for the test (26)-(30) to be used in the sequel is:

$$\max_{j>0}\left\{q_j^* \frac{f_j^{(\underline{c}^*)}(X)}{f_0^{(\underline{b}^*)}(X)}\right\} \quad \underset{\underline{H}_0}{\overset{\underline{H}_{j_{max}}}{\gtrless}} \quad \lambda^* \tag{31}$$

which is read as: if the left hand side exceeds the threshold λ^* t0hen decide $\underline{H}_{j_{max}}$ where $j_{max} \stackrel{def}{=} argmax_{j>0}\left\{q_j^* f_j^{(\underline{c}^*)}(X)/f_0^{(\underline{b}^*)}(X)\right\}$. In (31) the threshold is written as $\lambda^* \stackrel{def}{=} \lambda(\underline{b}^*, \underline{c}^*)$ to emphasize its dependency on the optimal weights. The test (31) can be recognized as a weighted generalized likelihood test (GLRT) constructed with the weighted average densities $f_0^{(\underline{b}^*)} = \sum_{\theta\in\Theta_0} b_\theta^* f_\theta$ and $f_j^{(\underline{c}^*)} = \sum_{\theta\in\Theta_j} c_\theta^* f_\theta$, $j = 1, \ldots, J$. The test performs detection at a given level α and if $\underline{H}_0$ is rejected, it classifies in such a way as to select the hypothesis $\underline{H}_{j_{max}}$ which maximizes the weighted likelihood ratio $q_j^* \frac{f_j^{(\underline{c}^*)}}{f_0^{(\underline{b}^*)}}$.

The next Corollary, proven in Appendix, specifies sufficient conditions on the weight vectors $\underline{b}$ and $\underline{c}$ such that the test $\underline{\phi}^{(\underline{b},\underline{c})}$ defined by (26)-(28) be a constrained min-max test. As such, it provides us with a verification condition as an alternative to the constructive definition (29)-(30) of the weight vectors $\underline{b}^*, \underline{c}^*, \underline{q}^*$.

Corollary 2 *For arbitrary $\underline{c} \in \mathcal{C}_M$ and $\underline{b} \in \mathcal{C}_{K-M+1}$, let $\underline{\phi}^{(\underline{b},\underline{c})} = [\phi_0^{(\underline{b},\underline{c})}, \ldots, \phi_J^{(\underline{b},\underline{c})}]^T$ be of the form given in (26)-(28). Suppose that there exist weight vectors $\underline{c}^* \in \mathcal{C}_M$ and $\underline{b}^* \in \mathcal{C}_{K-M+1}$ and a constant V such that:*

$$E_\theta[1 - \phi_0^{(\underline{b}^*,\underline{c}^*)}] = \alpha, \quad \forall\theta \in \Theta_0 \tag{32}$$

and

$$E_\theta[1 - \phi_{r_J(\theta)}^{(\underline{b}^*,\underline{c}^*)}] = V, \quad \forall\theta \notin \Theta_0 \tag{33}$$

Then $\underline{\phi}^* \stackrel{def}{=} \underline{\phi}^{(\underline{b}^*,\underline{c}^*)}$ is a constrained min-max test of level α for testing $\underline{H}_0, \ldots, \underline{H}_J$.

Corollary 2 states that the likelihood ratio test (31) is a constrained min-max test of level α provided one can find weights $\underline{b}, \underline{c}$ such that: i) the false alarm constraint $E_\theta[1 - \phi_0^{(\underline{b},\underline{c})}] \le \alpha$ is satisfied with equality uniformly for all elements $\theta \in \Theta_0$; and ii) the probability of erroneous classification $E_\theta[1 - \phi_{r_J(\theta)}^{(\underline{b},\underline{c})}]$ is constant over $\theta \notin \Theta_0$. Such a test is called an "equalizer rule" since it equalizes the decision error probabilities over the alternatives $\underline{H}_1, \ldots, \underline{H}_J$.

Combining Corollaries 1 and 2, the search for the constrained min-max test consists of two steps: i) construct the test function $\underline{\phi}^{(\underline{b},\underline{c})}$ of (26)-(28); ii) vary $\underline{b}$ and $\underline{c}$ maintaining the false alarm constraint for all $\theta \in \Theta_0$ until equalization of $E_\theta[1 - \phi_{r_J(\theta)}^{(\underline{b},\underline{c})}]$, $\theta \notin \Theta_0$ is achieved.

Corollary 1 provides a lower bound for the worst case probability of error of any other level-α test function $\underline{\phi}$. The following corollary formalizes this point and will be used in the next section to assess the detection versus classification tradeoff for joint detection and estimation in the multiple-component signal model.

Corollary 3 *Fix the level $\alpha \in [0, 1]$. Then for any level-α test function $\underline{\phi} = [\phi_0, \ldots, \phi_J]^T$, its worst case error probability over $\theta \notin \Theta_0$ satisfies the lower bound:*

$$\max_{\theta\notin\Theta_0} E_\theta[1 - \phi_{r_J(\theta)}] \ge V \stackrel{def}{=} \max_{\theta\notin\Theta_0} E_\theta[1 - \phi_{r_J(\theta)}^*] \tag{34}$$

where $\phi_{r_J(\theta)}^*$ is as defined in (26)-(29).

4 Discussion

Remark 1: Define the $(K+1)$-dimensional probability vector $\underline{p}^* = [p_0^*, \ldots, p_K^*]^T$, $\sum_{k=0}^K p_k^* = 1$:

$$p_\theta^* \stackrel{def}{=} \left\{ \begin{array}{ll} \frac{\lambda^* b_\theta^*}{1+\lambda^*}, & \theta \in \Theta_0 \\ \\ \frac{c_\theta^*}{1+\lambda^*}, & \theta \notin \Theta_0 \end{array} \right. .$$

Using the identities (13) and (15), the constrained min-max test $\underline{\phi}^*$ (31) is equivalent to:

$$\max_{j>0} \left\{ \frac{\sum_{\theta \in \Theta_j} c_\theta^* f_\theta(X)}{\lambda^* \cdot \sum_{\theta \in \Theta_0} b_\theta^* f_\theta(X)} \right\} \begin{array}{c} \underline{H}_{j_{max}} \\ > \\ < \\ \underline{H}_0 \end{array} 1.$$

Now dividing numerator and denominator by $1+\lambda^*$ we obtain:

$$\max_{j>0} \left\{ \frac{\sum_{\theta \in \Theta_j} p_\theta^* f_\theta(X)}{\sum_{\theta \in \Theta_0} p_\theta^* f_\theta(X)} \right\} \begin{array}{c} \underline{H}_{j_{max}} \\ > \\ < \\ \underline{H}_0 \end{array} 1. \tag{35}$$

Assume that θ is a random variable taking values $\theta_0, \ldots, \theta_M$ in Θ with probabilities $p_0^*, \ldots, p_M^*$. Define an integer valued random variable J by $J = j$ iff $\theta \in \Theta_j$ (i.e. $J = r_J(\theta)$). Then the random variable J has the prior distribution

$$P(J=j) = p(\theta \in \Theta_j) = \sum_{\theta \in \Theta_j} p_\theta^* \tag{36}$$

and the posterior distribution

$$P(J=j|X) = P(\theta \in \Theta_j|X) = \frac{\sum_{\theta \in \Theta_j} p_\theta^* f_\theta(X)}{\sum_{\theta \in \Theta_0} p_\theta^* f_\theta(X)}. \tag{37}$$

Therefore the constrained min-max test (35) is equivalent to the maximum *a posteriori* decision rule: Choose index j of $\underline{H}_j$ such that the posterior probability $P(J=j|X)$ given by (37) is maximized over j. This establishes that the constrained min-max test of level α is Bayes relative to a set of priors $\{p_{\theta^*}\}$ [2].

Remark 2: If $\alpha = 1$ in Corollary 1 then the constraint $\max_{\theta \in \Theta_0} E_\theta[1-\phi_0^*] \leq \alpha$ is satisfied for $\lambda = 0$ and $\xi = 0$ so that $\phi_0^*(X) = 0$ w.p.1. with respect to any set of probability densities $\{f_\theta(x) : \theta \in \Theta_0\}$ on $\mathcal{X}$. Let $\Theta_1, \ldots, \Theta_J$ be singleton sets so that the J alternative hypotheses are simple. In this case, $\sum_{j=1}^J \phi_j^* = 1$ and the test of Corollary 1 reduces to the minmax weighted maximum likelihood estimator [6], where for all $\underline{c} \in \mathcal{C}_J$

$$\phi_j^{(\underline{c})}(x) = \left\{ \begin{array}{ll} 1 & , \text{if } j = j_{max} = \arg\max_{\theta \notin \Theta_0} \{c_\theta f_\theta(x)\} \\ 0 & , \text{otherwise} \end{array} \right. \tag{38}$$

for $j = 1, \ldots, J$, and $\underline{c}^* \in \mathcal{C}_J$ is a vector of non-negative unit normalized weights which satisfy

$$\sum_{\theta \notin \Theta_0} c_\theta^* E_\theta[1 - \phi_\theta^{(\underline{c}^*)}] = \max_{\underline{c} \in \mathcal{C}_M} \left\{ \sum_{\theta \notin \Theta_0} c_\theta E_\theta[1 - \phi_\theta^{(\underline{c})}] \right\}. \tag{39}$$

Remark 3: For binary hypotheses $J = 1$, $\Theta = \{\Theta_0, \Theta_1\}$ and the constrained min-max test $\underline{\phi}^*$ of Corollary 1 reduces to the well known min-max detector [12]:

$$\frac{\sum_{\theta \in \Theta_1} c^*_\theta f_\theta(X)}{\sum_{\theta \in \Theta_0} b^*_\theta f_\theta(X)} \overset{\underline{H}_1}{\underset{\underline{H}_0}{\gtrless}} \lambda^*. \tag{40}$$

Here $\underline{c}^* \in \mathcal{C}_M$ and $\underline{b}^* \in \mathcal{C}_{K-M}$ are non-negative unity normalized weights which are obtained from the solution of a non-linear optimization problem, and the threshold λ^* and the randomization parameter ξ are selected to meet the false alarm constraint with equality: $E_\theta[1 - \phi_0] = \alpha$, $\theta \in \Theta_0$. In the case of simple null and alternative hypotheses Θ_0 and Θ_1 are singleton sets and the min-max detector reduces to the classical Neyman-Pearson test $f_{\theta_1}(X)/f_{\theta_0}(X) \overset{H_1}{\underset{H_0}{\gtrless}} \lambda$.

It is useful to compare the min-max optimal detector to the popular [25] *ad hoc* generalized likelihood ratio test (GLRT):

$$\frac{\max_{\theta \notin \Theta_0} f_\theta(X)}{\max_{\theta \in \Theta_0} f_\theta(X)} \overset{\underline{H}_1}{\underset{\underline{H}_0}{\gtrless}} \lambda. \tag{41}$$

Note that the GLRT (41) is not a min-max optimal detector except in the unlikely event that the ratio of $\max_{\theta \in \Theta_0} f_\theta$ and $\max_{\theta \in \Theta_1} f_\theta$ is equivalent to the ratio of weighted average densities in (40).
Remark 4: Let $J = M$ and let $\Theta_1, \ldots, \Theta_J$ be single-point sets, i.e. simple alternatives. Specializing Corollary 1 to this case we obtain the constrained min-max classifier:

$$\max_{\theta \notin \Theta_0} \left\{ \frac{c^*_\theta f_\theta(X)}{\sum_{\theta \in \Theta_0} b^*_\theta f_\theta(X)} \right\} \overset{H_{\theta_{max}}}{\underset{\underline{H}_0}{\gtrless}} \lambda^* \tag{42}$$

If H_0 is simple then the constrained min-max classifier (42) is equivalent to the weighted generalized likelihood ratio test and maximum likelihood classifier (GLRT-MLC):

$$\max_{\theta \notin \Theta_0} \left\{ \frac{c^*_\theta f_\theta(X)}{f_0(X)} \right\} \overset{H_{\theta_{max}}}{\underset{\underline{H}_0}{\gtrless}} \lambda^*. \tag{43}$$

If the optimal weight vector turns out to be uniform (i.e. $\underline{c}^* = [1/J, \ldots, 1/J]$) this is equivalent to the standard unweighted GLRT.
Remark 5: In Corollary 2, we specified sufficient conditions such that $\underline{\phi}^{(\underline{b},\underline{c})}$ defined by (26)-(28) be a constrained min-max test. More specifically, these conditions state that if $\underline{\phi}^{(\underline{b},\underline{c})}$ equalizes the decision error probabilities over *all of the alternatives* $\underline{H}_1, \ldots, \underline{H}_J$ and over $\underline{H}_0$, then it is a constrained min-max test. In some cases equalization of all of the decision error probabilities is not possible. More specifically, equalization over the entire set Θ_0 is not possible if the false alarm probabilities for some parameter values in Θ_0 always dominate the false alarm probabilities for the remaining parameter values in Θ_0, no matter what the weights are. Similarly, equalization over the entire set $\Theta - \Theta_0$ is not possible if the classification error probabilities for some parameter values in $\Theta - \Theta_0$ always dominate the classification error probabilities for the remaining parameter values in $\Theta - \Theta_0$, no matter what the weights are. In fact it can be shown [2] that equalization of only the dominating decision error probabilities provides a sufficient condition for optimality, provided the corresponding weight vectors assign zero weights for the remaining parameter values.

Remark 6: The dimension of the weight space over which a search must be performed to determine the optimal weights is the sum of the number of simple alternative hypotheses plus the number of the simple hypotheses composing the null hypothesis. For a composite null hypothesis, this latter number can be very large which severely complicates the computation of the value function. An alternative approach is to compress the composite null hypothesis into a simple null hypothesis by applying invariance principles [19], thereby reducing the number of weights to be determined. These principles involve mapping the observations to a lower dimensional space via a non-invertible transformation which renders the distribution of the resultant data set functionally independent of the unknown null hypothesis parameters. Such use of invariance principles was described in previous work [1]. The invariance approach has the advantage of simplifying the evaluation of the value function but usually at the expense of degradation of performance since it involves non-invertible transformations of the data [6].

Remark 7: In some applications it is possible to efficiently parameterize the weights and significantly reduce the number of unknowns in the weight space, facilitating the search for optimal weights satisfying the conditions of Corollary 2. One important case where such a reduction is possible is the case where the decision problem is *permutation invariant* [2], in which case the distribution of the likelihood ratio is invariant to permutations in the indices of the hypotheses. We make use of a special type of permutation invariance in the multiple-component signal application treated in the next section.

5 Applications

First we briefly discuss a simple application to changepoint joint detection and classification.

5.1 Detection and Classification of Changes in a Distribution

Consider the vector $\underline{X} = [X_1, \ldots, X_N]^T$ of independent random variables with a nominal marginal density $h_0(x)$ and an alternative "outlier" density $h_1(x)$. We say an outlier occurs when some X_i's have undergone a change in distribution from h_0 to h_1. The objective is to detect and identify any outliers. The change detection and classification problem has been addressed in [16] and [5]. It would be very interesting to compare the error performance of the algorithms proposed in these papers to the achievable lower bounds specified by our finite sample min-max decision rules described below.

A. Point Change Problem

Also known as the slippage problem [6], in the point change problem there is at most one outlier in the vector $\underline{X}$ which can occur at indices $1, \ldots, N$. Let θ denote the index $1, \ldots, N$ where this outlier occurs. Thus we have the null hypothesis and the $M = N$ alternative hypotheses:

$$\begin{array}{lll} H_0: & X \sim f_0 = \prod_{i=1}^N h_0(x_i) & \text{(no outliers)} \\ H_1: & X \sim f_1 = h_1(x_1) \prod_{i \neq 1} h_0(x_i) & (\theta = 1) \\ \vdots & \vdots & \vdots \\ H_N: & X \sim f_M = h_1(x_N) \prod_{i \neq N} h_0(x_i) & (\theta = N) \end{array} . \qquad (44)$$

The weighted likelihood ratios are:

$$c_\theta \frac{f_\theta(\underline{X})}{f_0(\underline{X})} = c_\theta \frac{h_1(X_\theta)}{h_0(X_\theta)}, \qquad \theta = 1, \ldots, M.. \tag{45}$$

We will assume that the likelihood ratios have continuous distributions under H_0 so that randomization is not needed to achieve the false alarm constraint. Using (40) in Remark 3 we obtain the detection optimal (DO) decision rule for min-max testing for outliers:

$$\sum_{\theta=1}^{N} c_\theta^D \frac{h_1(X_\theta)}{h_0(X_\theta)} \quad \underset{H_0}{\overset{H_1}{\gtrless}} \quad \lambda^D \tag{46}$$

where the threshold λ^D and the weight vector $\underline{c}^D$ are chosen such that the false alarm constraint is satisfied with equality ($E_0[1-\phi_0^D] = \alpha$) and the average miss probability $\sum_{\theta=1}^N c_\theta^D E_\theta[\phi_0^D]$ is maximized. Likewise the classification optimal (CO) rule for constrained min-max classification of θ over $\{1, \ldots, N\}$ is obtained from (43) in Remark 4 as:

$$\max_{\theta>0} c_\theta^C \frac{h_1(X_\theta)}{h_0(X_\theta)} \quad \underset{H_0}{\overset{H_{\theta_{max}}}{\gtrless}} \quad \lambda^C \ , \tag{47}$$

where the threshold λ^C and the weight vector $\underline{c}^C$ are chosen such that the false alarm constraint is satisfied with equality ($E_0[1-\phi_0^C] = \alpha$) and the average erroneous classification probability $\sum_{\theta=1}^N c_\theta^C E_\theta[1-\phi_\theta^C]$ is maximized.

The optimal weight vectors $\underline{c}^D$ and $\underline{c}^C$ are easily found by using the equalization condition for optimality given in Corollary 2. Let $\underline{c}^D$ be given by the uniform weighting $c_1^D = \ldots = c_N^D = \frac{1}{N}$. The miss probabilities $P_\theta(M)$, $\theta = 1, \ldots, N$, of the resulting detection rule (46) are simply computed:

$$E_\theta[\phi_0^D] = \int_{-\infty}^{\infty} S_0(N\lambda^D - x) g_1(x) dx \qquad \text{(independent of } \theta\text{)}, \tag{48}$$

where g_1 is the probability density function of the likelihood ratio $\frac{h_1(X_\theta)}{h_0(X_\theta)}$ under H_θ and S_0 is the cumulative distribution function of the sum of i.i.d. likelihood ratios $\{\frac{h_1(X_i)}{h_0(X_i)}\}_{i\neq\theta}$ under H_θ. Similarly, let $c_1^C = \ldots = c_N^C = \frac{1}{N}$ and we obtain the erroneous classification probabilities $P_\theta(EC)$, $\theta = 1, \ldots, N$, of the resulting detection/classification rule (47):

$$P_\theta(EC) = E_\theta[1-\phi_\theta^C] = 1 - \int_{N\lambda^C}^{\infty} G_0(x)^{N-1} g_1(x) dx \qquad \text{(independent of } \theta\text{)}, \tag{49}$$

where G_0 is the common cumulative distribution function of each of the $N-1$ likelihood ratios $\frac{h_1(X_i)}{h_0(X_i)}$, $i \neq \theta$, under H_θ.

The uniform weights $\underline{c}^D$ and $\underline{c}^C$ give the two decision rules for detection and detection/classification as, respectively:

$$\textstyle\sum_{\theta=1}^N \frac{h_1(X_\theta)}{h_0(X_\theta)} \quad \underset{H_0}{\overset{H_1}{\gtrless}} \quad N\lambda^D \ ; \qquad \max_{\theta>0} \frac{h_1(X_\theta)}{h_0(X_\theta)} \quad \underset{H_0}{\overset{H_{\theta_{max}}}{\gtrless}} \quad N\lambda^C \tag{50}$$

Now for any $\alpha \in [0,1]$, since the likelihood ratios have continuous distribution under H_0, there exist positive thresholds λ^D and λ^C such that the false alarm probability of each of the above decision rules is equal to α. Furthermore, since the above decision rules equalize $P_\theta(M)$ and $P_\theta(EC)$, respectively, Corollary 2 asserts that these two rules are in fact the DO and CO rules of level α.

The DO rule (50) is a weighted average likelihood ratio test and is not equivalent to the generalized likelihood ratio test (GLRT). The CO rule (50) is a thresholded maximum likelihood classifier equivalent to the GLRT-MLC (43) and is identical to Ferguson's outlier discriminator [6]. Therefore, in this example, the GLRT-MLC rule is min-max optimal and attains the lower bound on the worst case erroneous classification probability.

B. Step Change Problem

In the step change problem the objective is to detect a step change and estimate the time of change θ in the marginal density of X_i, $i = 1, \ldots, N$, where it is hypothesized that $X_i \sim h_0$, $\forall i \leq \theta$ and $X_i \sim h_1$, $\forall i > \theta$, $\theta \in \{1, \ldots, N\}$. Hence we must test between a null hypothesis and $M = N - 1$ simple alternatives:

$$\begin{array}{ll} H_0: & X \sim f_0 = \prod_{i=1}^N h_0(x_i) \\ H_1: & X \sim f_1 = h_0(x_1) \prod_{i=2}^N h_1(x_i) \\ \vdots & \vdots \\ H_M: & X \sim f_M = \left[\prod_{i=1}^{N-1} h_0(x_i)\right] h_1(x_N) \end{array} . \tag{51}$$

The weighted likelihood ratios for these $M+1$ hypotheses are

$$c_\theta \frac{f_\theta(\underline{X})}{f_0(\underline{X})} = c_\theta \prod_{i=\theta+1}^N \frac{h_1(X_i)}{h_0(X_i)}, \qquad \theta = 1, \ldots, M.. \tag{52}$$

The DO rule obtained from (40) in Remark 3 has the form:

$$\sum_{\theta=1}^N c_\theta^D \left[\prod_{i=\theta+1}^N \frac{h_1(X_i)}{h_0(X_i)}\right] \underset{H_0}{\overset{H_1}{\gtrless}} \lambda^D \tag{53}$$

where, similarly to the point change example, the threshold λ^D and the weight vector $\underline{c}^D$ are chosen such that $P(FA) = \alpha$ and the weighted average of miss probabilities is maximized. The CO rule obtained from (43) in Remark 4 has the form:

$$\max_{\theta > 0} \left\{ c_\theta^C \left[\prod_{i=\theta+1}^N \frac{h_1(X_i)}{h_0(X_i)}\right]\right\} \underset{H_0}{\overset{H_{\theta_{max}}}{\gtrless}} \lambda^C \tag{54}$$

where the threshold λ^C and the weight vector $\underline{c}^C$ are chosen such that $P(FA) = \alpha$ and the weighted average of erroneous classification probabilities is maximized. The equalization condition of Corollary 2 is difficult to apply for this example since the distribution of the likelihood ratio is not permutation invariant [2]. Thus the optimal weights $\underline{c}^D$ and $\underline{c}^C$ must be derived directly as the solution of the non-linear optimization problem (25) which maximizes average probability of miss and erroneous classification, respectively. In general, the uniform weight assignment is not min-max for the step change problem and therefore, unlike for the point change problem, the GLRT-MLC is not an optimal joint detection estimation rule.

An equivalent form for the CO decision rule (54) is to compare the running cusum statistic [15, 21]:

$$T_\theta(X) \stackrel{def}{=} \sum_{i=\theta+1}^{N} \ln \frac{h_1(X_i)}{h_0(X_i)},$$

to a curved boundary $\lambda^C - c_\theta$, $\theta = 1, \ldots, N$. If, as a function of θ, $T_\theta(x)$ crosses the curved boundary at time $\hat{\theta}$ then a step change is declared. Otherwise, if no boundary crossing occurs over $\theta = 1, \ldots, N$, H_0 is decided.

5.2 Detection, Order Selection and Classification for a Multi-component Signal in Noise

Here we consider the multi-component signal in noise model (1) introduced in Section 1. We have a set of N finite-energy orthonormal complex-valued signal components $\{S_1, \ldots, S_N\}$ each evolving over the time interval $[0, T]$, $S_i \stackrel{def}{=} \{S_i(t) : t \in [0, T]\}$. Available for measurement are L independent realizations (snapshots) $Y_k = \{Y_k(t) : t \in [0, T]\}$, $k = 1, \ldots, L$. Each waveform Y_k is composed of either a sum of randomly scaled versions of p of the N signal components plus noise, or noise alone. We assume that the noise $W_k(t)$ is a wideband complex Gaussian process and that the a_{lk}'s are i.i.d. zero-mean complex Gaussian random variables independent of the noise. We will also assume that an upper bound P on the number p of signal components is specified where $P < N$. The integer p, taking any value in $\{0, \ldots, P\}$, is unknown and, if $p > 0$, the signal indices $i_1, \ldots, i_p$ taking any p distinct values in $\{1, \ldots, N\}$ are also unknown. The noise variance $\sigma_W^2 = 1$ and the common variance σ_a^2 of the scaling factors $\{a_{lk}\}$ are known.

For each k let the continuous-time signal Y_k be passed through an energy detector consisting of a bank of N matched filters with impulse responses $S_1(-t), \ldots, S_N(-t)$ followed by a modulus squared operation as shown in Figure 1. The output of this energy detector is the statistic $X_k = [X_{k1}, \ldots, X_{kN}]$, where $X_{ki} = |< S_i, Y_k >|^2$, $k = 1, \ldots, L$, $i = 1, \ldots, N$. The statistical decisions will be based on the dimension $L \times N$ matrix statistic:

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_L \end{bmatrix}$$

It can be easily shown [2] that the X_{ki}'s are independent exponentially distributed random variables with parameters β_i, the value of β_i depending on the presence or absence of signal energy at the output of the i-th matched filter:

$$f_{X_{ki}}(x) = \begin{cases} h_1(x) = \beta_1 \exp(-\beta_1 x) u(x), & i = i_1, \ldots, i_p \\ h_0(x) = \beta_0 \exp(-\beta_0 x) u(x), & o.w. \end{cases} \tag{55}$$

where $u(x)$ is the unit step function, $\beta_1 = \frac{1}{\gamma}$, $\beta_0 = 1$, and $\gamma = 1 + \sigma_a^2$.

The decision problem consists of three related objectives based on observation of X: i) signal detection, which is to decide if $p > 0$; ii) signal power estimation (order selection) which, if $p > 0$, is to specify p; and iii) signal component estimation (classification) which, if $p = p_0$, is to identify the p_0 indices $i_1, \ldots, i_{p_0}$.

For fixed $p \in \{1, \ldots, P\}$, and for a given set of p indices $i_1, \ldots, i_p$, the signal components $\{a_l S_{i_l}\}$ have equal power σ_a^2. Therefore permutations of these components are indistinguishable. There are thus $\binom{N}{p}$ possible distinct arrangements of p signal components over N possible indices $1, \ldots, N$. Hence there are a total of $M \stackrel{def}{=} \sum_{p=1}^{P} \binom{N}{p}$ possible signal arrangements. We define the mapping between component indices $i_1, \ldots, i_p$, $p = 0, \ldots, P$, and a scalar θ as follows. For $j = 1, \ldots, N$, let θ_j be equal to a binary variable taking the value one if S_j is present and zero otherwise. Define the N-element vector $\underline{\theta} = [\theta_1, \ldots, \theta_N]^T$ as the the binary representation of the integer $\sum_{j=1}^{N} \theta_j 2^{j-1}$. Define the row vector $\underline{\psi}_k$ as the sequence of integers $\{\sum_{j=1}^{N} \theta_j 2^{j-1} : \sum_{j=1}^{N} \theta_j = k\}$ arranged in increasing order. Thus $\underline{\psi}_0 = 0$, $\underline{\psi}_1 = [1, 2, 4, 8, \ldots, 2^{(N-1)}]^T$, $\underline{\psi}_2 = [3, 5, 6, 9, \ldots, 2^{(N-2)} + 2^{(N-1)}]^T, \ldots, \underline{\psi}_P = [\sum_{j=1}^{P} 2^{j-1}, \ldots, \sum_{j=P}^{N} 2^{j-1}]^T$, corresponding to $p = 0$, 1, 2 and P signal components, respectively. Finally the mapping from the binary vector $\underline{\theta}$ to the scalar parameter θ is obtained by identifying θ as the successive indices $0, \ldots, M$ of the M-element vector $[\underline{\psi}_0, \underline{\psi}_1, \ldots, \underline{\psi}_P]$. The hypotheses are defined as before: $\{H_\theta\}_{\theta=0}^{M}$. Thus H_0 is the noise alone hypothesis ($\theta = p = 0$), each of $H_1, \ldots, H_N$ is a hypothesis that one of the N distinct arrangements of $p = 1$ signal components is present, each of $H_{\binom{N}{1}+1}, \ldots, H_{\binom{N}{1}+\binom{N}{2}}$ is a hypothesis that one of the $\binom{N}{2}$ distinct arrangements of $p = 2$ signal components are present, and so on.

Using the result (55) we have:

$$\begin{array}{rl} H_0: & X \sim \prod_{k=1}^{L} \prod_{i=1}^{N} h_0(x_{ik}) \\ H_1: & X \sim \prod_{k=1}^{L} h_1(x_{1k}) \prod_{i=2}^{N} h_0(x_{ik}) \\ \vdots & \quad \vdots \\ H_N: & X \sim \prod_{k=1}^{L} h_1(x_{Nk}) \prod_{i=1}^{N-1} h_0(x_{ik}) \\ H_{N+1}: & X \sim \prod_{k=1}^{L} \prod_{i=1}^{2} h_1(x_{ik}) \prod_{i=3}^{N} h_0(x_{ik}) \\ \vdots & \quad \vdots \\ H_M: & X \sim \prod_{k=1}^{L} \prod_{i=N-P+1}^{N} h_1(x_{ik}) \prod_{i=1}^{N-P} h_0(x_{ik}) \end{array} . \tag{56}$$

The likelihood ratios can be expressed in the form [2]:

$$\frac{f_\theta(X)}{f_0(X)} = \frac{1}{\gamma^{L\hat{p}(\underline{\theta})}} \exp\left[\frac{\gamma - 1}{\gamma} \underline{1}_L^T X \underline{\theta}\right], \qquad \theta = 1, \ldots, M \tag{57}$$

where $\underline{1}_L = [1, \ldots, 1]^T$ is an L-vector of 1's and $\hat{p}(\underline{\theta}) = \sum_{i=1}^{N} \theta_i$ is the number p of signal components under hypothesis H_θ.

The detection optimal (DO) rule obtained from (40) in Remark 3 has the form:

$$\sum_{\theta=1}^{M} c_\theta^D \frac{f_\theta(X)}{f_0(X)} \underset{H_0}{\overset{H_{\{p>0\}}}{\gtrless}} \lambda^D, \tag{58}$$

the order selection optimal (OSO) rule obtained from Theorem 2 has the form:

$$\max_{p>0} \left\{ \sum_{\theta:\hat{p}(\theta)=p} c_\theta^O \frac{f_\theta(X)}{f_0(X)} \right\} \underset{H_0}{\overset{H_{p_{max}}}{\gtrless}} \lambda^O, \tag{59}$$

and the classification optimal (CO) rule obtained from (43) in Remark 4 has the form:

$$\max_{\theta>0} \left\{ c_\theta^C \frac{f_\theta(X)}{f_0(X)} \right\} \underset{H_0}{\overset{H_{\theta_{max}}}{\gtrless}} \lambda^C. \tag{60}$$

The weights for the OSO and CO rules can be found by choosing $\underline{c}^O$ and $\underline{c}^C$ to equalize the probabilities of incorrect order selection and incorrect classification, respectively associated with the likelihood ratio tests (59) and (60), as indicated by Corollary 2. On the other hand, the miss probabilities of the DO rule (58) cannot be equalized for any choice of the weights $\{c_\theta\}$ [2] and the optimal weight vector must be found in the manner outlined in Remark 5. In particular, itcan be shown that the optimal weight vector $\{c_\theta\}$ equalizes the miss probabilities for $1 \leq \theta \leq N$ (single signal component) and assigns zero weights for $N < \theta \leq M$ (multiple signal components).

Denote the optimal weights generically as c_θ^* ($*$ = "D", or "O", or "C"). In [2], it is shown that due to the equal power signal component assumption, the distribution of the likelihood ratios (56) is invariant to permutations of the indices $i_1, \ldots, i_p$ for any fixed p in the sense that for any permutation matrix T operating on $\underline{\theta}$, the ordered set $\left\{\frac{f_\theta}{f_0}\right\}_{\underline{\theta}}$ has the same joint distribution under $H_{\underline{\theta}}$ as the joint distribution of $\left\{\frac{f_{T\theta}}{f_0}\right\}_{\underline{\theta}}$ under $H_{T\underline{\theta}}$. This implies that the optimal weights $\{c_\theta^*\}$ for (58)-(60) are constant for fixed p and satisfy [2]:

$$c_\theta^* = \begin{cases} \frac{q_1^*}{\binom{N}{1}}, & 1 \leq \hat{p}(\underline{\theta}) \leq N \\ \vdots & \vdots \\ \frac{q_P^*}{\binom{N}{P}}, & \sum_{j=1}^{P-1} \binom{N}{j} + 1 \leq \hat{p}(\underline{\theta}) \leq M \end{cases}, \tag{61}$$

where $q_p^* \in [0, 1]$, $p = 1, \ldots, P$, and $\sum_{p=1}^{P} q_p^* = 1$. The relation (61) corresponds to a significant reduction (from $M = \sum_{p=1}^{P} \binom{N}{p}$ to P) in the number of weights to be determined and greatly simplifies the search for the min-max constrained decision rules and associated lower bounds on detection, order selection, and classification error probability.

An Optimal Order Selection Rule for Classification

Using the q_k^* weight specification (61), the classification optimal rule (60) has the following equivalent form:

$$\max_{\theta > 0} \left\{ \ln \frac{f_\theta(X)}{f_0(X)} + \ln \frac{q^C_{\hat{p}(\underline{\theta})}}{\binom{N}{\hat{p}(\underline{\theta})}} \right\} \quad \overset{H_{\theta_{max}}}{\underset{H_0}{\gtrless}} \quad \ln \lambda^C, \tag{62}$$

where as defined above $\hat{p}(\underline{\theta}) = \sum_{i=1}^{N} \theta_i$ is the number of signal components subsumed by hypothesis H_θ. The form (62) establishes that the CO rule incorporates an order selection penalty, $g(p) = \ln \frac{q_p^C}{\binom{N}{p}} \leq 0$, which is added to the log-likelihood function $\ln \frac{f_\theta}{f_0}$ for classification. It is the equal power signal components assumption that makes the optimal penalty depend solely on the hypothesized number of signal components. In the general case, where the signal components can have unequal power, the min-max optimum penalty $g(p, \theta) = \ln c_\theta^C$ depends on the specific signal component indices assumed under H_θ. For comparison consider the unweighted GLRT-MLC for which $g(p)$ is a constant independent of p, Akaike's AIC for which $g(p) = -p$, and Rissanen's MDL for which $g(p) = -\frac{p}{2} \ln L$ where L is the number of snapshots. Unlike the CO penalty function, the GLRT-MLC, AIC and MDL penalty functions are not min-max optimal for constrained classification. On the other hand, while the CO penalty function typically depends on the parameters of the likelihood ratio distribution through the solution to the difficult minimization problem (25), these other penalty functions can be specified independently of any such parameters. Under certain asymptotic conditions, however, the CO penalty function also becomes independent of the process parameters. The following proposition establishes this fact.

Proposition 1 *Let the optimal weights $\{q_p^C\}_{p=1}^P$ satisfy for all fixed $p > 0$*

$$\frac{-\ln q_p^C}{\ln \binom{N}{p}} \to \beta \text{ as } N \to \infty, \tag{63}$$

where β is a finite constant. Then

$$\ln \left[\frac{q_p^C}{\binom{N}{p}}\right] \to -(1+\beta)p \ln N \leq 0. \tag{64}$$

Proof: Expressing

$$\ln \left[\frac{q_p^C}{\binom{N}{p}}\right] = -\ln\left[\binom{N}{p}\right]\left[1 - \frac{\ln q_p^C}{\ln \binom{N}{p}}\right] \tag{65}$$

using the large N Stirling approximation $\ln \binom{N}{p} \approx p \ln N$ and the boundedness of β, it is obvious that as $N \to \infty$ the right hand side of (65) becomes: $-(1+\beta)\ln p \ln N$. Since $c_\theta = q_p / \binom{N}{p} \leq 1$ for all $N > 0$ this limit is less than or equal to zero.

□

Proposition 1 shows that even if the weights $\{q_p^C\}$ decay to zero as $N \to \infty$, as long as the rate of decay is not too high, the CO penalties converge to $g(p) = -\frac{p}{2}\ln N$.

Numerical Comparisons

Here we quantify the performance of the DO, OSO and CO rules ((58) - (60)) to study the tradeoffs between detection, order selection, and estimation.

Since classification is performed over the finest partition of the parameter space Θ, the CO rule (60) specifies order selection and signal detection rules which are optimal for min-max classification. For example, when CO makes a decision that the signal components are at bins $i_1, \ldots, i_p$ it also specifies "signals detected" ($p > 0$) and "exactly p signal components present.". From (60) we see that the CO rule performs "classification optimal" order selection as:

$$\max_{p>0}\left\{\max_{\theta:\hat{p}(\underline{\theta})=p}\left\{c_\theta^C \frac{f_\theta(X)}{f_0(X)}\right\}\right\} \quad \underset{H_0}{\overset{H_{p_{max}}}{\gtrless}} \quad \lambda^C, \tag{66}$$

and the CO rule performs "classification optimal" signal detection as:

$$\max_{\theta>0}\left\{c_\theta^C \frac{f_\theta(X)}{f_0(X)}\right\} \quad \underset{H_0}{\overset{H_{\{p>0\}}}{\gtrless}} \quad \lambda^C. \tag{67}$$

On the other hand, in addition to providing an optimal order selection rule, the OSO rule (59) also specifies an "order selection optimal" detection rule:

$$\max_{p>0}\left\{q_p^O \frac{f_\theta^{(\underline{c}^O)}(X)}{f_0(X)}\right\} \quad \underset{H_0}{\overset{H_{\{p>0\}}}{\gtrless}} \quad \lambda^O. \tag{68}$$

Note that the OSO rule does not itself specify an post-order-selection classifier while the DO rule does not specify a post-detection order-selector or classifier.

To assess the impact of imposing classification optimality on best achievable detection performance and best achievable order selection performance we evaluate the difference between CO detector and DO detector error probabilities, and the difference between CO order selector and OSO order selector error probabilities. Since the DO detector and the OSO order selector are optimal for detection and order selection, respectively, these differences are always non-negative, the differences corresponding to performance losses associated with requiring classification optimality.

We also investigated the classification optimality of the strategies of gating an (unconstrained) conditionally min-max classifier with a OSO order selector and DO detector called the OSO-gated classification rules and the DO-gated classification rules, respectively. While in the case of DO gating there is a single conditionally min-max classifier, which classifies all signal indices given that the DO rule declares signals to be present, in the case of OSO gating there are P conditionally min-max classifiers, each one classifiying the indices of the number p of signals decided upon by the OSO rule. Applying Theorem 2 it is seen that the conditionally minmax classifiers are of the form of weighted maximum likelihood estimators (38). Since all signal components have equal power, it can be shown [2] that for each p the OSO-gated conditionally min-max classifier is an unweighted maximum likelihood estimator of θ over the $\binom{N}{p}$ configurations of p signals over N bins, while the DO-gated conditionally min-max classifier is a weighted maximum likelihood estimator of θ over all $\sum_{p=1}^{P}\binom{N}{p}$ possible configurations of $1,\ldots,P$ signals over N bins.

Using numerical integration and simulations we evaluated the performance of the classification optimal rule, order selection optimal rule, and detection optimal rule for the case where $P = 2$, i.e. where it is *a priori* known that there can be at most two signal components. The false alarm and erroneous classification probabilities of the CO rules were computed analytically [2] while the remaining decision error probabilities were determined via simulations. For each simulation a complex Gaussian 4×10 observation matrix X was generated corresponding $L = 4$ independent realizations and $N = 10$ possible orthogonal signal component indices. The signal to noise power ratio per observation per signal was set to $\gamma - 1 = 2$ (+3 dB). The false alarm probability was constrained to $\alpha = 0.1$. Successive columns of Table 1 show the worst case erroneous classification (EC) probability $\max_{\theta \notin \Theta_0} P_\theta(EC)$, worst case erroneous order selection (EOS) probability $\max_{\theta \notin \Theta_0} P_\theta(EOS)$, and worst case miss (M) probability $\max_{\theta \notin \Theta_0} P_\theta(M)$ for the CO, OSO, and DO rules. Since CO minimizes $\max_{\theta \notin \Theta_0} P_\theta(EC)$, OSO minimizes $\max_{\theta \notin \Theta_0} P_\theta(EOS)$, and DO minimizes $\max_{\theta \notin \Theta_0} P_\theta(M)$, the diagonal entries, 0.59, 0.52, and 0.41, of Table 1 provide us with respective lower bounds on the worst case erroneous classification probability, erroneous order selection probability, and miss probability which apply to any identically constrained ($P(FA) = \alpha$)) decision rule.

The important observation from Tables 1 and 2 is that requiring classification optimality necessarily entails a loss in order selection performance: the worst case erroneous order selection probability of the classification optimal rule is 0.56; in relative terms this is (0.56 - 0.52)/0.52 = 7.69% above the order selection lower bound. This result indicates that order selection optimality and classification optimality are not generally simultaneously achievable. Conversely, requiring order selection optimality under the uncoupled design strategy entails a loss in classification performance: the worst case erroneous classification probability of the order selection optimal rule is 0.62; in relative terms this is (0.62 - 0.59)/0.59 = 5.08% above the classification lower bound. Furthermore, requiring detection optimality under the uncoupled design strategy entails a significant loss in classification performance and order selection performance: the worst case erroneous classification probability of the detection optimal rule ((3, 1) entry, 0.70) is (0.70 - 0.59)/0.59 = 18.64% above the classification lower bound, and its worst case erroneous order selection probability ((3, 2) entry, 0.68) is (0.68 - 0.52)/0.52 = 30.77% above the order selection lower bound. These results are an indication that the commonly used uncoupled design approach can severely sacrifice order selection and classification performance. On the other hand, requiring classification optimality or order selection optimality entails only very little performance loss in detection performance: the worst case miss probabilities of the classification optimal and the order selection optimal rules both are (0.42 - 0.41)/0.41 = 2.44% above the detection lower bound. These relative performance losses are summarized in Table 2.

We also evaluated the loss in performance due to using a suboptimal order selection penalty function $g(p)$. Figure 2 is a plot, as a function of the number of snapshots L, of the associated worst case erroneous classification probabilities for: the optimal penalty $g(p) = \ln \frac{q_p^C}{\binom{N}{\hat{p}(\theta)}}$, the uniform (ML) penalty $g(p) = 0$, the Akaike AIC penalty $g(p) = -p$, and the Rissanen MDL penalty $g(p) = -\frac{p}{2} \ln L$. The worst case classification performance is lower bounded by the classification performance using our optimal penalty (solid line). Observe that the MDL penalty is near optimal for large L while the AIC is near optimal for small L with both curves lying at most 0.1 above the lower bound. On the other hand, the uniform penalty function, corresponding to the unweighted GLRT-MLC rule, entails a significant loss in performance over most of the range of L studied; e.g. as high as 0.25 above the lower bound.

Appendix: Proofs

Proof of Theorem 1:

By the assumption (18), $\underline{\phi}^* = \underline{\phi}^{(\underline{b}^*)}$ is obviously of level α for testing among the composite hypotheses $\underline{H}_0, \ldots, \underline{H}_J$. To prove that $\underline{\phi}^* = \underline{\phi}^{(b^*)}$ is a constrained min-max test of level α, we need to verify that for any other test function $\underline{\phi}$ of level α with respect to the composite null hypothesis $\underline{H}_0$,

$$\max_{\theta \notin \Theta_0} E_\theta[1 - \phi^*_{r_J(\theta)}] \le \max_{\theta \notin \Theta_0} E_\theta[1 - \phi_{r_J(\theta)}]. \tag{69}$$

Recall that by definition $\underline{\phi}^* = \underline{\phi}^{(\underline{b}^*)}$ minimizes the worst case classification error probability $\max_{\theta \notin \Theta_0} E_\theta[1 - \phi_{r_J(\theta)}]$ among all test functions of level α for the hypotheses $H_0^{(\underline{b}^*)}, \underline{H}_1, \ldots, \underline{H}_J$. It will therefore suffice to show that any test function $\underline{\phi}$ of level α for the hypotheses $\underline{H}_0, \underline{H}_1, \ldots, \underline{H}_J$ is also of level α for the hypotheses $H_0^{(\underline{b}^*)}, \underline{H}_1, \ldots, \underline{H}_J$. Since $E_\theta[1 - \phi_0] \le \alpha$ for all $\theta \in \Theta_0$, we have

$$\begin{aligned} E_0^{(\underline{b}^*)}[1 - \phi_0] &= \int_{\mathcal{X}} (1 - \phi_0(x)) f_0^{(\underline{b}^*)}(x) d\mu(x) \\ &= \int_{\mathcal{X}} (1 - \phi_0(x)) \sum_{\theta \in \Theta_0} b^*_\theta f_\theta(x) d\mu(x) \\ &= \sum_{\theta \in \Theta_0} b^*_\theta \int_{\mathcal{X}} (1 - \phi_0(x)) f_\theta(x) d\mu(x) \\ &= \sum_{\theta \in \Theta_0} b^*_\theta E_\theta[1 - \phi_0] \\ &\le \alpha. \end{aligned} \tag{70}$$

This shows that $\underline{\phi}$ is of level α also with respect to the simple null hypothesis $H_0^{(\underline{b}^*)}$ and therefore cannot have smaller worst case classification error than $\underline{\phi}^{(\underline{b}^*)}$.

As for the sufficient condition for existence of $\underline{b}^*$, we use the following:

$$\begin{aligned}
\max_{\theta\in\Theta_0} E_\theta[1-\phi_0^{(\underline{b}^*)}] &= \max_{\underline{b}\in\mathcal{C}_{K-M+1}} \sum_{\theta\in\Theta_0} b_\theta E_\theta[1-\phi_0^{(\underline{b}^*)}] \\
&= \max_{\underline{b}\in\mathcal{C}_{K-M+1}} \int_{\mathcal{X}} [1-\phi_0^{(\underline{b}^*)}(x)] \sum_{\theta\in\Theta_0} b_\theta f_\theta(x) d\mu(x) \\
&= \max_{\underline{b}\in\mathcal{C}_{K-M+1}} \int_{\mathcal{X}} [1-\phi_0^{(\underline{b}^*)}(x)] f_0^{(\underline{b})}(x) d\mu(x) \\
&\leq \int_{\mathcal{X}} [1-\phi_0^{(\underline{b}^*)}(x)] f_0^{(\underline{b}^*)}(x) d\mu(x) \\
&= \alpha,
\end{aligned} \tag{71}$$

where the last two lines follow directly from the assumed condition (20) and (19).

□

Proof of Theorem 2:

The proof mainly consists of establishing three facts: Existence of the test function $\underline{\phi}^* \in \mathcal{D}_\alpha$ defined by the relations (22)-(25), existence of a min-max strategy, the min-max property of $\underline{\phi}^*$. We start by showing the existence of $\underline{\phi}^*$.

<u>Existence of $\underline{\phi}^* \in \mathcal{D}_\alpha$</u>

By using the defining equations (22) and (23), we can readily verify that $\underline{\phi}^{(\underline{c})}$ satisfies $\underline{\phi}^{(\underline{c})} \in [0,1]^{(J+1)}$, $\forall \underline{c} \in \mathcal{C}_M$ and $\sum_{j=0}^{J} \phi_j^{(\underline{c})} = 1$. We next show that i) for any fixed $\underline{c}$, there exist $\lambda > 0$ and $\xi \in [0,1]$, both functions of $\underline{c}$, such that the false alarm constraint given by (24) is satisfied and, ii) a minimizing weight vector $\underline{c}^*$ exists that maximizes (25) over $\mathcal{C}_M$. But given assertion i), assertion ii) (and thus the existence of $\underline{\phi}^*$) follows directly due to the fact that the convex set $\mathcal{C}_M$ is compact. Hence it remains to justify assertion i).

For $j = 1, \ldots, J$, define $L_j \stackrel{def}{=} (q_j f_j^{(\underline{c})})/f_0$. If $f_0 \equiv 0$, define L_j to be unbounded. Let $F(z) \stackrel{def}{=} P_0(\{X : L_{j_{max}}(X) < z\})$, $\forall z \in \mathcal{R}$, where $j_{max} \stackrel{def}{=} \arg\max_{\theta\notin\Theta_0} q_j f_j^{(\underline{c})}$. Then using the definition of $\underline{\phi}$, we can express the false alarm constraint in (24) as:

$$F(\lambda) + \xi P_0(\{X : L_{j_{max}}(X) = \lambda\}) = 1 - \alpha. \tag{72}$$

It follows from the left continuity and monotonicity of the function $F(z)$ that there exists a point $z = \lambda$ such that $F(\lambda) \leq 1-\alpha$ and $F(\lambda^+) \geq 1-\alpha$, where $F(\lambda^+)$ denotes the right limit of F at λ, i.e $F(\lambda^+) = \lim_{z\downarrow\lambda} F(z)$. Now if $F(z)$ is continuous at λ, then $P_0(\{X : L_{j_{max}}(X) = \lambda\}) = 0$ and $F(\lambda) = 1-\alpha$ so the constraint (24) is satisfied for any ξ. If, on the other hand, $F(z)$ is not continuous at λ, then $P_0(\{X : L_{j_{max}}(X) = \lambda\}) = F(\lambda^+) - F(\lambda) > 0$ and we can satisfy (24) by choosing $\xi = \frac{1-\alpha-F(\lambda)}{F(\lambda^+)-F(\lambda)}$. This establishes the existence of a decision rule $\underline{\phi}^* \in \mathcal{D}_\alpha$ that satisfies (22)-(24).

We now proceed to the proof of the existence of a min-max strategy.

<u>Existence of a min-max strategy</u>

We must show that there exists a decision rule $\underline{\phi}^* \in \mathcal{D}_\alpha$ that achieves the infimum value

$$\max_{\theta \notin \Theta_0} E_\theta[1 - \phi^*_{r_J(\theta)}] = \inf_{\underline{\phi} \in \mathcal{D}_\alpha} \max_{\theta \notin \Theta_0} E_\theta[1 - \phi_{r_J(\theta)}]. \tag{73}$$

To prove this, we will use the method of risk sets [6].

Define the risk function $R(\theta, \underline{\phi})$ as the probability of error incurred by using the test function $\underline{\phi}$:

$$R(\theta, \underline{\phi}) \stackrel{def}{=} 1 - E_\theta[\phi_{r_J(\theta)}]. \tag{74}$$

In terms of the risk function, the min-max problem in (73) can be written as

$$\max_{\theta \notin \Theta_0} R(\theta, \underline{\phi}^*) = \inf_{\underline{\phi} \in \mathcal{D}_\alpha} \max_{\theta \notin \Theta_0} R(\theta, \underline{\phi}). \tag{75}$$

Now observe that from the definition of $\mathcal{C}_M$ we have:

$$\inf_{\underline{\phi} \in \mathcal{D}_\alpha} \max_{\theta \notin \Theta_0} R(\theta, \underline{\phi}) = \inf_{\underline{\phi} \in \mathcal{D}_\alpha} \max_{\underline{c} \in \mathcal{C}_M} \sum_{\theta \notin \Theta_0} c_\theta R(\theta, \underline{\phi}). \tag{76}$$

To conclude that the min-max problem on the right hand side of (76) admits a solution, we will need to use the min-max theorem [6, Thm. 1 of Sec. 2.9], which dictates that a min-max strategy exists if the risk set relevant to the problem is convex and compact. The relevant risk set in this case is the constrained risk set $\mathcal{S}_\alpha$ defined below:

$$\mathcal{S}_\alpha \stackrel{def}{=} \{[y_0, \ldots, y_M]^T : y_\theta = R(\theta, \underline{\phi}),\, \theta \in \Theta,\, y_0 \leq \alpha\}. \tag{77}$$

To show that $\mathcal{S}_\alpha$ is both convex and compact we proceed as follows. Consider the unconstrained risk set $\mathcal{S}$ defined by $\mathcal{S} \stackrel{def}{=} \{[y_0, \ldots, y_M]^T : y_\theta = R(\theta, \underline{\phi}), \theta \in \Theta\}$. From [6, Lemma 1 of Sec. 1.7], the set $\mathcal{S}$ is convex. Furthermore, since $\mathcal{S}$ is the convex hull of the risk set of nonrandomized decision rules, which in this case is finite and thus compact, it follows from [6, Thm. 2 of Sec. 2.4] that $\mathcal{S}$ is also compact. Now note that the constrained risk set $\mathcal{S}_\alpha$ is the intersection of $\mathcal{S}$ and the closed half-plane $\{y_0 = R(0, \underline{\phi}) \leq \alpha\}$. Hence the convexity and compactness of the constrained risk set $\mathcal{S}_\alpha$.

Therefore, by the min-max theorem cited above, the function $\sum_{\theta \notin \Theta_0} c_\theta R(\theta, \underline{\phi})$ possesses a saddle point over $\underline{c} \in \mathcal{C}_M$ and $\underline{\phi} \in \mathcal{D}_\alpha$:

$$\inf_{\underline{\phi} \in \mathcal{D}_\alpha} \max_{\underline{c} \in \mathcal{C}_M} \sum_{\theta \notin \Theta_0} c_\theta R(\theta, \underline{\phi}) = \max_{\underline{c} \in \mathcal{C}_M} \inf_{\underline{\phi} \in \mathcal{D}_\alpha} \sum_{\theta \notin \Theta_0} c_\theta R(\theta, \underline{\phi}), \tag{78}$$

and there exists an admissible min-max decision rule with an associated "least favorable distribution" over the alternatives. Furthermore, due to the existence of a test function achieving the infimum over $\mathcal{D}_\alpha$, we can change infimum to minimum over $\mathcal{D}_\alpha$.

<u>The min-max property of $\underline{\phi}^*$</u>

We must now show that, for fixed $\underline{c}$, the decision rule $\underline{\phi}^{(\underline{c})}$ defined by the relations (22)-(24) achieves the following minimum:

$$\sum_{\theta \notin \Theta_0} c_\theta R(\theta, \underline{\phi}^{(\underline{c})}) = \min_{\underline{\phi} \in \mathcal{D}_\alpha} \sum_{\theta \notin \Theta_0} c_\theta R(\theta, \underline{\phi}) \tag{79}$$

since, by (78), (76), (75), and the defining relation (25) of the weight vector $\underline{c}^*$, this will be equivalent to $\underline{\phi}^* = \underline{\phi}^{(\underline{c}^*)}$ being a constrained min-max test.

For an arbitrary test function $\underline{\phi} \in \mathcal{D}_\alpha$, let $\rho(\underline{\phi}) \stackrel{def}{=} \sum_{\theta \notin \Theta_0} c_\theta R(\theta, \underline{\phi})$. We will show that $\rho(\underline{\phi}) - \rho(\underline{\phi}^{(\underline{c})}) \geq 0$. Using the definition of risk (74) and the definition of the hypothetical pdf's $\{f_j^{(\underline{c})}\}$ (15), the identities $\phi_0 = 1 - \sum_{j=1}^J \phi_j$ and $\sum_{\theta \notin \Theta_0} c_\theta = 1$, the expression for the expectation $E_\theta[\phi_l] = \int_{\mathcal{X}} \phi_l(x) f_\theta(x) d\mu(x)$, and the false alarm constraint $E_0[\phi_0] \geq 1 - \alpha$, we can write the following lower bound on the difference $\rho(\underline{\phi}) - \rho(\underline{\phi}^{(\underline{c})})$:

$$\begin{aligned}\rho(\underline{\phi}) - \rho(\underline{\phi}^{(\underline{c})}) &\geq \sum_{k,j=1}^J \int_{\mathcal{X}} [q_k f_k^{(\underline{c})}(x) - q_j f_j^{(\underline{c})}(x)] \phi_k^{(\underline{c})}(x) \phi_j(x) d\mu(x) \\ &+ \sum_{j=1}^J \int_{\mathcal{X}} [q_j f_j^{(\underline{c})}(x) - \lambda f_0(x)] (\phi_j^{(\underline{c})}(x) \phi_0(x) - \phi_j(x) \phi_0^{(\underline{c})}(x)) d\mu(x) \\ & .\end{aligned} \tag{80}$$

To see that the right hand side of (80) is non-negative, we will consider the following partition of $\mathcal{X}$: $\mathcal{X}_0 = \{x \in \mathcal{X} : \max_{j>0} L_j(x) < \lambda\}$ and $\mathcal{X}_1 = \{x \in \mathcal{X} : \max_{j>0} L_j(x) \geq \lambda\}$. Now we write each integral over $\mathcal{X}$ in the right hand side of (80) as the sum of two integrals, over each of the $\mathcal{X}_i$'s, $i = 0, 1$:

$$\begin{aligned}\sum_{k,j=1}^J \int_{\mathcal{X}} [q_k f_k^{(\underline{c})}(x) - q_j f_j^{(\underline{c})}(x)] \phi_k^{(\underline{c})}(x) \phi_j(x) d\mu(x) &= \\ \sum_{k,j=1}^J \int_{\mathcal{X}_0} [q_k f_k^{(\underline{c})}(x) - q_j f_j^{(\underline{c})}(x)] \phi_k^{(\underline{c})}(x) \phi_j(x) d\mu(x) & \\ + \sum_{k,j=1}^J \int_{\mathcal{X}_1} [q_k f_k^{(\underline{c})}(x) - q_j f_j^{(\underline{c})}(x)] \phi_k^{(\underline{c})}(x) \phi_j(x) d\mu(x), &\end{aligned} \tag{81}$$

$$\begin{aligned}\sum_{j=1}^J \int_{\mathcal{X}} [q_j f_j^{(\underline{c})}(x) - \lambda f_0(x)] (\phi_j^{(\underline{c})}(x) \phi_0(x) - \phi_j(x) \phi_0^{(\underline{c})}(x)) d\mu(x) &= \\ \sum_{j=1}^J \int_{\mathcal{X}_0} [q_j f_j^{(\underline{c})}(x) - \lambda f_0(x)] (\phi_j^{(\underline{c})}(x) \phi_0(x) - \phi_j(x) \phi_0^{(\underline{c})}(x)) d\mu(x) & \\ + \sum_{j=1}^J \int_{\mathcal{X}_1} [q_j f_j^{(\underline{c})}(x) - \lambda f_0(x)] (\phi_j^{(\underline{c})}(x) \phi_0(x) - \phi_j(x) \phi_0^{(\underline{c})}(x)) d\mu(x). &\end{aligned} \tag{82}$$

Now, by the defining equations (22)-(23), $\phi_k^{(\underline{c})}(x)$ vanishes on $\mathcal{X}_0$ for $k > 0$. Therefore the first term in (81) is zero and, since $-\phi_k(x) \phi_0^{(\underline{c})}(x) \leq 0, \forall x \in \mathcal{X}$, the first term in (82) is non-negative.

For the integrals over $\mathcal{X}_1$, observe that $\forall x \in \mathcal{X}_1$, $\phi_k^{(\underline{c})}(x) = 0$ for $k \neq j_{max}, k = 0, \ldots, J$ and $q_{j_{max}} f_{j_{max}}^{(\underline{c})}(x) - q_j f_j^{(\underline{c})}(x) \geq 0$. Thus we have

$$\sum_{k,j=1}^J \int_{\mathcal{X}_1} [q_k f_k^{(\underline{c})}(x) - q_j f_j^{(\underline{c})}(x)] \phi_k^{(\underline{c})}(x) \phi_j(x) d\mu(x)$$

$$= \sum_{j=1}^{J} \int_{\mathcal{X}_1} [q_{j_{max}} f_{j_{max}}^{(\underline{c})}(x) - q_j f_j^{(\underline{c})}(x)] \phi_{j_{max}}^{(\underline{c})}(x) \phi_j(x) d\mu(x)$$
$$\geq 0, \tag{83}$$

and

$$\sum_{j=1}^{J} \int_{\mathcal{X}_1} [q_j f_j^{(\underline{c})}(x) - \lambda f_0(x)](\phi_j^{(\underline{c})}(x)\phi_0(x) - \phi_j(x)\phi_0^{(\underline{c})}(x)) d\mu(x)$$
$$= \int_{\mathcal{X}_1} [q_{j_{max}} f_{j_{max}}^{(\underline{c})}(x) - \lambda f_0(x)] \phi_{j_{max}}^{(\underline{c})}(x) \phi_0(x) d\mu(x)$$
$$\geq 0. \tag{84}$$

This concludes the proof of (79).

Now substituting (79) in (78) and using (76) in conjunction with (75), we obtain:

$$\min_{\underline{\phi} \in \mathcal{D}_\alpha} \max_{\theta \notin \Theta_0} R(\theta, \underline{\phi}) = \max_{\underline{c} \in \mathcal{C}_M} \sum_{\theta \notin \Theta_0} c_\theta R(\theta, \underline{\phi}^{\underline{c}^*}), \tag{85}$$

which, by the definition of the risk functions, given in (74), leads to

$$\max_{\underline{\phi} \in \mathcal{D}_\alpha} \min_{\theta \notin \Theta_0} E_\theta[\underline{\phi}_{r_J(\theta)}] = \min_{\underline{c} \in \mathcal{C}_M} \sum_{\theta \notin \Theta_0} c_\theta E_\theta[\phi_{r_J(\theta)}^{(\underline{c})}] \tag{86}$$

so that $\underline{\phi}^*$ is a constrained min-max test of level α.

□

Proof of Corollary 1:

By Theorem 2, $\underline{\phi}^* = \underline{\phi}^{(\underline{b}^*, \underline{c}^*)}$ is a constrained min-max test of level α for testing among the hypotheses $H_0^{(\underline{b}^*)}, \underline{H}_1, \ldots, \underline{H}_J$. Furthermore, if condition (30) holds, i.e. if $\underline{\phi}^*$ is of level α also with respect to the composite hypotheses $\underline{H}_0, \ldots, \underline{H}_J$, then the condition (18) of Theorem 1 is satisfied. In this case $\underline{\phi}^*$ is also a constrained min-max test of level α for testing among the composite hypotheses $\underline{H}_0, \ldots, \underline{H}_J$.

□

Proof of Corollary 2:

Condition (32) ensures that $\max_{\theta \in \Theta_0} E_\theta[1 - \phi_0^{(\underline{b}^*, \underline{c}^*)}] = \alpha$ so that $\underline{\phi}^*$ is of level α for testing between the composite hypotheses $\underline{H}_0, \ldots, \underline{H}_J$. Let $\underline{\phi}$ be an arbitrary level α test for testing among the composite hypotheses $\underline{H}_0, \ldots, \underline{H}_J$. To establish minmaxity, we need to show the following:

$$\max_{\theta \notin \Theta_0} E_\theta[1 - \phi_{r_J(\theta)}^*] \leq \max_{\theta \notin \Theta_0} E_\theta[1 - \phi_{r_J(\theta)}]. \tag{87}$$

We have directly from (33)

$$\sum_{\theta \notin \Theta_0} c_\theta^* E_\theta[1 - \phi_{r_J(\theta)}^*] = \max_{\theta \notin \Theta_0} E_\theta[1 - \phi_{r_J(\theta)}^*] = V. \tag{88}$$

On the other hand, it can be shown (see (79) in the proof of Theorem 2) that $\underline{\phi}^*$ satisfies

$$\sum_{\theta \notin \Theta_0} c_\theta^* E_\theta[1 - \phi_{r_J(\theta)}^*] \leq \sum_{\theta \notin \Theta_0} c_\theta^* E_\theta[1 - \phi_{r_J(\theta)}]. \tag{89}$$

Combining (88) and (89) we obtain

$$\begin{aligned}\max_{\theta\notin\Theta_0} E_\theta[1-\phi^*_{r_J(\theta)}] &\leq \sum_{\theta\notin\Theta_0} c^*_\theta E_\theta[1-\phi_{r_J(\theta)}] \\ &\leq \max_{\theta\notin\Theta_0} E_\theta[1-\phi_{r_J(\theta)}], \end{aligned} \tag{90}$$

which establishes (87).

□

Proof of Corollary 3:

Let $\underline{\phi}$ be an arbitrary level α test for testing among the composite hypotheses $\underline{H}_0,\ldots,\underline{H}_J$. Since $\underline{\phi}^*$ defined by (26)-(29) is a constrained min-max test of level α for the same hypotheses, by definition the following inequality is satisfied.

$$\max_{\theta\notin\Theta_0} E_\theta[1-\phi_{r_J(\theta)}] \geq \max_{\theta\notin\Theta_0} E_\theta[1-\phi^*_{r_J(\theta)}]. \tag{91}$$

□

References

[1] Baygün, B., Hero, A. O., "An order selection criterion via optimal joint estimation/detection theory," in *Proc. IEEE Workshop on Spectrum Est. and Modeling*, Rochester, New York, 1990.

[2] Baygün, B., "Optimal strategies and tradeoffs for joint detection and estimation," Ph.D. Thesis, Dept. of EECS, The University of Michigan, Ann Arbor, 48109-2122, 1992.

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

[4] Breiman, L, Friedman, J. H., Olshen, R. A., and Stone, C. J., *Classification and Regression Trees*, Wadsworth, California, 1984.

[5] Deshayes, J. and Picard, D., "Off-line statistical analysis of change-point models using non-parametric and likelihood methods," in *Detection of Abrupt Changes in Signals and Systems,* Benveniste, A. and Basseville, M. (eds.), pp. 103-168, Springer-Verlag, New York, 1977.

[6] Ferguson, T. S., *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, Florida, 1967.

[7] Fredriksen, A., Middleton, D., and Vandelinde, D., "Simultaneous signal detection and estimation under multiple hypotheses," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 5, pp. 760-768, November 1972.

[8] Hero, A. O. and Kim J. K., "Simultaneous signal detection and classification under a false alarm constraint," in Proc. IEEE ICASSP-90, Albuqerque, New Mexico, April 1990.

[9] Ibragimov, I. A. and Has'minskii, R. Z., *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, New York, 1981.

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

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

[12] Lehmann, E. L., *Testing Statistical Hypotheses*, John Wiley and Sons, New York, 1959.

[13] Merhav, N. and Ephraim, Y., "A Bayesian classification approach with application to speech recognition," *IEEE Trans. Signal Processing*, vol. 39, no. 10, pp. 2157-2166, October 1991.

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

[15] Page, E. S., "Continuous inspection schemes," Biometrika, vol. 42, pp. 100-115, 1954.

[16] Picard, D., "Testing and estimating change-points in time series," *Adv. Appl. Prob.*, vol. 17, pp. 841-867, 1985.

[17] Proakis, J. G., *Digital Communications*, McGraw-Hill, Tokyo, 1983.

[18] Rissanen, J., "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465-471, 1978.

[19] Scharf, L. L., *Statistical Signal Processing: Theory and Application*, Prentice-Hall, New Jersey, 1988.

[20] Schwartz, R. E., "Minimax CFAR detection in additive Gaussian noise of unknown covariance, *IEEE Trans. Inform. Theory*, vol. IT-15, no. 6, pp. 722-725, November 1969.

[21] Siegmund, D., *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag, New York, 1985.

[22] Stuller, J. A., "Generalized likelihood signal resolution," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 3, May 1975.

[23] Willsky, A. S., "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pp. 601-611, 1976.

[24] Zacks, S., *The Theory of Statistical Inference*, John Wiley and Sons, New York, 1971.

[25] Zeitouni, O., Ziv, J., and Merhav, N., "When is generalized likelihood ratio test optimal?," *IEEE Trans. Inform. Theory*, vol. IT-38, no. 5, pp. 1597-1602, September 1992.

[26] Zhang, Q. T., Wong, K. M., Patrick, C. Y., and Reilly, J. P., "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, no. 10, pp. 1557-1567, October 1989.

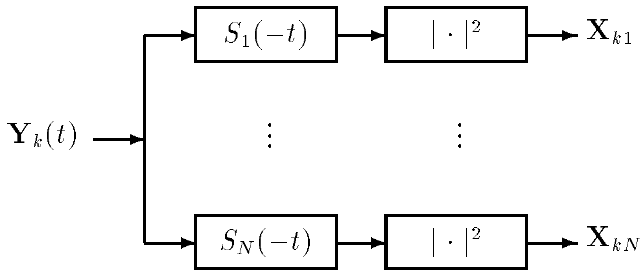

Figure 1: A bank of energy detectors for each of the signal components $\{S_1, \ldots, S_N\}$.

	$\max_{\theta \notin \Theta_0} P_\theta(EC)$	$\max_{\theta \notin \Theta_0} P_\theta(EOS)$	$\max_{\theta \notin \Theta_0} P_\theta(M)$
CO	0.59	0.56	0.42
OSO	0.62	0.52	0.42
DO	0.70	0.68	0.41

Table 1: Worst case performance comparisons for $L = 4, N = 10, \gamma = 3, \alpha = 0.1$.

	classification	order selection	detection
CO	0%	7.69%	2.44%
OSO	5.08%	0%	2.44%
DO	18.64%	30.77%	0%

Table 2: Relative performance losses for $L = 4, N = 10, \gamma = 3, \alpha = 0.1$.

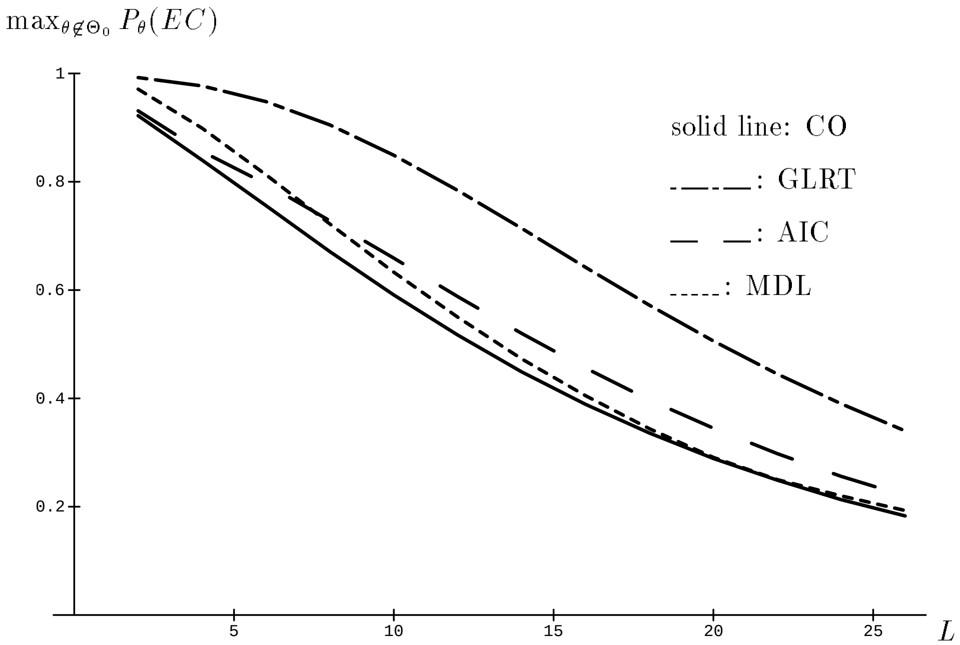

Figure 2: Worst case misclassification probabilities of the classification optimal rule, unweighted GLRT, AIC, and MDL as a function of number of snapshots L. $N = 10$, $\gamma = 2$, $\alpha = 0.1$.