SRC TR 87-156-r1

On Robust Continuous-Time Discrimination

By

E. Geraniotis
and
Y. A. Chau

ON ROBUST CONTINUOUS-TIME DISCRIMINATION

Evaggelos Geraniotis and Yawgeng A. Chau

Department of Electrical Engineering
and Systems Research Center
University of Maryland
College Park, MD 20742

ABSTRACT

Target discrimination problems are modeled as the testing of binary hypotheses characterized by continuous-time observations which (i) consist of distinct signals in additive white Gaussian noise, or (ii) are the output of stochastic dynamical systems driven by white Gaussian noise, and in both cases have partially known statistics. In particular, the signals in the first model, the parameters of the dynamical systems in the second model, and the autocorrelation functions of the noise in both models belong to one of the following distinct uncertainty classes: classes determined by 2-alternating capacities and classes with minimum or maximum elements. Robust discrimination tests with a fixed observation interval and sequential tests are derived whose likelihood ratios depend on the least-favorable pairs of parameters in the aforementioned uncertainty classes and are shown to have an acceptable level of performance despite the uncertainty. For tests with a fixed observation interval the performance measures considered are the actual error probabilities and the Chernoff upper bounds on them; the latter are shown to preserve their desirable asymptotic properties in the presence of the uncertainties. For sequential tests the performance measures are the error probabilities and the average required length of the observation interval under each hypothesis.

This research was supported in part by the Office of Naval Research under contract N00014-86-K-0013 and in part by the Systems Research Center at the University of Maryland, College Park, through the National Science Foundation's Engineering Research Centers Program: NSF CDR 8803012.

I. INTRODUCTION

In contrast to discrete-time robust detection problems (refer to the tutorial [1]), continuous-time robust detection problems–although involved in several situations of practical interest–have not received sufficient attention in the literature. The work of [2] on minimax continuous-time discrimination from Poisson observations constitutes an exception to this. Motivated by practical situations encountered in target discrimination, we investigate in this paper detection problems from continuous-time observations with statistical uncertainty and develop detectors (discriminators) with either **fixed-observation-interval** or **sequential** processing which are immune (robust) to the uncertainty in the observations.

Two types of problems of testing binary hypotheses are considered here: the first type pertains to observations that are signals in additive white Gaussian noise, namely

$$\mathcal{P}_1 \quad H_i \ : \ Y_t = s_i(t) + N_t; \quad i = 0, 1$$

where $0 \leq t \leq T$, $s_i(t)$ represents the signal under hypothesis H_i, and N_t is a Gaussian noise process with covariance $E[N_t N_s] = \sigma(t)\sigma(s)\delta(t-s)$.

The other type pertains to observations that constitute the path of the state in a dynamical system driven by white Gaussian noise described by the following stochastic equations under the two hypotheses

$$\mathcal{P}_2 \qquad H_i \ : \quad Y_t = X_t$$

$$dX_t = s_i(t) X_t dt + N_t; \quad i = 0, 1$$

for $0 \leq t \leq T$, where N_t is as described above, X_0 is a normal process $X_0 \sim \mathcal{N}(0,1)$, and $s_i(t)$ is the parameter of the dynamical system. Throughout this paper we place time variables in the subscript for the random processes and in the argument for the deterministic functions of time.

Both of the above formulations are motivated from problems of practical interest in radar target discrimination or target identification. In such problems, the output of the signal processor must indicate which of several targets is present or to discriminate between an actual target and a decoy or chaff cloud. Usually, this discrimination is to be performed after the initial decision that some object is present has taken place. As a special case of the two models above, the radar detection of the presence of a target could also be considered; in this case the signal under the null hypothesis is identically equal to zero. The continuous-time fixed-observation-interval or sequential processing considered here can model the processing of (i) each individual pulse of the radar return or (ii) of several consecutive pulses of the radar return.

If the characteristics of the channel and the system are known, then $s_i(t)$ and $\sigma(t)$ (defined above) are given. In reality, however, due to uncertainties in the the modeling of the targets and the medium through which the radar signal and the radar return signal propagate, the waveforms $s_i(t)$ and the second-order characteristics of the additive noise are not completely known. To model the available partial knowledge about $s_i(t)$ and the noise, we consider structured uncertainty classes and design minimax robust discriminators for problems $\mathcal{P}_1$ and $\mathcal{P}_2$.

We consider two distinct types of uncertainty classes: (i) classes determined by two alternating capacities and (ii) classes with minimum or maximum elements. The former uncertainty models have been very popular among the robust statisticians and engineers (see [1]-[6]) because they include several useful models of uncertainty such as ϵ-contaminated models, total variation models, band models, and p-point models and result in closed form expressions for the least-favorable elements in the class for a variety of performance measures. The latter uncertainty models also describe useful classes of uncertainty and result in convenient closed form expressions.

The design philosophy that we pursue is that of minimax robustness, which has received considerable attention for in the past fifteen years (see the tutorial [1]). According to it, the worst-case operational conditions are identified with regard to some performnace criterion of the decision design and the optimal such design for these conditions is derived. Subsequently this decision design is employed independently of actual conditions (which are not known, except for the fact that they belong to some structured uncertainty classes) and it is shown that it achieves desirable performance despite the uncertainty. In this paper the design criteria for problem $\mathcal{P}_1$ are the probabilities of false alarm and miss. For problem $\mathcal{P}_2$, since there is no closed form for the error probabilities under the two hypotheses, we derive the Chernoff bounds on the error probabilities, identify their connection with the discrimination measures known as the divergences, and use the divergences as the design criterion.

This paper is organized as follows. In Section II, we first review the definition and known properties of the uncertainty classes characterized by 2-alternating capacities and derive some new useful results for our problems. Then we describe the uncertainty classes with minimum and maximum elements. Finally, we cite the likelihood ratio functions for problems $\mathcal{P}_1$ and $\mathcal{P}_2$. In Section III, we derive the minimax robust fixed-observation-interval (block) detector for $\mathcal{P}_1$ under the probability of error criterion; we also show that the error probabilities of the robust test converge to zero exponentially as the duration of the observation interval increases, despite the uncertianty. In Section IV, we derive the minimax robust block test for $\mathcal{P}_2$ when the performance measures are the Chernoff bounds on the error probabilities. In Sections III and IV both the uncertainty classes determined by 2-alternating capacities and uncertainty classes with minimum and maximum elements are considered. In Section V, we derive the minimax robust sequential detector for a homogeneous $\mathcal{P}_1$–for which the signals $s_i(t)$ $(i = 0, 1)$ and the noise variance $\sigma(t)$ are time invariant; the uncertainty classes considered here are of the minimum/maximum-element type. Finally, in Section VI the results of the paper are summarized and conclusions are drawn.

II. PRELIMINARIES

In this section we describe the two uncertainty class models that we use in this paper and the associated results of minmax robustness. We also derive present the likelihood-ratios for problems $\mathcal{P}_1$ and $\mathcal{P}_2$.

II.A Uncertainty Classes Determined by 2-Alternating Capacities

First we provide a brief but self-contained review of known results from the theory of 2-alternating capacities and the associated theory of robustness of [6]. This review is necessary to introduce the concepts, notation, and fundamental results of minimax robustness within these uncertainty classes, which we use repeatedly in this paper. Then we derive a new result, Proposition 1, which extends the results of [6] and is used extensively in the proofs of our results on robust discrimination described in Sections III, IV, and V.

Definition: A positive finite set function v on a sample s space Ω and associated σ-field $\mathcal{F}$ is called a **2-alternating capacity**, if it is increasing, continuous from below, continuous from above on a closed set, and satisfies the conditions $v(\emptyset) = 0$ and $v(A \cup B) + v(A \cap B) \leq v(A) + v(B)$; the set function v may not be a measure.

Suppose now that $\mathbf{M}$ is the class of measures on $(\Omega, \mathcal{F})$ and $m \in \mathbf{M}$ is a measure. Consider the **uncertainty class** which is determined by the 2-alternating capacity v as follows

$$\mathbf{M_v} = \{m \in \mathbf{M} | m(A) \leq v(A),\ \forall A \in \mathcal{F},\ m(\Omega) = v(\Omega)\} \tag{1}$$

Thus all measures in the above uncertainty class have the same upper set function v. When Ω is compact, several popular uncertainty models like ϵ-contaminated neighborhoods [3], total variation neighborhoods [3], band classes [4], and p-point classes [5] are special cases of this model. The complete description of one of these classes, the ϵ-contaminated mixture model will be provided later in this section. Fundamental properties of the uncertainty model (1) have been studied by Huber and Strassen [6] which enable the robustification of the average risk criterion to these uncertainties. We state the relevant properties as Lemma 1.

Lemma 1: Suppose v_0 and v_1 are 2-alternating capacities on $(\Omega, \mathcal{F})$, $(\mathbf{M}_0, \mathbf{M}_1)$ are the uncertainty classes determined by (v_0, v_1) as in (1), and the two classes **do not overlap.** Then there exists a Lebesgue-measurable function $\pi_v : \Omega \longrightarrow [0, \infty]$ such that the average (Bayes) risk is minimized

$$\theta v_0(\{\pi_v > \theta\}) + v_1(\{\pi_v \leq \theta\}) \leq \theta v_0(A) + v_1(A^c) \tag{2}$$

for all $A \in \mathcal{F}$ and $\theta \geq 0$; θ can be interpreted as the ratio of the prior probabilities of the two hypotheses H_0 and H_1, whereas $\{\pi_v > \theta\}$ can be interpreted as the **likelihood ratio test** for v_0 versus v_1. Clearly, (2) together with (1) imply that

$$\theta m_0(\{\pi_v > \theta\}) + m_1(\{\pi_v \leq \theta\}) \leq \theta v_0(\{\pi_v > \theta\}) + v_1(\{\pi_v \leq \theta\}) \leq \theta v_0(A) + v_1(A^c) (*)$$

for all $m_0 \in \mathbf{M_0}$, $m_1 \in \mathbf{M_1}$, and $A \in \mathcal{F}$; this inequality establishes the **minimax robustness** of the test based on π_v. Furthermore, there exist measures $(\hat{m}_0, \hat{m}_1)$ in $\mathbf{M}_0 \times \mathbf{M}_1$ such that

$$\hat{m}_0(\{\pi_v > \theta\}) = v_0(\{\pi_v > \theta\}) \geq m_0(\{\pi_v > \theta\}) \tag{3}$$

$$\hat{m}_1(\{\pi_v \leq \theta\}) = v_1(\{\pi_v \leq \theta\}) \geq m_1(\{\pi_v \leq \theta\}) \tag{4}$$

The quantity π_v, which is termed the **Huber-Strassen derivative** of the classes $\mathbf{M_0} \times \mathbf{M_1}$, is a version of $d\hat{m}_1/d\hat{m}_0$ and is unique a.e. $[\hat{m}_0]$; it plays the role of the **worst-case likelihood ratio** for the two uncertainty classes $\mathbf{M_0}$ and $\mathbf{M_1}$. The dominance properties

(3) - (4) establish the existence of measures in the classes $\mathbf{M_0}$ and $\mathbf{M_1}$ that achieve the upper values provided by v_0 and v_1 for sets of the form $\{\pi_v \leq \theta\}$ and their complements. The measures $(\hat{m}_0, \hat{m}_1)$ are termed the **least-favorable** measures over $\mathbf{M_0} \times \mathbf{M_1}$.

Example: Consider the ϵ-contaminated mixture uncertainty classes of probability measures described by [2]

$$\mathbf{M_j} = \{m_j \in \mathbf{M} | m_j = (1-\epsilon_j)m_j^0 + \epsilon_j \tilde{m}_j, \tilde{m}_j(\Omega) = m_j^0(\Omega) = 1\},\ j = 0, 1.$$

which are determined by the known **nominal measures** m_0^0 and m_1^0 and the **degrees of uncertainty** ϵ_0 and ϵ_1 ($0 \leq \epsilon_i \leq 1$ for $i = 0, 1$); the unknown probability measures $\tilde{m}_j$ ($j = 0, 1$) are allowed to take any arbitrary values. This uncertainty class is appropriate for modeling situations in which the probability measures (or pdfs) governing the observations are convex combinations of known probability measures (pdfs) and arbitrary probability measures (pdfs). Then the associated 2-alternating capacities are for $j = 0, 1$

$$v_j(A) = \begin{cases} (1-\epsilon_j)m_j^0(A) + \epsilon_j, & A \neq \emptyset \\ 0, & A = \emptyset \end{cases}$$

and the least-favorable distributions are

$$d\hat{m}_0/d\lambda = \begin{cases} (1-\epsilon_0)dm_0^0/d\lambda, & dm_1^0/dm_0^0 \leq c_0 \\ \frac{1-\epsilon_0}{c_0} dm_1^0/d\lambda, & c_0 < dm_1^0/dm_0^0, \end{cases}$$

$$d\hat{m}_1/d\lambda = \begin{cases} (1-\epsilon_1)dm_1^0/d\lambda, & c_1 < dm_1^0/dm_0^0 \\ c_1(1-\epsilon_1)dm_0^0/d\lambda, & dm_1^0/dm_0^0 \leq c_1, \end{cases}$$

where λ is the Lebesgue measure and $0 \leq c_1 \leq c_0 < \infty$ are constants such that $\hat{m}_1(\Omega) = \hat{m}_0(\Omega) = 1$. The Huber-Strassen derivative π_v has the form

$$\pi_v = d\hat{m}_1/d\hat{m}_0 = \frac{1-\epsilon_1}{1-\epsilon_0} \min\{c_0, \max(c_1, dm_1^0/dm_0^0)\}.$$

which consists of a censored version of the nominal likelihood-ratio dm_1^0/dm_0^0.

The following proposition extends the dominance properties (3) and (4) of [6] to more general functionals than the error probabilities of the robust test:

Proposition 1: Suppose that the measures (m_0, m_1) on $(\Omega, \mathcal{F})$ belong to $\mathbf{M_0} \times \mathbf{M_1}$ characterized by (1) and that x is a real variable.

(i) If one of the following situations holds:

(a) both $g(\pi_v)$ and $h(x)$ are nonnegative, increasing functions of π_v (the Huber-Strassen derivative) and x, respectively;

(b) $g(\pi_v)$ is a nonnegative, decreasing function of π_v and $h(x)$ is a nonpositive and increasing function of x;

(c) $g(\pi_v)$ is a nonpositive, increasing function of π_v and $h(x)$ is a nonnegative and decreasing function of x;

(d) both $g(\pi_v)$ and $h(x)$ are nonpositive and decreasing functions of π_v and x, respectively,

then

$$\int_\Omega g(\pi_v(x))h(x)m_0(dx) \leq \int_\Omega g(\pi_v(x))h(x)\hat{m}_0(dx) \tag{5}$$

$$\int_\Omega g(\pi_v(x))h(x)m_1(dx) \geq \int_\Omega g(\pi_v(x))h(x)\hat{m}_1(dx) \tag{6}$$

(ii) If one of the following situations holds:

(a) both $g(\pi_v)$ and $h(x)$ are nonnegative, decreasing functions of π_v and x, respectively;

(b) $g(\pi_v)$ is a nonnegative, increasing function of π_v and $h(x)$ is a nonpositive, decreasing function of x;

(c) $g(\pi_v)$ is a nonpositive, decreasing function of π_v and $h(x)$ is a nonnegative and increasing function of x;

(d) both $g(\pi_v)$ and $h(x)$ are nonpositive and increasing functions of π_v and x, respectively, then

$$\int_\Omega g(\pi_v(x))h(x)m_0(dx) \geq \int_\Omega g(\pi_v(x))h(x)\hat{m}_0(dx) \tag{7}$$

$$\int_\Omega g(\pi_v(x))h(x)m_1(dx) \leq \int_\Omega g(\pi_v(x))h(x)\hat{m}_1(dx) \tag{8}$$

where $\hat{m}_0$ and $\hat{m}_1$ are singled out by Lemma 1.

Proof: We only prove (6) of part (i) for situation (a), since (6) for other situations, (5) and part (ii) can be proved in a similar way. Applying the transformation $z = \pi_v(x)$ we have

$$\int_\Omega g(\pi_v(x))h(x)m_1(dx) = \int_\Omega g(z)h(\pi_v^{-1}(z))m_1(dz)$$

for $i = 0, 1$, where π_v^{-1} is the inverse image of π_v. In our notation, $m_1(dx) = m_1(\{x < X \leq x + dx\}) = dM_1(x)$, where the associated distribution function is $M_1(x) = m_1(\{X \leq x\})$. Now, assuming that the conditions in (i) hold (i.e., $g'(x) \geq 0$, $h'(x) \geq 0$, $g(x) \geq 0$ and $h(x) \geq 0$, with the prime denoting a derivative with respect to the argument) and using the above transformation and integration by parts, we have

$$\begin{aligned}
&\int_\Omega g(\pi_v(x))h(x)m_1(dx) - \int_\Omega g(\pi_v(x))h(x)\hat{m}_1(dx) \\
= &\int_\Omega g(z)h(\pi_v^{-1}(z))m_1(dz) - \int_\Omega g(z)h(\pi_v^{-1}(z))\hat{m}_1(dz) \\
= &\int_\Omega \hat{m}_1(\{Z \leq z\})d[g(z)h(\pi_v^{-1}(z))] - \int_\Omega m_1(\{Z \leq z\})d[g(z)h(\pi_v^{-1}(z))] \\
= &\int_\Omega (\hat{m}_1(\{Z \leq z\}) - m_1(\{Z \leq z\}))h(\pi_v^{-1}(z))dg(z) \\
&+ \int_\Omega (\hat{m}_1(\{Z \leq z\}) - m_1(\{Z \leq z\}))g(z)dh(\pi_v^{-1}(z)) \\
= &\int_\Omega (\hat{m}_1(\{\pi_v(X) \leq z\}) - m_1(\{\pi_v(X) \leq z\}))h(\pi_v^{-1}(z))g'(z)dz \\
&+ \int_\Omega (\hat{m}_1(\{\pi_v(X) \leq \pi_v(t)\}) - m_1(\{\pi_v(X) \leq \pi_v(t)\}))g(\pi_v(t))h'(t)dt \geq 0
\end{aligned}$$

where $Z = \pi_v(X)$; the last inequality derives from (4) since $z = \pi_v(t) \geq 0$ for all t.

Remark 1: If either $g(x) = 1$ or $h(x) = 1$ for all x, i.e., one of the functions g or x is absent from the integrands of (5)-(8), the inequalities in (5)-(8) still hold; in this case, the nonnegativity or nonpositivity of the function involved is not a necessary condition.

Remark 2: Lemma 1 and Proposition 1 hold even if the 2-alternating capacity v_0 is itself a measure. In this case, the uncertainty class M_0 has a single element v_0.

II.B Uncertainty Classes with Minimum or Maximum Elements

In our cases of interest these uncertainty classes are of the form

$$\mathbf{S_{0,U}} = \{s_0(t) | s_0(t) \leq s_{0U}(t),\ 0 \leq t \leq T\}, \tag{9}$$

$$\mathbf{S_{1,L}} = \{s_1(t) | s_1(t) \geq s_{1L}(t),\ 0 \leq t \leq T\}, \tag{10}$$

and

$$\mathbf{\Sigma_U} = \{\sigma(t) | \sigma(t) \leq \sigma_U(t),\ 0 \leq t \leq T\}. \tag{11}$$

Uncertainty classes (9) and (11) are characterized by maximum elements $s_U(t)$ and $\sigma_U(t)$, respectively, while (10) is characterized by the minimum element $s_L(t)$. These uncertainty classes are easy to characterize and for the performance measures of interest their **least-favorable elements** turn out to be the maximum and minimum elements involved in the definition of the classes. These classes are appropriate in situations in which upper and/or lower bounds on the range of the system and noise parameters are available (can be estimated). To guarantee that the classes $\mathbf{S_{0,U}}$ and $\mathbf{S_{1,L}}$ **do not overlap** we impose the condition:

$$s_{1L}(t) \geq s_{0U}(t),\ 0 \leq t \leq T \tag{12}$$

II.C Likelihood Ratios for Problems $\mathcal{P}_1$ and $\mathcal{P}_2$

In the following we derive the likelihood ratio functions for $\mathcal{P}_1$ and $\mathcal{P}_2$. Notice that the observation process in $\mathcal{P}_1$ can be written in differential form as follows:

$$H_i: \; dY_t = s_i(t)dt + dW_t \, , \;\; i = 0, 1; \; 0 \leq t \leq T$$

where W_t is a Wiener process with zero mean function and intensity function $\sigma(t)$. Using this differential form of observations and manipulations analogous to the derivation of generalized likelihood ratio (see [7] and [8]), we give Lemma 2 without proof for the likelihood ratio function of $\mathcal{P}_1$.

Lemma 2: The likelihood ratio function for $\mathcal{P}_1$ has the following form

$$\ln L_T = \ln \frac{dP_1}{dP_0} = \int_0^T \frac{s_1(t) - s_0(t)}{[\sigma(t)]^2} Y_t dt - \int_0^T \frac{[s_1(t)]^2 - [s_0(t)]^2}{2[\sigma(t)]^2} dt \tag{13}$$

where P_i for $i = 0, 1$ are the joint probabilities of the observations $\{Y_t, \, 0 \leq t \leq T\}$ under H_i.

By using the explicit form of the solution for X_t (and thus for Y_t) to the stochastic differential equation in $\mathcal{P}_2$ (see Section 4.4 of [9]), we can derive the likelihood ratio function of $\mathcal{P}_2$, which is stated here as another lemma and its proof is omitted.

Lemma 3: The likelihood ratio function for $\mathcal{P}_2$ has the following form

$$\ln L_T = \ln \frac{dP_1}{dP_0} = \int_0^T \frac{s_1(t) - s_0(t)}{[\sigma(t)]^2} Y_t dY_t - \int_0^T \frac{[s_1(t)]^2 - [s_0(t)]^2}{2[\sigma(t)]^2} Y_t^2 dt \tag{14}$$

where P_i denotes the joint probabilities of the observations $\{Y_t, \, 0 \leq t \leq T\}$ under H_i, for $i = 0, 1$.

The likelihood ratio test is used as the decision rule, since it minimizes the expected risk and provides one side of the inequality for the saddle-point description of the minimax robust scheme [refer to equation (*)]. The criterion used for deriving the minimax robust scheme for $\mathcal{P}_1$ is the probability of false alarm, $\alpha(m_0, L_T)$, and the probability of miss, $\beta(m_1, L_T)$. Under mismatch, that is when the statistics $\hat{m}_0$ and $\hat{m}_1$ for which $\hat{L}_T = d\hat{m}_1/\hat{m}_0$ is derived are different than the actual m_0 and m_1 governing the observations, the probabilities of false alarm and miss become $\alpha(m_0, \hat{L}_T)$ and $\beta(m_1, \hat{L}_T)$. For $\mathcal{P}_2$ we use the Chernoff bounds on α and β as the criterion for robustness.

To prove minimax robustness one has to show that

$$\alpha(m_0, \hat{L}_T) \leq \alpha(\hat{m}_0, \hat{L}_T) \leq \alpha(\hat{m}_0, G_T) \tag{15}$$

and

$$\beta(m_1, \hat{L}_T) \leq \beta(\hat{m}_1, \hat{L}_T) \leq \beta(\hat{m}_1, G_T) \tag{16}$$

over all $m_0 \in \mathbf{M_0}$, $m_1 \in \mathbf{M_1}$, and arbitrary decision tests G_T based on the observation interval $[0, T]$; the right-hand inequalities in the equations above are satisfied immediately once the likelihood-ratio $\hat{L}_T$ has been derived. Therefore, in the following sections we only need to prove the left-hand inequalities in (15)-(16) for the various cases of interest in order to establish the minmax robustness of the likelihood ratio tests.

For convex classes $\mathbf{M_0}$ and $\mathbf{M_1}$ the above inequalities are equivalent to

$$max_{m_0}\alpha(m_0, L_T), \;\; max_{m_1}\beta(m_1, L_T) \tag{17}$$

As shown in [6] the least-favorable pair $(\hat{m}_0, \hat{m}_1)$ for general classes of probability measures $\mathbf{M_0}$ and $\mathbf{M_1}$ of the form of (1) for the problem

$$\max_{m_0} \alpha(m_0, L), \;\; \max_{m_1} \beta(m_1, L)$$

where $L = dm_1/dm_0$, is also the least-favorable pair for each one of the following problems:

(i) $\max\{\max_{m_0} \alpha(m_0, L), \max_{m_1} \beta(m_1, L)\}$;

(ii) $\max_{m_1} \beta(m_1, L)$ subject to $\max_{m_0} \alpha(m_0, L) \leq \alpha_0$;

(iii) $\max_{m_0, m_1}(\lambda\alpha(m_0, L) + (1-\lambda)\beta(m_1, L))$ with λ the known prior probability of H_0.

In the above discussion we omitted s_0, s_1 and σ in the argument of α and β. However, when we consider the uncertainty classes (9)-(11), s_0, s_1 and σ are the variables to be robustified and the above statements are still valid. In the following, we will show explicitly in our notation only the arguments of α and β that are to be robustified.

III. MINIMAX ROBUST BLOCK DISCRIMINATION FOR $\mathcal{P}_1$

III.A Uncertainty in the Signal

1. Uncertainty Class Characterized by 2-Alternating Capacities

In order to apply Lemma 1 and Proposition 1, the space $(\Omega, \mathcal{F}, m_i)$ $i = 0, 1$ is identified as the observation interval $[0, T]$ and the σ-field $\mathcal{F}$ is the one generated by all subsets of $[0, T]$. In this subsection, we consider the particular case where $s_0(t)$ is a known nonnegative time-varying function; thus there is now uncertainty under hypothesis H_0. Then the class $\mathbf{S_0}$ contains a single measure and is defined as

$$\mathbf{S_0} = \{s_0 | m_0(A) = \int_A s_0(t)\lambda(dt),\ \forall A \in \mathcal{F},\ \} \tag{18}$$

where $A \in \mathcal{F}$ and λ is the Lebesgue measure. Therefore, in this case, the probability of false alarm remains unchanged over $\mathbf{S_0}$: $\alpha(s_0(t), \hat{L}_T) = \alpha(\hat{s}_0(t), \hat{L}_T)$, so that we only need to consider the probability of miss. The uncertainty class under hypothesis H_1 is defined as

$$\mathbf{S_1} = \{s_1(t) | m_1(A) = \int_A |s_1(t)|\lambda(dt) \leq v_1(A),\ \forall A \in \mathcal{F},\ \int_\Omega |s_1(t)|dt = v_1(\Omega) = p_s\} \tag{19}$$

This definition allows for $s_1(t)$ to take both positive and negative values, since it is $|s_1(t)|$ which is involved in the definition of the measure m_1 and not $s_1(t)$ itself. It is assumed that $s_1(t) \neq s_0(t)$ for some $t \in [0, T]$ for all elements of the class (19) so that there is no overlap between $\mathbf{S_1}$ and $\mathbf{S_0}$.

As mentioned in Section II, the uncertainty class (19) includes several useful uncertainty models (including four of the most popular uncertainty class models). In adition to those classes, the following two other uncertainty classes based on norms can be considered:

$$\mathbf{S_2} = \{s_1(t) | \int_A |s_1(t) - s_1^0(t)|dt \leq \delta,\ \forall A \in \mathcal{F}\}\ \ (L^1 norm) \tag{20}$$

and

$$\mathbf{S_3} = \{s_1(t) | \sup_t |s_1(t) - s_1^0(t)| \leq \delta \} \ \ (L^\infty norm) \tag{21}$$

We can show that $\mathbf{S_j} \subset \mathbf{S_1}$ for $j = 2, 3$ as follows. For any $s_1(t) \in \mathbf{S_2}$ we have

$$\int_A |s_1(t)| dt - \int_A |s_1^0(t)| dt \leq \int_A |s_1(t) - s_1^0(t)| dt \leq \delta,$$

therefore, if we define $v_2(A) = \int_A |s_1^0(t)| dt + \delta$, then $s_1(t) \in \mathbf{S_1}$ as well. Similarly, for any $s_1(t) \in \mathbf{S_3}$ we have

$$\int_A |s_1(t)| dt - \int_A |s_1^0(t)| dt \leq \int_A |s_1(t) - s_1^0(t)| dt \leq \sup_t |s_1(t) - s_1^0(t)| \lambda(A) \leq \delta \lambda(A),$$

therefore, if we define $v_3(A) = \int_A |s_1^0(t)| dt + \delta\lambda(A)$, where $\lambda(A) \leq \lambda(\Omega) = T$, then $s_1(t) \in \mathbf{S_1}$, as well. Subsequently, since $\mathbf{S_j} \subset \mathbf{S_1}$ for $j = 2, 3$, the following inequality holds

$$\max_{s_1(t) \in \mathbf{S_j}} \beta(s_1(t), \hat{L}_T) \leq \max_{s_1(t) \in \mathbf{S_1}} \beta(s_1(t), \hat{L}_T), \ j = 2, 3. \tag{22}$$

where $\hat{s}_1(t) = |\hat{s}_1(t)|$ and $s_0(t)$ are the least-favorable densities for $\mathbf{S_1}$ vesus $\mathbf{S_0}$ and $\hat{L}_T$ is the likelihhod ratio for signals $\hat{s}_1(t)$ and $s_0(t)$ obtained from (13). Since there is no uncertainty under H_0, the probability of false alarm trivially satisfies

$$\max_{s_1(t) \in \mathbf{S_j}} \alpha(s_0(t), \hat{L}_T) \leq \max_{s_1(t) \in \mathbf{S_1}} \alpha(\hat{s}_0(t), \hat{L}_T), \ j = 2, 3,$$

with equality for $\hat{s}_0(t) = s_0(t)$.

Proposition 2 provides the minimax robust test for signal uncertainty in $\mathcal{P}_1$.

Proposition 2: For problem $\mathcal{P}_1$ and signal uncertainty within classes (19) and (18), if $0 \leq \hat{s}_1(t) < s_0(t)$, $s_0'(t) \leq 0$, and $\sigma'(t) \geq 0$ for $0 \leq t \leq T$, then the likelihood ratio test based on

$$\ln \hat{L}_T = \ln \frac{d\hat{P}_1}{d\hat{P}_0} = \int_0^T \frac{\hat{s}_1(t) - s_0(t)}{[\sigma(t)]^2} Y_t dt - \int_0^T \frac{[\hat{s}_1(t)]^2 - [s_0(t)]^2}{2[\sigma(t)]^2} dt \tag{23}$$

where $\hat{s}_1(t)$ and $\hat{s}_0(t) = s_0(t)$ are the least-favorable elements for $\mathbf{S_1}$ vs $\mathbf{S_0}$ of (19) and (18) singled out by Lemma 1, is minimax robust; that is, the two types of error probabilities for this test satisfy

$$\alpha(s_0(t), \hat{L}_T) \leq \alpha(\hat{s}_0(t), \hat{L}_T)$$

and

$$\beta(s_1(t), \hat{L}_T) \leq \beta(\hat{s}_1(t), \hat{L}_T) \tag{24}$$

Proof: From Lemma 2 we can obtain (23) under worst-case operational conditions. Since $s_0(t)$ is known, the inequality about the probability of false alarm is trivially satified with equality. Next we prove (24). Let

$$r_1 = \int_0^T \frac{[\hat{s}_1(t)]^2 - [s_0(t)]^2}{2[\sigma(t)]^2} dt; \;\; r_2 = \int_0^T \{\frac{\hat{s}_1(t) - s_0(t)}{[\sigma(t)]^2}\}^2 [\sigma(t)]^2 dt,$$

and define

$$f_1 = \int_0^T \frac{\hat{s}_1(t) - s_0(t)}{[\sigma(t)]^2} s_1(t) dt; \;\; \hat{f}_1 = \int_0^T \frac{\hat{s}_1(t) - s_0(t)}{[\sigma(t)]^2} \hat{s}_1(t) dt.$$

From $\hat{s}_1(t) \leq s_0(t)$ we have

$$\begin{aligned} f_1 &= \int_0^T \frac{\hat{s}_1(t) - s_0(t)}{[\sigma(t)]^2} s_1(t) dt \\ &\geq \int_0^T \frac{\hat{s}_1(t) - s_0(t)}{[\sigma(t)]^2} |s_1(t)| dt. \end{aligned}$$

If $s_0(t) \neq 0$, then since $\sigma'(t) \geq 0$ and $s_0'(t) \leq 0$ from (c) of part (i) in Proposition 1, we use (6) to obtain

$$\begin{aligned} f_1 &\geq \int_0^T \frac{\hat{s}_1(t) - s_0(t)}{[\sigma(t)]^2} |s_1(t)| dt \\ &= \int_0^T \left(\frac{\hat{s}_1(t)}{s_0(t)} - 1\right) \frac{s_0(t)}{[\sigma(t)]^2} |s_1(t)| dt \\ &\geq \int_0^T \left(\frac{\hat{s}_1(t)}{s_0(t)} - 1\right) \frac{s_0(t)}{[\sigma(t)]^2} \hat{s}_1(t) dt = \hat{f}_1. \end{aligned}$$

Finally, since $f_1 \geq \hat{f}_1$, it is easily observed that

$$\begin{aligned}\beta(s_1(t), \hat{L}_T) &= P_1\left(\ln\frac{d\hat{P}_1}{d\hat{P}_0} \leq \ln\eta\right) = \Phi\left(\frac{\ln\eta + r_1 - f_1}{\sqrt{r_2}}\right) \\ &\leq \Phi\left[\frac{\ln\eta + r_1 - \hat{f}_1}{\sqrt{r_2}}\right] = \beta(\hat{s}_1(t), \hat{L}_T)\end{aligned}$$

since $\Phi[(\ln\eta + r_1 - x)/\sqrt{r_2}]$ is a decreasing function of x.

2. Uncertainty Classes with Minimal and Maximal Elements

The minimax robust discriminator for the uncertainty classes (9) and (10) is given by the following proposition.

Proposition 3: For $\mathcal{P}_1$ and the uncertainty classes (9) and (10), if $s_{1L}(t) \geq s_{0U}(t)$ for $0 \leq t \leq T$, then the likelihood ratio test based on

$$\ln\hat{L}_T = \ln\frac{d\hat{P}_1}{d\hat{P}_0} = \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\sigma(t)]^2} Y_t dt - \int_0^T \frac{[\hat{s}_1(t)]^2 - [\hat{s}_0(t)]^2}{2[\sigma(t)]^2} dt \tag{25}$$

where $\hat{s}_0(t) = s_{0U}(t)$ and $\hat{s}_1(t) = s_{1L}(t)$, is minimax robust, that is, the error probabilities satisfy

$$\alpha(s_0(t), \hat{L}_T) \leq \alpha(\hat{s}_0(t), \hat{L}_T) \tag{26}$$

and

$$\beta(s_1(t), \hat{L}_T) \leq \beta(\hat{s}_1(t)\hat{L}_T), \tag{27}$$

Proof: We only give the proof of (26), since (27) can be proved in an identical manner. Let

$$r_1 = \int_0^T \frac{[\hat{s}_1(t)]^2 - [\hat{s}_0(t)]^2}{2[\sigma(t)]^2} dt; \;\; r_2 = \int_0^T \{\frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\sigma(t)]^2}\}^2 [\sigma(t)]^2 dt$$

and define

$$f_0 = \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\sigma(t)]^2} s_0(t) dt; \;\; \hat{f}_0 = \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\sigma(t)]^2} \hat{s}_0(t) dt.$$

Then $f_0 \le \hat{f}_0$ and

$$\begin{aligned}\alpha(s_0(t), \hat{L}_T) &= P_0\left(\ln \frac{d\hat{P}_1}{d\hat{P}_0} > \ln \eta\right) = Q\left[\frac{\ln \eta + r_1 - f_0}{\sqrt{r_2}}\right] \\ &\le Q\left[\frac{\ln \eta + r_1 - \hat{f}_0}{\sqrt{r_2}}\right] = \alpha(\hat{s}_0(t), \hat{L}_T)\end{aligned}$$

where we used the fact that $Q[(\ln \eta + r_1 - x)/\sqrt{r_2}]$ is an increasing function of x and that $\hat{s}_1(t) = s_{1L}(t) \ge s_{0U}(t) = \hat{s}_0(t)$.

Notice that the distance of the least-favorable pair $|\hat{s}_1(t) - \hat{s}_0(t)| = s_{1L}(t) - s_{0U}(t)$ is the shortest of all distances of the pair $(s_0(t), s_1(t)) \in \mathbf{S_{0,U}} \times \mathbf{S_{1,L}}$. This result is intuitively pleasing, since the closer the two signals $(s_0(t), s_1(t))$ are, the higher the error probabilities are expected to be.

III.B Uncertainty in the Noise Variance

1. Uncertainty Classes Characterized by 2-Alternating Capacities

In this subsection, we consider the following uncertainty class for $\sigma(t)$

$$\Sigma = \{\sigma(t) | m_(A) = \int_A [\sigma(t)]^2 \lambda(dt) \le v(A), \forall A \in \mathcal{F},\ \int_\Omega [\sigma(t)]^2 \lambda(dt) = v(\Omega) = p_N\} \quad (28)$$

where $m(A)$ defines a measure on $(\Omega, \mathcal{F})$ with Ω and $\mathcal{F}$ interpreted as in III.A. In this case $\mathbf{M_1} = \Sigma$ and $\mathbf{M_0} = \{\lambda\}$ and we assume that $\sigma(t) \ne 1$ for some $t \in [0, T]$ such that there is no overlap between $\mathbf{M_1}$ and $\mathbf{M_0}$. Note that the constraint $\int_\Omega [\sigma(t)]^2 \lambda(dt) = v(\Omega) = p_N$ implies a fixed noise power. The minimax robust discriminator for $\mathcal{P}_1$ is characterized by the following proposition.

Proposition 4: For problem $\mathcal{P}_1$ and noise uncertainty within the class (28), if (i) $s_1(t) \ge s_0(t)$ and $s_1'(t) \le s_0'(t)$ or (ii) $s_1(t) \le s_0(t)$ and $s_1'(t) \ge s_0'(t)$ for $0 \le t \le T$, then the likelihood ratio test based on

$$\ln \hat{L}_T = \ln \frac{d\hat{P}_1}{d\hat{P}_0} = \int_0^T \frac{s_1(t) - s_0(t)}{[\hat{\sigma}(t)]^2} Y_t dt - \int_0^T \frac{[s_1(t)]^2 - [s_0(t)]^2}{2[\hat{\sigma}(t)]^2} dt \quad (29)$$

where $[\hat{\sigma}(t)]^2$ is the Huber-Strassen derivative for $\mathbf{M_1} = \Sigma$ vs $\mathbf{M_0} = \{\lambda\}$ singled out by Lemma 1, is minimax robust; that is, the error probabilities satisfy

$$\alpha(\sigma(t), \hat{L}_T) \leq \alpha(\hat{\sigma}(t), \hat{L}_T) \tag{30}$$

and

$$\beta(\sigma(t), \hat{L}_T) \leq \beta(\hat{\sigma}(t), \hat{L}_T) \tag{31}$$

for all thresholds η satisfying

$$|\ln \eta| \leq \int_\Omega \frac{[s_1(t) - s_0(t)]^2}{2[\hat{\sigma}(t)]^2} dt \tag{32}$$

Proof: We only prove the part with assumption (i), since under (ii) it can be proved in a similar way. Let

$$r_1 = \int_0^T \frac{[s_1(t)]^2 - [s_0(t)]^2}{2[\hat{\sigma}(t)]^2} dt, \quad r_2 = \int_0^T \{\frac{s_1(t) - s_0(t)}{[\hat{\sigma}(t)]^2}\}^2 [\sigma(t)]^2 dt,$$
$$f_0 = \int_0^T \frac{s_1(t) - s_0(t)}{[\hat{\sigma}(t)]^2} s_0(t) dt, \quad f_1 = \int_0^T \frac{s_1(t) - s_0(t)}{[\hat{\sigma}(t)]^2} s_1(t) dt,$$

and

$$\hat{r}_2 = \int_0^T \{\frac{s_1(t) - s_0(t)}{[\hat{\sigma}(t)]^2}\}^2 [\hat{\sigma}(t)]^2 dt.$$

Under the assumption (i) $[s_1(t) - s_0(t)]^2$ is a decreasing function of t, thus from (a) of part (ii) in Proposition 1 we use (8) to obtain $r_2 \leq \hat{r}_2$. Then (30) follows by noticing that

$$\begin{aligned} \alpha(\sigma(t), \hat{L}_T) &= P_0\left(\ln \frac{d\hat{P}_1}{d\hat{P}_0} > \ln \eta\right) = Q\left[\frac{\ln \eta + r_1 - f_0}{\sqrt{r_2}}\right] \\ &\leq Q\left[\frac{\ln \eta + r_1 - f_0}{\sqrt{\hat{r}_2}}\right] = \alpha(\hat{\sigma}(t), \hat{L}_T) \end{aligned}$$

where we used the fact that, under (32), $\ln \eta + r_1 - f_0 \geq 0$ and thus $Q[(\ln \eta + r_1 - f_0)/x]$ is an increasing function of x. The inequality in (31) can be proved in a similar way by using the fact that $r_2 \leq \hat{r}_2$ and that, under (32), $\ln \eta + r_1 - f_1 \leq 0$, which in turn implies that $\Phi[(\ln \eta + r_1 - f_1)/x]$ is an increasing function of x.

2. Uncertainty Classes with a Maximal Element

The minimax robust scheme for the uncertainty class (11) is given as the following proposition.

Proposition 5: For problem $\mathcal{P}_1$ and the uncertainty class (11) the likelihood ratio test based on

$$\ln \hat{L}_T = \ln \frac{d\hat{P}_1}{d\hat{P}_0} = \int_0^T \frac{s_1(t) - s_0(t)}{[\hat{\sigma}(t)]^2} Y_t dt - \int_0^T \frac{[s_1(t)]^2 - [s_0(t)]^2}{2[\hat{\sigma}(t)]^2} dt \tag{33}$$

where $\hat{\sigma}(t) = \sigma_U(t)$, is a minimax robust test; that is

$$\alpha(\sigma(t), \hat{L}_T) \leq \alpha(\hat{\sigma}(t), \hat{L}_T) \tag{34}$$

and

$$\beta(\sigma(t), \hat{L}_T) \leq \beta(\hat{\sigma}(t), \hat{L}_T) \tag{35}$$

for all thresholds η satisfying

$$|\ln \eta| \leq \int_\Omega \frac{[s_1(t) - s_0(t)]^2}{2[\hat{\sigma}(t)]^2} dt \tag{36}$$

Proof: We prove (34) and (35) following steps similar to those in the proof of Proposition 4 and using the inequality

$$r_2 = \int_0^T \{\frac{s_1(t) - s_0(t)}{[\hat{\sigma}(t)]^2}\}^2 [\sigma(t)]^2 dt \leq \int_0^T \{\frac{s_1(t) - s_0(t)}{[\hat{\sigma}(t)]^2}\}^2 [\hat{\sigma}(t)]^2 dt = \hat{r}_2$$

III.C Uncertainty in the Signal and Noise Variance

In this subsection, we treat the general case where the signals and the variance of the noise are only known to belong to the uncertainty classes described in II. By examining

Propositions 2-5 we notice that the minimax robust scheme for this general case can be derived directly from the combination of their results. Let us now address the minimax robust tests for different combinations (different uncertainty classes corresponding to the signal and the noise) as four corollaries. Their proofs are omitted, since they can be deduced from the proofs of Propositions 2-5.

Corollary 1: For problem $\mathcal{P}_1$ with signal and noise uncertainty within classes (19) and (28), suppose that $s_0(t)$ is a known time-varying function, $0 \le \hat{s}_1(t) < s_0(t)$, $\hat{s}_1'(t) \ge s_0'(t)$, $s_0'(t) \le 0$, and $\hat{\sigma}'(t) \ge 0$ for $0 \le t \le T$, where $\hat{s}_1(t)$ and $\hat{\sigma}(t)$ are the least favorable elements of classes (19) and (28) singled out by Lemma 1 [applied to (19) vs (18) and (28) vs $\{\lambda\}$], then the likelihood ratio test based on

$$\ln \hat{L}_T = \ln \frac{d\hat{P}_1}{d\hat{P}_0} = \int_0^T \frac{\hat{s}_1(t) - s_0(t)}{[\hat{\sigma}(t)]^2} Y_t dt - \int_0^T \frac{[\hat{s}_1(t)]^2 - [s_0(t)]^2}{2[\hat{\sigma}(t)]^2} dt \tag{37}$$

is minimax robust; this implies that the error probabilities satisfy

$$\alpha(\sigma(t), \hat{L}_T) \le \alpha(\hat{\sigma}(t), \hat{L}_T) \tag{38}$$

and

$$\beta(s_1(t), \sigma(t), \hat{L}_T) \le \beta(\hat{s}_1(t), \hat{\sigma}(t), \hat{L}_T) \tag{39}$$

for all thresholds η satisfying

$$|\ln \eta| \le \int_\Omega \frac{[\hat{s}_1(t) - s_0]^2}{2[\hat{\sigma}(t)]^2} dt. \tag{40}$$

Corollary 2: For problem $\mathcal{P}_1$ with signal and noise uncertainty within classes (19) and (11), suppose that $s_0(t)$ is known, $0 \le \hat{s}_1(t) < s_0(t)$, $s_0'(t) \le 0$ $\hat{\sigma}'(t) \ge 0$ for $0 \le t \le T$, where $\hat{s}_1(t)$ is the least-favorable element of (19) [for testing (19) vs (18)] singled out by Lemma 1 and $\hat{\sigma}(t) = \sigma_U(t)$, then the likelihood ratio test based on (37) is a minimax robust test; that is, the error probabilities satisfy (38) and (39) for all thresholds η satisfying (40)

Corollary 3: For problem $\mathcal{P}_1$ with signal and noise uncertainty within classes (9), (10), and (28), if $s_{1L}(t) > s_{0U}(t)$ and $s'_{1L}(t) \leq s'_{0U}(t)$ for $0 \leq t \leq T$, where $\hat{s}_0(t) = s_{0U}(t)$, $\hat{s}_1(t) = s_{1L}(t)$ and $\hat{\sigma}(t)$ is singled out by Lemma 1 [for testing (28) vs $\{\lambda\}$], then the likelihood ratio test based on

$$\ln \hat{L}_T = \ln \frac{d\hat{P}_1}{d\hat{P}_0} = \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\hat{\sigma}(t)]^2} Y_t dt - \int_0^T \frac{[\hat{s}_1(t)]^2 - [\hat{s}_0(t)]^2}{2[\hat{\sigma}(t)]^2} dt \tag{41}$$

is minimax robust; in this case the error probabilities satisfy

$$\alpha(s_0(t), \sigma(t), \hat{L}_T) \leq \alpha(\hat{s}_0(t), \hat{\sigma}(t), \hat{L}_T) \tag{42}$$

and

$$\beta(s_1(t), \sigma(t), \hat{L}_T) \leq \beta(\hat{s}_1(t), \hat{\sigma}(t), \hat{L}_T) \tag{43}$$

for all thresholds η satisfying

$$|\ln \eta| \leq \int_\Omega \frac{[\hat{s}_1(t) - \hat{s}_0(t)]^2}{2[\hat{\sigma}(t)]^2} dt \tag{44}$$

Corollary 4: For problem $\mathcal{P}_1$ with signal and noise uncertainty within classes (9)-(11), suppose $s_{1L}(t) \geq s_{0U}(t)$ for $0 \leq t \leq T$, then the likelihood ratio test based on (41) with $\hat{s}_0(t) = s_{0U}(t), \hat{s}_1(t) = s_{1L}(t)$ and $\hat{\sigma}(t) = \sigma_U(t)$, is minimax robust; that is, (42) and (43) hold for all thresholds satisfying (44).

III.D Asymptotic Performance

It is interesting to examine for the general case of III.C the behavior of the two types of error probability as T (the length of the observation interval) increases. It is particularly desirable to ascertain that the robust likelihood ratio tests described in Corollary 1-4 maintain their robustness asymptotically (for large T). Under mismatch and for the threshold η, the **Chernoff bounds** on the two error probabilities have the forms

$$\alpha_U(s, s_0(t), \sigma(t), \hat{L}_T) = \exp\{-s\eta - TC_0(s, \hat{L}_T)\} \tag{45}$$

and

$$\beta_U(s, s_1(t), \sigma(t), \hat{L}_T) = \exp\{s\eta - TC_1(s, \hat{L}_T)\} \tag{46}$$

for all $s \in [0,1]$. In (45) and (46), the **Chernoff distances** $C_j(s, \hat{L}_T)$ ($j = 0,1$) under mismatch are defined as

$$C_0(s, \hat{L}_T) = -\ln[E_0\{\hat{L}_T^s\}]/T \tag{47}$$

and

$$C_1(s, \hat{L}_T) = -\ln[E_1\{\hat{L}_T^{-s}\}]/T. \tag{48}$$

For $\mathcal{P}_1$ we can easily obtain that $E_j\{\hat{L}^s\}$ are given by

$$E_j\{\hat{L}^s\} = E_j\{\exp[s\int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\hat{\sigma}(t)]^2} Y_t dt]\} \exp[-s\int_0^T \frac{[\hat{s}_1(t)]^2 - [\hat{s}_0(t)]^2}{2[\hat{\sigma}(t)]^2} dt] \tag{49}$$

for $j = 0,1$. If we define $Z_T = \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\hat{\sigma}(t)]^2} Y_t dt$, then Z_T is normally distributed with mean

$$\int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\hat{\sigma}(t)]^2} s_j(t) dt$$

under H_j, and variance

$$\int_0^T \{\frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\hat{\sigma}(t)]^2}\}^2 [\sigma(t)]^2 dt.$$

Upon substitution into (47)-(49) we obtain

$$C_0(s,\hat{L}_T)=\frac{1}{T}[c_1s^2+c_2s-s\int_0^T\frac{\hat{s}_1(t)-\hat{s}_0(t)}{[\hat{\sigma}(t)]^2}s_0(t)dt] \tag{50}$$

and

$$C_1(s,\hat{L}_T)=\frac{1}{T}[c_1s^2-c_2s+s\int_0^T\frac{\hat{s}_1(t)-\hat{s}_0(t)}{[\hat{\sigma}(t)]^2}s_1(t)dt] \tag{51}$$

where

$$c_1=-\frac{1}{2}\int_0^T\{\frac{\hat{s}_1(t)-\hat{s}_0(t)}{[\hat{\sigma}(t)]^2}\}^2[\sigma(t)]^2dt,\ c_2=\int_0^T\frac{[\hat{s}_1(t)]^2-[\hat{s}_0(t)]^2}{2[\hat{\sigma}(t)]^2}dt. \tag{52}$$

The following proposition guarantees the robustness of the likelihood ratio tests in Corollaries 1-4 for the Chernoff bounds on the error probabilities.

Proposition 6: For problem $\mathcal{P}_1$ with the uncertainty models for the signal and the noise described in Corollaries 1-4, the likelihood ratio tests and the least favorable pairs provided there are minimax robust and least-favorable, respectively, for the Chernoff bounds on error probabilities; that is, the following inequalities are satisfied for all s in $[0,1]$:

$$\alpha_U(s,s_0(t),\sigma(t),\hat{L}_T)\le\alpha_U(s,\hat{s}_0(t),\hat{\sigma}(t),\hat{L}_T), \tag{53}$$

and

$$\beta_U(s,s_1(t),\sigma(t),\hat{L}_T)\le\beta_U(s,\hat{s}_1(t),\hat{\sigma}(t),\hat{L}_T). \tag{54}$$

Proof: In order to prove (53) and (54) under the conditions of Corollary 1, it suffices to show that the following inequality involving the Chernoff distances is true

$$C_j(s,\hat{L}_T)\ge\hat{C}_j(s,\hat{L}_T) \tag{55}$$

for $j=0,1$, where $\hat{C}_j(s,\hat{L}_T)$ are the Chernoff distances for the matched case. Indeed, this is satisfied for $j=0,1$, because as shown in the proofs of Propositions 4 and 2:

$$\int_0^T\{\frac{\hat{s}_1(t)-\hat{s}_0(t)}{[\hat{\sigma}(t)]^2}\}^2[\sigma(t)]^2dt\le\int_0^T\{\frac{\hat{s}_1(t)-\hat{s}_0(t)}{[\hat{\sigma}(t)]^2}\}^2[\hat{\sigma}(t)]^2dt$$

and

$$\int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\hat{\sigma}(t)]^2} s_1(t)dt \geq \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\hat{\sigma}(t)]^2} \hat{s}_1(t)dt$$

where $\hat{s}_0(t) = s_0(t)$.

Similarly we can prove the above proposition under the conditions of Corollaries 2, 3, and 4. The following proposition holds for any of the robust tests described in Corollaries 1-4.

Proposition 7: For problem $\mathcal{P}_1$ with the uncertainty models for the signal and the noise described in Corollaries 1-4, if the likelihood ratio tests of (37) or (41) are employed, then the Chernoff upper bounds on the error probabilities of approach zero exponentially with increasing T despite the uncertainty .

Proof: Because of the definitions (45)-(46) and (55) (which was essential for proving Proposition 6), it suffices to show that

$$\hat{C}_j(s, \hat{L}_T) > 0 \tag{56}$$

for all s in $[0,1]$. Indeed, from (50) with $c1$ and c_2 as defined there but with $s_j(t)$ replaced by $\hat{s}_j$ for $j = 0, 1$ we obtain

$$\begin{aligned}
\hat{C}_0(s, \hat{L}_T) &= \frac{1}{T}[c_1 s^2 + c_2 s - s\int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\hat{\sigma}(t)]^2}\hat{s}_0(t)dt] \\
&= \frac{1}{T}s(1-s)\int_0^T \frac{[\hat{s}_1(t) - \hat{s}_0(t)]^2}{2[\hat{\sigma}(t)]^2}dt > 0
\end{aligned}$$

and similarly from (51)

$$\begin{aligned}
\hat{C}_1(s, \hat{L}_T) &= \frac{1}{T}[c_1 s^2 - c_2 s + s\int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\hat{\sigma}(t)]^2}\hat{s}_1(t)dt] \\
&= \frac{1}{T}s(1-s)\int_0^T \frac{[\hat{s}_1(t) - \hat{s}_0(t)]^2}{2[\hat{\sigma}(t)]^2}dt > 0.
\end{aligned}$$

Notice that the Chernoff distances in the matched case are strictly nonzero for $\hat{s}_1(t) \neq \hat{s}_0(t)$, $0 \leq t \leq T$. Also, $\hat{C}_1(s, \hat{L}_T) = \hat{C}_0(s, \hat{L}_T)$ for all s in $[0,1]$.

If the limit $\lim_{T\to\infty} \frac{1}{T} \int_0^T \frac{[\hat{s}_1(t)-\hat{s}_0(t)]^2}{[\hat{\sigma}(t)]^2} dt$ exists, then $\hat{C}_j(s, \hat{L})$ $(j = 0, 1)$ becomes independent of T and thus the Chernoff bounds on the error probabilities decrease exponentially to zero with constant rate, as T increases.

IV. MINIMAX ROBUST BLOCK DISCRIMINATION FOR $\mathcal{P}_2$

The criteria employed for the robust design in this section are the Chernoff bounds on the error probability and the divergences induced by them. The divergences under mismatch are defined as follows:

$$I_0(\hat{L}) = C_0'(s, \hat{L})|_{s=0} = -\int \ln \frac{d\hat{P}_1}{d\hat{P}_0} dP_0, \tag{57}$$

and

$$I_1(\hat{L}) = C_1'(s, \hat{L})|_{s=0} = \int \ln \frac{d\hat{P}_1}{d\hat{P}_0} dP_1, \tag{58}$$

where $'$ denotes derivative with respect to s; similarly the regular (matched) divergences are defined as

$$\hat{I}_0(\hat{L}) = \hat{C}_0'(s, \hat{L})|_{s=0} = -\int \ln \frac{d\hat{P}_1}{d\hat{P}_0} d\hat{P}_0 > 0, \tag{59}$$

and

$$\hat{I}_1(\hat{L}) = \hat{C}_1'(s, \hat{L})|_{s=0} = \int \ln \frac{d\hat{P}_1}{d\hat{P}_0} d\hat{P}_1 > 0, \tag{60}$$

The strict positivity of the matched divergences for $\hat{P}_1 \neq \hat{P}_0$ can be easily established with the help of Jensen's inequality. In [10] the divergences were considered in the context of robust discrimination problems from discrete-time observations. The usefulness of the divergence distance measures lies in (i) their relation to the Chernoff bounds on the error probabilities of the hypothesis testing problems, which enables us to carry over inequalities valid for the divergencxes to similar inequalities on the Chernoff distances, and (ii) the simplicity of their form which enables us to prove dominance properties (inequalities for minimax robustness) for them, which are impossible to prove directly for the error probabilities or the Chernoff bounds on them for a variety of problems with statistical uncertainty.

From the form of the Chernoff bounds (45)-(48), we notice that the the exponents of the bounds are convex functions of s in $[0,1]$. Therefore, in order to prove inequalities of the form (53)-(54) which are equivalent to inequalities of the form (55)

$$C_j(s, \hat{L}_T) \geq \hat{C}_j(s, \hat{L}_T)$$

$\forall s \in [0, \bar{s}]$ for some $\bar{s} > 0$, we need to prove the equivalent inequalities between the mismatch and matched divergences:

$$I_j(\hat{L}_T) = C_j'(s, \hat{L}_T)|_{s=0} \geq \hat{C}_j'(s, \hat{L}_T)|_{s=0} = \hat{I}_j(\hat{L}_T); \quad j = 0, 1, \tag{61}$$

Thus the final criterion turns out to be the divergences under two hypotheses, which can be easily obtained from the likelihood ratio function (14) and the following form of solution of Y_t in $\mathcal{P}_2$ (see [4.4 of 9]) under H_j $(j = 0, 1)$

$$Y_t = \exp[\int_0^t s_i(t)dt]X_0 + \int_0^t \exp[\int_\tau^t s_i(\tau_1)d\tau_1]dW_\tau, \;\; 0 \leq t \leq T \tag{62}$$

X_0 has a standard normal distribution and Y_t is independent of the noise $N_\tau = \dot{W}(\tau)$ for $\tau \geq t$ under either hypothesis. Substituting (62) into (14) and using the independence of Y_t and $\dot{W}_\tau$ for $0 \leq \tau \leq t$, we have the following general forms for the mismatch divergences of $\mathcal{P}_2$ under the assumption $E_j[Y_t^2] < \infty$ for $j = 0, 1$:

$$\begin{aligned}
I_0(s_0(t), s_1(t), \sigma(t), \hat{L}_T) &= -E_0\{\ln \hat{L}_T\} \\
&= -E_0\{-\int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{2[\hat{\sigma}(t)]^2}[\hat{s}_1(t) + \hat{s}_0(t) - 2s_0(t)]Y_t^2 dt \\
&\qquad + \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{\hat{\sigma}(t)} Y_t \dot{W}_t dt\} \\
&= \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{2[\hat{\sigma}(t)]^2}[\hat{s}_1(t) + \hat{s}_0(t) - 2s_0(t)]E_0\{Y_t^2\}dt \\
&= \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{2[\hat{\sigma}(t)]^2}[\hat{s}_1(t) + \hat{s}_0(t) - 2s_0(t)]
\end{aligned}$$

$$\{\exp[2\int_0^t s_0(\tau_1)d\tau_1] + \int_0^t \exp[2\int_{\tau_2}^t s_0(\tau_1)d\tau_1][\sigma(\tau_2)]^2 d\tau_2\} dt$$

(63)

$$\begin{aligned}
I_1(s_0(t), s_1(t), \sigma(t), \hat{L}_T) &= E_1\{\ln \hat{L}_T\} \\
&= E_1[\int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{2[\hat{\sigma}(t)]^2}[2s_1(t) - \hat{s}_1(t) - \hat{s}_0(t)]Y_t^2 dt \\
&\qquad + \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{\hat{\sigma}(t)} Y_t \dot{W}_t dt\} \\
&= \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{2[\hat{\sigma}(t)]^2}[2s_1(t) - \hat{s}_1(t) - \hat{s}_0(t)]E_1\{Y_t^2\} dt \\
&= \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{2[\hat{\sigma}(t)]^2}[2s_1(t) - \hat{s}_1(t) - \hat{s}_0(t)] \\
&\qquad \{\exp[2\int_0^t s_1(\tau_1)d\tau_1] + \int_0^t \exp[2\int_{\tau_2}^t s_1(\tau_1)d\tau_1][\sigma(\tau_2)]^2 d\tau_2\} dt
\end{aligned}$$

(64)

where $E_j[\dot{W}_u \dot{W}_v] = \sigma(u)\sigma(v)\delta(u - v)$ $(j = 0, 1)$. Furthermore, the matched divergences are

$$\begin{aligned}
I_0(\hat{s}_0(t), \hat{s}_1(t), \hat{\sigma}(t), \hat{L}_T) &= -\hat{E}_0\{\ln \hat{L}_T\} \\
&= \int_0^T \frac{[\hat{s}_1(t) - \hat{s}_0(t)]^2}{2[\hat{\sigma}(t)]^2} \\
&\qquad \{\exp[2\int_0^t \hat{s}_0(\tau_1)d\tau_1] + \int_0^t \exp[2\int_{\tau_2}^t \hat{s}_0(\tau_1)d\tau_1][\hat{\sigma}(\tau_2)]^2 d\tau_2\} dt
\end{aligned}$$

(65)

$$\begin{aligned}
I_1(\hat{s}_0(t), \hat{s}_1(t), \hat{\sigma}(t), \hat{L}_T) &= \hat{E}_1\{\ln \hat{L}_T\} \\
&= \int_0^T \frac{[\hat{s}_1(t) - \hat{s}_0(t)]^2}{2[\hat{\sigma}(t)]^2} \\
&\qquad \{\exp[2\int_0^t \hat{s}_1(\tau_1)d\tau_1] + \int_0^t \exp[2\int_{\tau_2}^t \hat{s}_1(\tau_1)d\tau_1][\hat{\sigma}(\tau_2)]^2 d\tau_2\} dt
\end{aligned}$$

(66)

Within this framework, we pursue the least-favorable pair within the class of likelihood ratio tests for the divergences, and thus for the Chernoff bounds on the two error probabilities, for different uncertainty classes. It turns out that we can prove the following proposition, in which the signal $s_0(t)$ under hypothesis H_0 is a known time function and the noise quantity $\sigma(t)$ in $\mathcal{P}_2$ belongs to the class:

$$\Sigma_L = \{\sigma(t)|\sigma(t) \geq \sigma_L(t), \ \ \forall t \in [0,T]\} \tag{67}$$

In comparing this class to that of (11) we notice that it is characterized by a minimum rather than a maximum element.

Proposition 8: For problem $\mathcal{P}_2$ with signal uncertainty within the class (10) and noise uncertainty within the class (67), if $\hat{s}_1(t) = s_{1,L}(t) \geq s_0(t) = \hat{s}_0(t)$ and $\hat{\sigma}(t) = \sigma_L(t)$, then the likelihood ratio test based on

$$\ln \hat{L}_T = \ln \frac{d\hat{P}_1}{d\hat{P}_0} = \int_0^T \frac{\hat{s}_1(t) - \hat{s}_0(t)}{[\hat{\sigma}(t)]^2} Y_t dY_t - \int_0^T \frac{[\hat{s}_1(t)]^2 - [\hat{s}_0(t)]^2}{2[\hat{\sigma}(t)]^2} Y_t^2 dt, \tag{68}$$

is minmax robust for the divergences and thus for the Chernoff bounds on the error probabilities under two hypotheses; that is

$$I_j(s_0(t), s_1(t)\sigma(t), \hat{L}_T) \geq I_j(\hat{s}_0(t), \hat{s}_1(t), \hat{\sigma}(t), \hat{L}_T); \ \ j = 0, 1 \tag{69}$$

Proof: By comparing (63) with (65) and (64) with (66), respectively, we can readily establish the desired inequalities (69) for $j = 0, 1$.

The following proposition follows directly from Proposition 8 in the same way that Proposition 7 followed from Proposition 6:

Proposition 9: Under the conditions of Proposition 8 the Chernoff upper bounds on the error probabilities approach zero exponentially with increasing T despite the uncertainty.

V. MINIMAX ROBUST SEQUENTIAL DISCRIMINATION FOR HOMOGENEOUS $\mathcal{P}_1$

Here we design a minimax robust sequential discriminator based on a robust version of the **sequential probability ratio test (SPRT)** for the case that the signals and the variance of noise in $\mathcal{P}_1$ are time invariant, that is, when $s_0(t) = s_0$, $s_1(t) = s_1$ and $\sigma(t) = \sigma$; in this case the observations follow the equation

$$dY_t = s_i dt + dW_t, \quad i = 0, 1$$

where W_t is a Wiener process with zero mean and intensity function σ. As in the previous sections, symbols with a hat denote least-favorable pairs of probabilities and the likelihood ratios based on them, symbols without a hat denote the actual operating conditions of $\mathcal{P}_1$. From Lemma 2 [eq. (13)] we know that for this particular case, the likelihood ratio function has the following form

$$\begin{aligned} \ln \hat{L}_\tau &= \ln \frac{d\hat{P}_1}{d\hat{P}_0} \\ &= \frac{\hat{s}_1 - \hat{s}_0}{\hat{\sigma}^2} \int_0^\tau Y_t dt - \frac{\hat{s}_1^2 - \hat{s}_0^2}{2\hat{\sigma}^2} \tau \end{aligned} \tag{70}$$

for observations over the interval $[0, \tau]$. In this section we also assume, without loss of generality, that $Y_0 = 0$. Let $\hat{\alpha}$ and $\hat{\beta}$ be the error probabilities under hypotheses H_0 and H_1, respectively, and $\hat{A}, \hat{B}$ be the two thresholds involved in the SPRT when the actual operating conditions are matched to the assumed signals and noise variance $\hat{s}_1, \hat{s}_0$, and $\hat{\sigma}$.

The SPRT operates as follows: if $\hat{L}_\tau \geq \hat{B}$ hypothesis H_1 is decided; if $\hat{L}_\tau \leq \hat{A}$ hypothesis H_0 is favored; finally, if $\hat{A} < \hat{L}_\tau < \hat{B}$ more observations are collected. The time τ that tany of the two thresholds is crossed is termed the stopping time. According to [11] for continuous-time SPRTs, $\hat{\alpha}, \hat{\beta}, \hat{A}$ and $\hat{B}$ are related as follows

$$\hat{A} = \frac{\hat{\beta}}{1-\hat{\alpha}} \tag{71}$$

and

$$\hat{B} = \frac{1-\hat{\beta}}{\hat{\alpha}} \tag{72}$$

As usual, we make the regularity condition that

$$\hat{\alpha} + \hat{\beta} < 1 \tag{73}$$

i.e., the false-alarm probability (α) is smaller than the power $(1-\beta)$ of the test. From (71)-(73) one obtains that $0 < \hat{A} < 1 < \hat{B}$.

Under mismatch, we have the following definitions for the error probabilities

$$\alpha = P_0(\hat{L}_\tau \geq \hat{B}),\ \beta = P_1(\hat{L}_\tau \leq \hat{A}),$$

while for the matched case

$$\hat{\alpha} = \hat{P}_0(\hat{L}_\tau \geq \hat{B}),\ \hat{\beta} = \hat{P}_1(\hat{L}_\tau \leq \hat{A})$$

Under mismatch the **expected stopping times** under the two hypotheses are denoted by $E_j\{\tau|\hat{L}_\tau\}$ $(j = 0,1)$; for the matched case they are denoted by $\hat{E}_j\{\tau|\hat{L}_\tau\}$ $(j = 0,1)$. Next we derive the following proposition for the situation characterized by mismatch:

Proposition 10: Suppose $\hat{A}$ and $\hat{B}$ $(0 < \hat{A} < 1 < \hat{B})$ are the thresholds of the SPRT for a homogeneous $\mathcal{P}_1$. If the signals and the noise variance $(\hat{s}_0, \hat{s}_1, \hat{\sigma})$ are used in determining the likelihood ratio of (70) which is employed by the SPRT, while the actual operating conditions are characterized by (s_0, s_1, σ), then the following identities hold

$$E_0\{\tau|\hat{L}_\tau\} = \frac{-2\hat{\sigma}^2\omega(\hat{\alpha},\hat{\beta},\alpha)}{(\hat{s}_1-\hat{s}_0)(2s_0-\hat{s}_1-\hat{s}_0)}, \tag{74}$$

$$E_1\{\tau|\hat{L}_\tau\} = \frac{2\hat{\sigma}^2\omega(\hat{\beta},\hat{\alpha},\beta)}{(\hat{s}_1-\hat{s}_0)(2s_1-\hat{s}_1-\hat{s}_0)}, \tag{75}$$

$$\alpha(s_0,\sigma,\hat{L}_\tau) = \frac{\hat{A}^{\gamma_0}-1}{(\hat{B}^{-1}\hat{A})^{\gamma_0}-1} \tag{76}$$

and

$$\beta(s_1,\sigma,\hat{L}_\tau) = \frac{\hat{B}^{\gamma_1}-1}{(\hat{A}^{-1}\hat{B})^{\gamma_1}-1} \tag{77}$$

where

$$\omega(\hat{x},\hat{y},x) = (1-x)\ln\frac{1-\hat{x}}{\hat{y}} + x\ln\frac{\hat{x}}{1-\hat{y}}, \tag{78}$$

$$\gamma_0 = \frac{\hat{\sigma}^2(2s_0-\hat{s}_1-\hat{s}_0)}{\sigma^2(\hat{s}_1-\hat{s}_0)}, \tag{79}$$

and

$$\gamma_1 = \frac{\hat{\sigma}^2(2s_1-\hat{s}_1-\hat{s}_0)}{\sigma^2(\hat{s}_1-\hat{s}_0)}; \tag{80}$$

Proof: First we prove (74) and (75). Let

$$I_t = \begin{cases} 1 & \text{if } t \leq \tau \\ 0 & \text{if } t > \tau \end{cases}$$

be the indicator function of the set $\{t \leq \tau\}$, then

$$\begin{aligned} E_0\{\int_0^\tau Y_t dt\} &= E_0\{\int_0^\infty Y_t I_t dt\} \\ &= \int_0^\infty E_0\{Y_t\}E_0\{I_t\}dt \\ &= s_0\int_0^\infty P_0(\tau \geq t)dt \\ &= s_0 E_0\{\tau\} \end{aligned}$$

where we used the fact that Y_t and I_t are independent (this is a result of the independence of Y_t and Y_u for $t \neq u$, the stopping time τ depends on all Y_u for $0 \leq u < \tau$ but not on the current value Y_t). Then, under the actual operation conditions, we have

$$\begin{aligned} E_0\{\ln \frac{d\hat{P}_1}{d\hat{P}_0}\} &= E_0\{\frac{\hat{s}_1 - \hat{s}_0}{\hat{\sigma}^2}\int_0^\tau Y_t dt - \frac{\hat{s}_1^2 - \hat{s}_0^2}{2\hat{\sigma}^2}\tau\} \\ &= \frac{(\hat{s}_1 - \hat{s}_0)(2s_0 - \hat{s}_1 - \hat{s}_0)}{2\hat{\sigma}^2} E_0\{\tau | \hat{L}_\tau\} \end{aligned}$$

and

$$\begin{aligned} E_0\{\ln \frac{d\hat{P}_1}{d\hat{P}_0}\} &= P_0(\hat{L}_\tau \geq \hat{B}) \ln \hat{B} + P_0(\hat{L}_\tau \leq \hat{A}) \ln \hat{A} \\ &= \alpha \ln \frac{1-\hat{\beta}}{\hat{\alpha}} + (1-\alpha) \ln \frac{\hat{\beta}}{1-\hat{\alpha}} \\ &= -\omega(\hat{\alpha}, \hat{\beta}, \alpha) \end{aligned}$$

Thus (74) follows from the above two expressions for $E_0\{\ln \hat{L}_\tau\}$. Eq. (75) can be shown in a similar way; the proof is omitted.

Next we prove (76) by using the method provided in Section 6.4 of [12] and omit the proof of (77). Let us define $h_1 = (\hat{s}_1 - \hat{s}_0)/\hat{\sigma}^2$ and $h_2 = (\hat{s}_1^2 - \hat{s}_0^2)/2\hat{\sigma}^2$. Then under the assumption $Y_0 = 0$, we have

$$\begin{aligned} \alpha(s_0, \sigma; \hat{s}_0, \hat{s}_1, \hat{\sigma}) &= P_0(\{\hat{L}_\tau \text{ } crosses \text{ } \ln \hat{B} \text{ } before \text{ } \ln \hat{A}\} | Y_0 = 0) \\ &= P_0(\{Z_t \text{ } crosses \text{ } \ln \hat{B} \text{ } before \text{ } \ln \hat{A}\} | Z_0 = 0) \end{aligned}$$

where $Z_t = h_1 \int_0^t Y_u du - h_2 t$ and, under H_0, Z_t has normal distribution with mean $(h_1 s_0 - h_2)t$ and variance $h_1^2 \sigma^2 t$. Then according to Section 6.4 of [11] we can think of α as $\alpha(0)$, where $\alpha(z)$ is a function of the initial condition $Z_0 = z$ and write the following expression

$$\alpha(z) = E_0[\alpha(z + Z)] + o(h)$$

where $o(h)$ denotes the probability that the process would have already crossed one of the barriers by the time h and $Z = Z_h - Z_0 = Z_h - z$. Notice that Z has a normal distribution with mean $(h_1 s_0 - h_2)h - z$ and variance $h_1^2\sigma^2 h$. Expanding $\alpha(z+Z)$ around $z+Z=z$ inside the expectation $E_0\{\}$ in the previous equation we obtain

$$\alpha(z) = E_0[\alpha(z) + \alpha'(z)Z + \alpha''(z)Z^2/2 + \cdots] + o(h)$$

where $\alpha'(z), \alpha''(z)\cdots$ denote the nth-derivative $(n = 1, 2, \cdots)$ of $\alpha(z)$ with respect to z. Letting $h \to 0$ and taking the expectation, we have the following differential equation

$$\frac{\hat{\sigma}(2s_0 - \hat{s}_0 - \hat{s}_1)}{2\sigma^2(\hat{s}_1 - \hat{s}_0)}\alpha'(z) + \frac{1}{2}\alpha''(z) = 0$$

By solving this differential equation with respect to $\alpha(z)$ and using the appropriate boundary conditions for the two barriers, $\alpha(\ln \hat{B}) = 1$, $\alpha(\ln \hat{A}) = 0$, we obtain for $z = Z_0 = 0$

$$\alpha = \alpha(0) = \frac{\hat{A}^{\gamma_0} - 1}{(\hat{B}^{-1}\hat{A})^{\gamma_0} - 1}$$

with γ_0 defined in (79). This completes the proof of this proposition.

Equipped with the above proposition, we now state and prove the main result of this section for the minimax robust discriminator of homogeneous $\mathcal{P}_1$.

Proposition 11: For problem $\mathcal{P}_1$ with signal and noise uncertainty within the homogeneous classes (9)-(11), suppose $\hat{s}_1 = s_{1L} > s_{0U} = \hat{s}_0$ and $\hat{\sigma} = \sigma_U$, and that the (73) and the following conditions on the desired error probabilities of the SPRT are satisfied:

$$\ln[(1-\hat{\beta})/\hat{\alpha}] \gg \hat{\beta}\ln[(1-\hat{\alpha})(1-\hat{\beta})/(\hat{\alpha}\hat{\beta})] \tag{81}$$

$$\ln[(1-\hat{\alpha})/\hat{\beta}] \gg \hat{\alpha}\ln[(1-\hat{\alpha})(1-\hat{\beta})/(\hat{\alpha}\hat{\beta})] \tag{82}$$

then the SPRT based on the likelihood ratio of (70) is minimax robust with respect to both the error probabilities and the expected stopping times under the two hypotheses; that is

$$\alpha(s_0, \sigma, \hat{L}_T) \leq \alpha(\hat{s}_0, \hat{\sigma}, \hat{L}_T) = \hat{\alpha} = \frac{1-\hat{A}}{\hat{B}-\hat{A}}, \tag{83}$$

$$\beta(s_1, \sigma, \hat{L}_\tau) \leq \beta(\hat{s}_1, \hat{\sigma}, \hat{L}_\tau) = \hat{\beta} = \frac{\hat{A}(\hat{B}-1)}{\hat{B}-\hat{A}}, \tag{84}$$

and

$$E_j\{\tau|\hat{L}_\tau\} \preceq \hat{E}_j\{\tau|\hat{L}_\tau\}, \;\; j = 0, 1 \tag{85}$$

Remark 1: Conditions (81)-(82) are not too restrictive, they are easily satisifed if both $\hat{\alpha}$ and $\hat{\beta}$ are smaller than 10^{-2}.

Remark 2: The inequality in (85) is not an inequality in the strict sense; it is however a valid inequality for all practical purposes when (81) and (82) are satisified.

Remark 3: As $\hat{\alpha} \to 0$ and $\hat{\beta} \to 0$ the stopping time $\tau \to \infty$ and (85) become inequalities in the strict sense:

$$E_0\{\tau|\hat{L}_\tau\} = \frac{2\hat{\sigma}^2 \ln \hat{A}}{(\hat{s}_1 - \hat{s}_0)(2s_0 - \hat{s}_1 - \hat{s}_0)} \leq \frac{-2\hat{\sigma}^2 \ln \hat{A}}{(\hat{s}_1 - \hat{s}_0)^2} = \hat{E}_0\{\tau|\hat{L}_\tau\} \tag{86}$$

$$E_1\{\tau|\hat{L}_\tau\} = \frac{2\hat{\sigma}^2 \ln \hat{B}}{(\hat{s}_1 - \hat{s}_0)(2s_1 - \hat{s}_1 - \hat{s}_0)} \leq \frac{2\hat{\sigma}^2 \ln \hat{B}}{(\hat{s}_1 - \hat{s}_0)^2} = \hat{E}_1\{\tau|\hat{L}_\tau\} \tag{87}$$

The above expected stopping times under the conditions $\hat{\alpha} \to 0$ and $\hat{\beta} \to 0$ are termed the asymptotic speeds of the SPRT.

Proof: We only show (83) and (85) for $j = 1$; a similar proof can be obtained for (84) and (85) for $j = 0$. First we prove (83). From Proposition 10 we notice that α depends on (s_0, σ) only through γ_0. Taking the derivative with respect to γ_0, we have

$$\frac{d\alpha(\gamma_0)}{d\gamma_0} = \frac{(\hat{B}^{-1}\hat{A})^{\gamma_0}}{\{(\hat{B}^{-1}\hat{A})^{\gamma_0} - 1\}^2}[(1 - \hat{B}^{\gamma_0}) \ln \hat{A} + (\hat{A}^{\gamma_0} - 1) \ln \hat{B}].$$

Let $G(\gamma_0) = (1 - \hat{B}^{\gamma_0}) \ln \hat{A} + (\hat{A}^{\gamma_0} - 1) \ln \hat{B}$. Then the sign of the above derivative is determined by the sign of $G(\gamma_0)$. Since the derivative of $G(\gamma_0)$ with respect to γ_0 is nonpositive

$$\frac{dG(\gamma_0)}{d\gamma_0} = \ln \hat{A} \ln \hat{B}(\hat{A}^{\gamma_0} - \hat{B}^{\gamma_0}) \leq 0$$

$$\frac{dG(\gamma_0)}{d\gamma_0} = \ln \hat{A} \ln \hat{B}(\hat{A}^{\gamma_0} - \hat{B}^{\gamma_0}) \leq 0$$

for $\gamma_0 \leq 0$ ($\ln \hat{A} < 0 < \ln \hat{B}$), then $G(\gamma_0)$ is a decreasing function of γ_0 for $\gamma_0 \leq 0$, and thus $G(\gamma_0) \geq G(0) = 0$ for $\gamma_0 \leq 0$. Consequently, $\alpha(\gamma_0)$ is an increasing function of γ_0 for $\gamma_0 \leq 0$. From the uncertainty classes (9) and (11) in the homogeneous case: $s_0 \leq s_{0U} = \hat{s}_0$ and $\sigma \leq \sigma_U = \hat{\sigma}$, and thus we can easily show that

$$\gamma_0 \leq -1$$

where the equality holds when $(s_0, \sigma) = (\hat{s}_0, \hat{\sigma})$. Therefore,

$$\alpha = \alpha(\gamma_0) \leq \alpha(-1) = \hat{\alpha}$$

which proves (83). Similarly, to prove (84) we first show that β is a decreasing function of γ_1 and that $\gamma_1 \geq 1$.

Next we show (85) for $j = 1$. Since from (84) $\beta \leq \hat{\beta}$ we use the assumption (81) to obtain

$$\ln[(1-\hat{\beta})/\hat{\alpha}] \gg \hat{\beta} \ln[(1-\hat{\alpha})(1-\hat{\beta})/(\hat{\alpha}\hat{\beta})] \geq \beta \ln[(1-\hat{\alpha})(1-\hat{\beta})/(\hat{\alpha}\hat{\beta})]$$

Then by using this inequality and the form of $E_1[\tau|\hat{L}_\tau]$ from Proposition 10, we have

$$\begin{aligned} E_1[\tau|\hat{L}_\tau] &= \frac{2\hat{\sigma}^2}{(\hat{s}_1 - \hat{s}_0)(2s_1 - \hat{s}_1 - \hat{s}_0)}[(1-\beta)\ln\frac{1-\hat{\beta}}{\hat{\alpha}} + \beta \ln \frac{\hat{\beta}}{1-\hat{\alpha}}] \\ &\approx \frac{2\hat{\sigma}^2}{(\hat{s}_1 - \hat{s}_0)(2s_1 - \hat{s}_1 - \hat{s}_0)} \ln \frac{1-\hat{\beta}}{\hat{\alpha}} \end{aligned}$$

Using (81) directly and the form of $\hat{E}_1[\tau|\hat{L}_\tau]$ from Proposition 10 for the matched case, we obtain

$$\begin{aligned} \hat{E}_1[\tau|\hat{L}_\tau] &= \frac{2\hat{\sigma}^2}{(\hat{s}_1 - \hat{s}_0)^2}[(1-\hat{\beta})\ln\frac{1-\hat{\beta}}{\hat{\alpha}} + \hat{\beta} \ln \frac{\hat{\beta}}{1-\hat{\alpha}}] \\ &\approx \frac{2\hat{\sigma}^2}{(\hat{s}_1 - \hat{s}_0)^2} \ln \frac{1-\hat{\beta}}{\hat{\alpha}} \end{aligned}$$

Both the above equations are approximations. However, as noted in Remark 1 above, conditions (81)-(82) are easy to satisfy, and this guarantees that the deviations of the numerators in both of the above equations from the common value $\ln \frac{1-\beta}{\alpha}$ are negligible. Since the denominators are expected to be substantially different for almost all elements s_0 and s_1 in the uncertainty classes, for all practical purposes the above approximations maintain the distinct features of the two quantities and allow us to compare them with each other without any measurable loss in accuracy. Indeed, by using the fact that $s_1 \geq s_{1L} = \hat{s}_1$ we obtain

$$E_1[\tau|\hat{L}_\tau] \preceq \hat{E}_1[\tau|\hat{L}_\tau]$$

which completes the proof.

VI. CONCLUSIONS

In this paper motivated by practical situations in target discrimination we considered binary hypothesis testing problems in continuous-time characterized by observations which (i) consist of distinct signals in additive white Gaussian noise or (ii) are the output of stochastic dynamical systems driven by white Gaussian noise. The statistics of the observations in both problems were only partially known. The goal of the paper was to robustify the continuous-time discrimination tests against statistical uncertainty in the observations. In particular, the signals in the first model, the parameters of the dynamical systems in the second model, and the autocorrelation functions of the noise in both models were modeled to belong to one of the following distinct uncertainty classes: (1) classes determined by 2-alternating capacities and (2) classes with minimum or maximum elements. These uncertainty classes include many popular models as subcases and can model effectively many practical situations. In the course of reviewing the theory of 2-alternating capacities we also extended Huber's results on minmax robustness within such classes to more general functionals than the average risk.

We then derived minimax robust discrimination tests (a) with a fixed observation interval and (b) sequential tests. The likelihood ratios of all the robust tests depended on the least-favorable pairs of parameters in the aforementioned uncertainty classes and were shown to have an acceptable level of performance despite the uncertainty. For the robust tests with a fixed observation interval the performance measures considered for problem (i) were the actual error probabilities under the two hypotheses and the Chernoff upper bounds on them; the latter were shown to preserve their desirable asymptotic properties (in the sense that they approach zero exponentially as the observation interval increases) in the presence of the uncertainties. For problem (ii) we provided results only for the Chernoff bounds and the associated divergences. Finally, we derived the robust sequential test for problem (i) with time-invariant parameters; in this case we used as performance measures the error probabilities and the average required length of the observation interval under each hypothesis.

References

[1] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: a survey," *Proc. IEEE*, pp. 433-481, Mar. 1985.

[2] E. A. Geraniotis and H. V. Poor, "Minimax discrimination for observed Poisson processes with uncertain rate functions," *IEEE Transactions on Information Theory*, Vol. IT-31, pp. 660-669, September 1985.

[3] P. J. Huber, "A robust version of the probability ratio test," *Ann. Math. Stat.*, vol. 36, pp. 1753-1758, 1965.

[4] S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 242-247, 1981.

[5] K. S. Vastola and H. V. Poor, "On the *p*-point uncertainty class," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 374-376, Mar. 1984.

[6] P. J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities," *Ann. Statist.*, vol. 1, pp. 251-265, 1973.

[7] T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 350-361, May 1969.

[8] T. E. Duncan, "Probability densities for diffusion processes with applications to nonlinear filtering theory and detection theory," Ph. D. dissertation, Dep. Elec. Eng., Stanford Univ., Stanford, Calif., May 1967.

[9] H. J. Larson and B. O. Shubert, *Probabilistic Models in Engineering Sciences: Volume II*. New York: John Wiley & Sons 1979.

[10] E. A. Geraniotis, "Performance bounds for discrimination problems with uncertainty statistics," *IEEE Trans. Inform. Theory,* vol. IT-31, pp. 703-707, September 1985.

[11] A. Wald, *Sequential Analysis.* New York: John Wiley & Sons 1974.

[12] S. M. Ross, *Stochastic Processes.* New York: John Wiley & Sons 1983.