Robust hypothesis testing with a relative entropy tolerance

Bernard C. Levy

Abstract

This paper considers the design of a minimax test for two hypotheses where the actual probability densities of the observations are located in neighborhoods obtained by placing a bound on the relative entropy between actual and nominal densities. The minimax problem admits a saddle point which is characterized. The robust test applies a nonlinear transformation which flattens the nominal likelihood ratio in the vicinity of one. Results are illustrated by considering the transmission of binary data in the presence of additive noise.

Index Terms

Robust hypothesis testing, Kullback-Leibler divergence, min-max problem, saddle point, least favorable densities.

I. Introduction

Robust hypothesis testing and signal detection problems have been examined in detail over the last 40 years [1], [2]. The purpose of such studies is to design tests or detectors which are insensitive to modelling errors. Specifically, whereas standard Bayesian or Neyman-Pearson tests are designed for nominal observation probability distributions, their performance may degrade rapidly when the actual model deviates only moderately from the nominal model. To guard against modelling errors, a minimax framework is usually adopted for selecting tests or detectors. In this context, the goal is to design a test that minimizes the worst-case performance for all observation models in a properly specified neighborhood of the nominal model. For robust hypothesis testing, when the neighborhood of the nominal model under each hypothesis corresponds either to a contamination model or a proximity model based on the Kolmogorov metric or a variant thereof, Huber [2]–[4] showed that the minimax detector applies a clipping transformation to the nominal likelihood ratio function. The clipping effect is achieved by shifting small portions of the probability mass under each hypothesis to the tail sections where errors occur. This relatively minute shift of probability mass can result in a significant degradation in test performance.

We adopt here a minimax formulation of the robust hypothesis testing problem of the same type as [2]–[4]. The only difference is that the neighborhood where the actual observation probability density is located under each hypothesis is formed by placing an upper bound on the relative entropy of the actual density with respect to the nominal density. To justify the choice of the relative entropy as a measure of proximity between statistical models, observe that Huber's work addresses primarily situations where statistical models are obtained directly from imperfect data, possibly contaminated by outliers. However there exists also situations where the densities employed in hypothesis testing are model based, arising from physical considerations,

B. C. Levy is with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616, USA (e-mail: levy@ece.ucdavis.edu).

possibly with a few unknown parameters which are estimated from the data. In this context, the relative entropy is a natural metric for model mismatch, since it provides the underlying metric for establishing the convergence of the expectation-maximization method [5] of mathematical statistics. In fact from a differential geometric viewpoint, it is argued in [6] that the relative entropy forms a natural 'distance' between statistical models. More recently, in the context of estimation and filtering it was shown in [7], [8] that minimax filters based on a relative entropy tolerance take the form or risk-sensitive Wiener or Kalman filters, which have well known robustness properties. By selecting the relative entropy as a measure of model mismatch, a risk-sensitive viewpoint was also adopted recently in [9] for developing robust macroeconomic policies. Given relative entropy neighborhoods of the nominal densities for the two hypotheses, it is easy to verify that a saddle point exists for the resulting minimax hypothesis testing problem. To identify the saddle point, two assumptions are made. First as in [3], it is assumed that the nominal likelihood ratio function (LR) is monotone increasing. Second, it is required that the nominal densities under the two hypotheses should be symmetric with respect to each other. This allows the parametrization of the robust test and least-favorable densities in terms of a single parameter which can be selected uniquely so that the relative entropy tolerance is satisfied. The least-favorable LR is expressed as a nonlinear transformation of the nominal LR. But, unlike [2]–[4], the transformation is not a clipping transformation. Instead, it attempts to drive the LR to a value as close one as possible. The least-favorable densities are divided into three segments. The extreme segments are scaled versions of the nominal densities, where the scaling aims at shifting some probability mass to tails where errors occur. But the middle segment is a section of the "mid-way density" on the geodesic linking the two nominal densities, where the mid-way density is characterized by the property that it has the same relative entropy with respect to each of the nominal densities.

The robust hypothesis testing problem we consider is also related to the worst-case noise detection problem examined in [10], [11], where given a binary communication system with additive noise, with the actual noise density located within a prespecified relative entropy bound of the nominal noise density, it is required to find the ML detector for the worst-case noise in the neighborhood of the nominal noise. Thus the difference between the problem we consider and [11] is that we allow the additive noise statistics to be different under each hypothesis, instead of forcing them to be the same. Finally, it is worth noting that [12] also examines robust hypothesis testing by using the relative entropy as a mismatch metric between actual and nominal densities, but it does so asymptotically as the number of measurements becomes infinite, so its results take a very different form.

The paper is organized as follows. Section II describes the minimax hypothesis problem with a relative entropy constraint. The saddle point of the problem is characterized in Section III, and examples are presented in Section IV. Finally, Section V gives some conclusions.

II. PROBLEM FORMULATION

Consider a binary hypothesis testing problem where under hypothesis H_j, with $j = 0, 1$, the random observation $Y \in \mathbb{R}$ admits $f_j(y)$ as nominal probability density. The actual density $g_j(y)$ of Y under H_j is not known exactly and belongs to the neighborhood

$$\mathcal{F}_j = \{g_j : D(g_j|f_j) \leq \epsilon_j\}, \tag{2.1}$$

where

$$D(g|f) = \int_{-\infty}^{\infty} \ln\left(\frac{g(y)}{f(y)}\right) g(y) dy \tag{2.2}$$

denotes the Kullback-Leibler (KL) divergence or relative entropy of probability densities $g(y)$ and $f(y)$. Note that the KL divergence is not a true distance since it is not symmetric, i.e., $D(g|f) \neq D(f|g)$, it does not satisfy the triangle inequality, but $D(g|f) \geq 0$ with equality if and only if $g = f$. Also, since $x \ln(x)$ is a convex function for $x \geq 0$, $D(g|f)$ is convex in g, which implies that neighborhood $\mathcal{F}_j$ is convex for $j = 0, 1$.

Let $\mathcal{D}$ denote the class of pointwise randomized decision rules $\delta(y)$ such that if $Y = y$, we select H_1 with probability $\delta(y)$ and H_0 with probability $1 - \delta$, where $0 \leq \delta(y) \leq 1$. Clearly $\mathcal{D}$ is convex, since if $\delta_1(y)$ and $\delta_2(y)$ are two decision rules of $\mathcal{D}$, then for $0 \leq \alpha \leq 1$,

$$\delta(y) = \alpha \delta_1(y) + (1 - \alpha)\delta_2(y)$$

also belongs to $\mathcal{D}$.

Let

$$P_F(\delta, g_0) = \int_{-\infty}^{\infty} \delta(y)g_0(y)dy \tag{2.3}$$

$$P_M(\delta, g_1) = \int_{-\infty}^{\infty} (1 - \delta(y))g_1(y)dy \tag{2.4}$$

denote respectively the probability of false alarm and the probablity of a miss for decision rule $\delta \in \mathcal{D}$ when the densities of Y under H_0 and H_1 are g_0 and g_1, respectively. Note that $P_F(\delta, g_0)$ is separately linear in δ and g_0. Similarly $P_M(\delta, g_1)$ is separately linear in δ and g_1. If we assume that the two hypotheses are equally likely, the probability of error of $\delta \in \mathcal{D}$ is given by

$$P_E(\delta, g_0, g_1) = \frac{1}{2}[P_F(\delta, g_0) + P_M(\delta, g_1)] . \tag{2.5}$$

We seek to solve the minimax problem

$$\min_{\delta \in \mathcal{D}} \max_{(g_0, g_1) \in \mathcal{F}_0 \times \mathcal{F}_1} P_E(\delta, g_0, g_1) \tag{2.6}$$

Note that $P_E(\delta, g_0, g_1)$ is linear and thus convex in δ. Similarly, it is linear and thus concave in g_0 and g_1. The set $\mathcal{F}_0 \times \mathcal{F}_1$ is convex and compact, $\mathcal{D}$ is convex and since

$$||\delta||_\infty = \max_{y \in \mathbb{R}} \delta(y)$$

for all $\delta \in \mathcal{D}$, $\mathcal{D}$ is compact with respect to the infinity norm. So according to the Von Neumann minimax theorem [13, p. 319], there exists a saddle point $(\delta_R, (g_0^L, g_1^L))$ for the minimax problem (2.6). Here δ_R is the robust/minimax test, whereas g_0^L and g_1^L are the least favorable densities in $\mathcal{F}_0 \times \mathcal{F}_1$. The saddle point is characterized by the property

$$P_E(\delta, g_0^L, g_1^L) \geq P_E(\delta_R, g_0^L, g_1^L) \geq P_E(\delta_R, g_0, g_1) \tag{2.7}$$

for all $\delta \in \mathcal{D}$, $g_0 \in \mathcal{F}_0$ and $g_1 \in \mathcal{F}_1$.

While it is nice to know that a saddle point exists, exhibiting a test δ_R and least favorable densities g_j^L, $j = 0, 1$ satisfying (2.7) is a nontrivial task. Before doing so, it is worth pointing out that the minimax problem (2.6) is of the same type as considered by Huber in [2]–[4]. The only difference is that the neighborhoods $\mathcal{F}_j$ differ from those considered in [2] which included contamination models or proximity models based on the Kolmogorov metric as special cases.

The problem (2.6) is also closely related to the worst-case noise detection problem considered in [11], where for hypotheses

$$
\begin{aligned}
H_0 &: \quad Y = -1 + N \\
H_1 &: \quad Y = 1 + N \,,
\end{aligned}
\tag{2.8}
$$

and a nominal probability density $f_N(n)$ for noise N, it was desired to construct a minimum probability of error detector for the least-favorable noise density $g_N(n)$ located in the KL ball specified by $D(g_N|f_N) \leq \epsilon$. Thus the problem (2.6) differs from the one examined in [10], [11] by the fact that we allow the least-favorable noise distribution to be different under hypotheses H_0 and H_1, instead of insisting they should be the same.

III. SADDLE POINT SPECIFICATION

The first inequality of the saddle point characterization (2.7) indicates that the robust test δ_{R} must be the optimum Bayesian test for the least-favorable pair $(g_0^{\mathrm{L}}, g_1^{\mathrm{L}})$. So if

$$
L_{\mathrm{L}}(y) = \frac{g_1^{\mathrm{L}}(y)}{g_0^{\mathrm{L}}(y)}
\tag{3.1}
$$

denotes the LR function for the pair $(g_0^{\mathrm{L}}, g_1^{\mathrm{L}})$, we need to have

$$
\delta_{\mathrm{R}}(y) = \begin{cases} 1 & \text{for} \quad L_{\mathrm{L}}(y) > 1 \\ \text{arbitrary} & \text{for} \quad L_{\mathrm{L}}(y) = 1 \\ 0 & \text{for} \quad L_{\mathrm{L}}(y) < 1 \,. \end{cases}
\tag{3.2}
$$

Consider now the second inequality of (2.7). Because of the form (2.5) of $P_E(\delta, g_0, g_1)$, it is equivalent to

$$
\begin{aligned}
P_F(\delta_{\mathrm{R}}, g_0^{\mathrm{L}}) &\geq P_F(\delta_{\mathrm{R}}, g_0) \\
P_M(\delta_{\mathrm{R}}, g_1^{\mathrm{L}}) &\geq P_M(\delta_{\mathrm{R}}, g_1)
\end{aligned}
$$

for all g_0 and g_1 in $\mathcal{F}_0$ and $\mathcal{F}_1$, respectively.

So, given δ_{R}, the least-favorable density g_0^{L} is obtained by maximizing $P_F(\delta_{\mathrm{R}}, g_0)$ for all functions $g_0 \in \mathcal{F}_0$ such that

$$
I(g_0) = \int_{-\infty}^{\infty} g_0(y)dy = 1 \,.
\tag{3.3}
$$

Since $P_F(\delta_{\mathrm{R}}, g_0)$ is concave in g_0 and the domain $\mathcal{F}_0$ is convex, the maximization can be accomplished by using the method of Lagrange multipliers [14, Chap. 5]. Consider the Lagrangian

$$
\begin{aligned}
L(g_0, \lambda, \mu) &= P_F(\delta_{\mathrm{R}}, g_0) + \lambda(\epsilon_0 - D(g_0|f_0)) + \mu(1 - I(g_0)) \\
&= \int_{-\infty}^{\infty} \left[\delta_{\mathrm{R}}(y) - \mu - \lambda \ln\left(\frac{g_0}{f_0}(y)\right)\right] g_0(y)dy + \lambda\epsilon_0 + \mu \,,
\end{aligned}
\tag{3.4}
$$

where Lagrange multiplier $\lambda \geq 0$ is associated to the inequality constraint $D(g_0|f_0) \leq \epsilon_0$, whereas multiplier μ corresponds to equality constraint (3.3). Note that the non-negativity constraint $g_0(y) \geq 0$ for the density function g_0 is not introduced explicitly, since the solution obtained below by maximizing L satisfies this constraint automatically.

The Gateaux derivative [14, p. 17] of L with respect to g_0 in the direction of an arbitrary function z is given by

$$
\begin{aligned}
\nabla_{g_0,z} L(g_0, \lambda, \mu) &= \lim_{h \to 0} \frac{1}{h} \big[L(g_0 + hz, \lambda, \mu) - L(g_0, \lambda, \mu) \big] \\
&= \int_{-\infty}^{\infty} \big[\delta_{\mathrm{R}} - (\lambda + \mu) - \lambda \ln \big(\tfrac{g_0}{f_0} \big) \big] z \, dy,
\end{aligned} \tag{3.5}
$$

and since $z(y)$ is arbitrary, this implies

$$
\delta_{\mathrm{R}}(y) - (\lambda + \mu) - \lambda \ln \big(\tfrac{g_0}{f_0} \big)(y) = 0. \tag{3.6}
$$

In addition, the Karush-Kuhn-Tucker (KKT) condition

$$
\lambda(\epsilon_0 - D(g_0 | f_0)) = 0 \tag{3.7}
$$

needs to be satisfied. Assume $\lambda > 0$, so $D(g_0 | f_0) = \epsilon_0$, i.e., g_0 is on the boundary of $\mathcal{F}_0$. Then (3.6) implies

$$
g_0^{\mathrm{L}}(y) = \frac{1}{Z_0} \exp(\alpha_0 \delta_{\mathrm{R}}(y)) f_0(y) \tag{3.8}
$$

with

$$
Z_0 \triangleq \exp(1 + \tfrac{\mu}{\lambda}) \quad, \quad \alpha_0 \triangleq \frac{1}{\lambda}.
$$

Note that since the nominal density $f_0(y) \geq 0$ for all y, the least-favorable density $g_0^{\mathrm{L}}(y)$ specified by (3.8) is also non-negative, so that the non-negativity constraint on g_0 is satisfied automatically. Proceeding in a similar manner, we find that the least-favorable density under H_1 can be expressed as

$$
g_1^{\mathrm{L}}(y) = \frac{1}{Z_1} \exp(\alpha_1 (1 - \delta_{\mathrm{R}}(y))) f_1(y) \tag{3.9}
$$

with $Z_1 > 0$.

Together, the expressions (3.2) for δ_{R} and (3.8)–(3.9) for $(g_0^{\mathrm{L}}, g_1^{\mathrm{L}})$ provide some guidelines for guessing a saddle point satisfying inequalities (2.7). We exhibit below a saddle point with the desired structure under the following assumptions.

Assumptions:

i) The nominal likelihood ratio

$$
L(y) = \frac{f_1(y)}{f_0(y)} \tag{3.10}
$$

is a monotone increasing function of y. This implies that $\ell = L(y)$ admits an inverse function $y = L^{-1}(\ell)$.

ii) $f_0(y)$ and $f_1(y)$ admit the symmetry

$$
f_1(y) = f_0(-y). \tag{3.11}
$$

This assumption implies

$$
L(-y) = \frac{f_1(-y)}{f_0(-y)} = \frac{f_0(y)}{f_1(y)} = \frac{1}{L(y)},
$$

and thus $L(0) = 1$.

Remarks:

a) The motonicity assumption for $L(y)$ appears also in [2]. The symmetry condition (3.11) has the effect of symmetrizing the KL divergence of f_0 and f_1, since it ensures

$$D(f_1|f_0) = D(f_0|f_1) \,.$$

Furthermore, for $0 \le u \le 1$, if we consider the geodesic

$$f_u(y) = \frac{f_1^u(y) f_0^{1-u}(y)}{Z(u)}$$

linking nominal densities f_0 and f_1, where

$$Z(u) = \int_{-\infty}^{\infty} f_1^u(y) f_0^{1-u}(y) dy \,,$$

the assumption ii) ensures that the density $f_{1/2}$ is located mid-way between f_0 and f_1 in terms of the KL divergence, since

$$D(f_{1/2}|f_0) = D(f_{1/2}|f_1) \,.$$

We refer the reader to [15, Chap. 4.] and [6] for a detailed discussion of the differential geometric structure of statistical models.

b) For model (2.8), the above assumptions are satisfied if under both hypotheses N admits a generalized Gaussian density

$$f_N(n) = a \exp(-|n/b|^{\alpha})$$

with $\alpha > 1$, where the constants a and b are adjusted to fix the variance of the distribution and normalize its total probability mass. The case $\alpha = 2$ corresponds to a standard Gaussian distribution. On the other hand, if N is Cauchy distributed, it is easy to verify that

$$L(y) = \frac{f_N(y-1)}{f_N(y+1)}$$

is not monotone increasing so Assumption i) is not satisfied.

c) The assumptions allow the consideration of nonsymmetric noise distributions. For example, consider model (2.8) where under H_1, N admits the asymmetric Laplace density

$$f_L(n) = \begin{cases} c \exp(-an) & n \ge 0 \\ c \exp(bn) & n \le 0 \end{cases} \tag{3.12}$$

with $b > a > 0$ and $c = (a^{-1} + b^{-1})^{-1}$, and under H_0, N admits the flipped density $f_L(-n)$. Then

$$f_1(y) = f_L(y-1) \quad \text{and} \quad f_0(y) = f_L(-(y+1))$$

satisfy the symmetry condition (3.11) and the log-likelihood ratio

$$\ln L(y) = \ln(f_1(y)/f_0(y)) = \begin{cases} (b-a)y + (b+a) & y \ge 1 \\ 2by & -1 \le y \le 1 \\ (b-a)y - (b+a) & y \le -1 \end{cases} \tag{3.13}$$

is monotone increasing. Note that this property requires $b > a$, which ensures that the fat tails of $f_1(y)$ and $f_0(y)$ are located on the opposite side of the location parameter of the competing hypothesis. For example under H_1, the location parameter (the constant additive

term in (2.8)) is 1 so the fat tail extends over $[1, \infty)$, which is on the opposite side of the location parameter -1 of the competing hypothesis H_0.

We can now prove the following result.

Theorem 1: Assume that constants ϵ_j specifying neighborhoods $\mathcal{F}_j$ with $j = 0, 1$ are such that $\epsilon_0 = \epsilon_1 = \epsilon$, where

$$0 < \epsilon < D(f_{1/2}|f_0) \,. \tag{3.14}$$

This requirement ensures that $\mathcal{F}_0$ and $\mathcal{F}_1$ do not intersect. Then under assumptions i)-ii) consider the decision rule

$$\delta_{\mathrm{R}}(y) = \begin{cases} 1 & y > y_U \\ \frac{1}{2}\left[1 + \frac{\ln L(y)}{\ln \ell_U}\right] & -y_U \leq y \leq y_U \\ 0 & y < -y_U \,, \end{cases} \tag{3.15}$$

and the least-favorable pair

$$g_0^{\mathrm{L}}(y) = \begin{cases} \ell_U f_0(y)/Z(y_U) & y > y_U \\ \ell_U^{1/2} f_1^{1/2}(y) f_0^{1/2}(y)/Z(y_U) & -y_U \leq y \leq y_U \\ f_0(y)/Z(y_U) & y < -y_U \end{cases} \tag{3.16}$$

$$g_1^{\mathrm{L}}(y) = \begin{cases} f_1(y)/Z(y_U) & y > y_U \\ \ell_U^{1/2} f_1^{1/2}(y) f_0^{1/2}(y)/Z(y_U) & -y_U \leq y \leq y_U \\ \ell_U f_1(y)/Z(y_U) & y < -y_U \end{cases} \tag{3.17}$$

which are parametrized by $y_U > 0$ and $\ell_U = L(y_U) > 1$. Here the normalizing constant $Z(y_U)$ is selected such that

$$I(g_0^{\mathrm{L}}) = I(g_1^{\mathrm{L}}) = 1 \,. \tag{3.18}$$

There exists a unique $y_U > 0$ such that

$$D(g_0^{\mathrm{L}}|f_0) = D(g_1^{\mathrm{L}}|f_1) = \epsilon \,, \tag{3.19}$$

and the corresponding δ_{R} and densities $(g_0^{\mathrm{L}}, g_1^{\mathrm{L}})$ form a saddle point of minimax problem (2.6).

Before proving the result, it is worth noting that the least-favorable LR

$$L_{\mathrm{L}}(y) = \begin{cases} \dfrac{L(y)}{\ell_U} > 1 & y > y_U \\ 1 & -y_U \leq y \leq y_U \\ \ell_U L(y) < 1 & y < -y_U \end{cases} \tag{3.20}$$

can be viewed as obtained by applying a nonlinearity $q(\cdot)$ to the nominal likelihood ratio L. Specifically, we have

$$L_{\mathrm{L}} = q(L) = \begin{cases} L/\ell_U & L > \ell_U \\ 1 & \ell_U^{-1} \leq L \leq \ell_U \\ \ell_U L & L < \ell_U^{-1} \end{cases} \tag{3.21}$$

where the nonlinearity $q(\cdot)$ is sketched in Fig. 1 below. This nonlinearity is different from the clipping transformation obtained by Huber [2]–[4] which truncates high and low values of the nominal likelihood ratio. Instead, the transformation $q(\cdot)$ attempts to force the transformed values

L_L to be as close to 1 as possible, where a LR value $L_\mathrm{L} = 1$ corresponds to a situation where observation $Y = y$ is uninformative in terms of making a decision between H_1 and H_0.

Fig. 1. Nonlinearity $q(\cdot)$ relating the nominal and least-favorable likelihood ratios.

Proof: Observe first that since the least-favorable LR is given by (3.20), the decision rule δ_R specified by (3.15) has the form (3.2). Note that since $\ell_U^{-1} \le L(y) \le \ell_U$ for $-y_U \le y \le y_U$, we have

$$-1 \le \frac{\ln L(y)}{\ln \ell_U} \le 1$$

for $-y_U \le y \le y_U$, which ensures $0 \le \delta_\mathrm{R}(y) \le 1$ for $-y_U \le y \le y_U$.

Next, with δ_R given by (3.15). it is easy to verify that the least favorable densities g_0^L and g_1^L given by (3.16) and (3.17) admit the forms (3.8) and (3.9) with $Z_0 = Z_1 = Z(y_U)$ and

$$\alpha_0 = \alpha_1 = \ln \ell_U .$$

To ensure that the normalization condition (3.18) holds we only need to select

$$Z(y_U) = \int_{-\infty}^{-y_U} f_0(y)dy + \ell_U^{1/2} \int_{-y_U}^{y_U} f_1^{1/2}(y) f_0^{1/2}(y)dy + \ell_U \int_{y_U}^{\infty} f_0(y)dy .$$

Then if $g_0^\mathrm{L}(\cdot|y_U)$ represents the function (3.16), where the parametrization by $y_U \ge 0$ is written explicitly, let

$$
\begin{aligned}
D(y_U) &\overset{\triangle}{=} D(g_0^\mathrm{L}(\cdot|y_U)|f_0) \\
&= -\ln Z(y_U) + \frac{1}{Z(y_U)}\Big[\ell_U \ln \ell_U \int_{y_U}^{\infty} f_0(y)dy \\
&\qquad\qquad + \ell_U^{1/2} \ln \ell_U \int_0^{y_U} f_1^{1/2}(y) f_0^{1/2}(y)dy \Big]
\end{aligned}
\tag{3.22}
$$

denote its KL divergence with respect to the nominal density f_0. For $y_U = 0$, we have $g_0^\mathrm{L}(\cdot|0) = f_0$, so $D(0) = 0$. Furthermore for $y_U = +\infty$, we have $g_0^\mathrm{L}(\cdot|+\infty) = f_{1/2}$, so $D(+\infty) = D(f_{1/2}|f_0)$, where as noted earlier the density $f_{1/2}$ represents the mid-way point on the geodesic linking f_0 to f_1.

8

Taking the derivative of $D(y_U)$ with respect to y_U gives

$$
\begin{aligned}
\frac{dD}{dy_U} &= -Z^{-1}(y_U)\frac{dZ}{dy_U} \\
&\quad -Z^{-2}(y_U)\frac{dZ}{dy_U}\left[\ell_U \ln \ell_U \int_{y_U}^{\infty} f_0(y)dy + l_U^{1/2}\ln \ell_U \int_{-y_U}^{y_U} f_1^{1/2}(y)f_0^{1/2}(y)dy\right] \\
&\quad +Z^{-1}(y_U)\left[\frac{d}{dy_U}(\ell_U \ln \ell_U)\int_{y_U}^{\infty} f_0(y)dy + \frac{d}{dy_U}(\ell_U^{1/2}\ln \ell_U)\int_0^{y_U} f_1^{1/2}(y)f_0^{1/2}(y)dy\right] \\
&= \frac{N(y_U)}{Z^2(y_U)}\frac{dL}{dy_U}\,, \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (3.23)
\end{aligned}
$$

where

$$
\begin{aligned}
N(y_U) &= \ln \ell_U \int_{y_U}^{\infty} f_0(y)dy \int_{y_U}^{\infty} f_1(y)dy \\
&\quad + \frac{1}{2}\ln \ell_U \left[\ell_U^{1/2}\int_0^{y_U} f_1^{1/2}(y)f_0^{1/2}(y)dy + \int_{y_U}^{\infty}\left(f_0(y)+\frac{f_1(y)}{\ell_U}\right)dy\right] > 0
\end{aligned}
$$

for $y_U > 0$. Since $L(y)$ is monotone increasing, we have $dL/dy_U > 0$ in (3.23), so $dD/dy_U > 0$. Consequently, $D(y_U)$ is monotone increasing from $D(0) = 0$ for $y_U = 0$ to $D(f_{1/2}|f_0)$ for $y_U = \infty$. Accordingly, given ϵ satisfying (3.14), there exists a unique y_U such that $D(y_U) = \epsilon$. For this choice of y_U, the least favorable densities g_0^{L} and g_1^{L} satisfy KKT condition (3.9), so the second inequality of (2.8) is satisfied, and δ_{R} together with $(g_0^{\mathrm{L}}, g_1^{\mathrm{L}})$ form the desired saddle point. $\qquad\square$

Worst case test performance: By taking into account the symmetries

$$
\begin{aligned}
1 - \delta_{\mathrm{R}}(y) &= \delta_{\mathrm{R}}(-y) \\
g_1^{\mathrm{L}}(y) &= g_0^{\mathrm{L}}(-y) \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (3.24)
\end{aligned}
$$

of the robust test and least favorable densities, which are a consequence of the symmetry assumption (3.11), we find that the worst-case probabilities of false alarm and of a miss for test δ_{R} satisfy

$$
P_F(\delta_{\mathrm{R}}, g_0^{\mathrm{L}}) = P_M(\delta_{\mathrm{R}}, g_1^{\mathrm{L}}) = P_E(\delta_{\mathrm{R}}, g_0^{\mathrm{L}}, g_1^{\mathrm{L}})\,,
$$

where

$$
\begin{aligned}
P_F(\delta_{\mathrm{R}}, g_0^{\mathrm{L}}) &= \int_{-y_U}^{y_U} \delta_{\mathrm{R}}(y)g_0^{\mathrm{L}}(y)dy + Z^{-1}(y_U)\ell_U \int_{y_U}^{\infty} f_0(y)dy \\
&= Z^{-1}(y_U)\left[\ell_U^{1/2}\int_0^{y_U} f_1^{1/2}(y)f_0^{1/2}(y)dy + \ell_U \int_{y_U}^{\infty} f_0(y)dy\right]. \quad (3.25)
\end{aligned}
$$

IV. EXAMPLES

Example 1: Consider the case where under H_0 and H_1, Y admits the nominal distributions

$$
\begin{aligned}
f_0(y) &= \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left(-\frac{(y+1)^2}{2\sigma^2}\right) \\
f_1(y) &= \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left(-\frac{(y-1)^2}{2\sigma^2}\right). \quad\quad\quad\quad\quad\quad (4.1)
\end{aligned}
$$

This corresponds to a model of the form (2.8) where the additive noise N has a $N(0, \sigma^2)$ nominal distribution. The signal to noise ratio (SNR) for this detection problem is $\text{SNR} = 1/\sigma^2$. The likelihood ratio

$$L(y) = \frac{f_1(y)}{f_0(y)} = \exp\left(\frac{2y}{\sigma^2}\right)$$

is clearly monotone increasing, and the nominal densities $f_j(y)$, $j = 0, 1$ admit the symmetry (3.11), so the assumptions of Theorem 1 are satisfied. In this case, it is interesting to note that the mid-way density

$$f_{1/2}(y) = \frac{f_1^{1/2}(y) f_0^{1/2}(y)}{Z(1/2)} = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{y^2}{2\sigma^2}\right),$$

is $N(0, \sigma^2)$ distributed, which makes sense since f_0 and f_1 have opposite means ∓ 1 but the same variance σ^2.

If we consider the parametrization (3.16) of the least favorable density $g_0^{\mathrm{L}}(y)$, we find that it is continuous and formed by three segments. Over $(-\infty, -y_U)$, g_0^{L} is an attenuated version of the nominal $N(-1, \sigma^2)$ density. Over $[-y_U, y_U]$, it is a scaled version of the mid-way $N(0, \sigma^2)$ density, and for (y_U, ∞) it is an amplified version of the nominal $N(-1, \sigma^2)$ density. Thus g_0^{L} can be viewed as obtained from the nominal density f_0 by shifting a portion of its probability mass to the middle segment where g_0^{L} and g_1^{L} are equal, and to the right tail where hypothesis H_1 is selected, which has the effect of increasing the probability of false alarm.

Fig. 2.   Plot of function $D(y_U)$ for 0dB SNR.

To illustrate the construction of $g_0^{\mathrm{L}}(y)$, let the relative entropy tolerance be $\epsilon = 0.1$. Then for a nominal SNR equal to 0dB ($\sigma = 1$), the function $D(y_U)$ measuring the KL divergence of $g_0^{\mathrm{L}}(\cdot|y_U)$ with respect to f_0 is plotted in Fig. 2. As expected, it is monotone increasing and

attains the desired tolerance value $\epsilon = 0.1$ for $y_U = 0.6080$. The least-favorable density $g_0^{\mathrm{L}}(y)$ is plotted together with the nominal density $f_0(y)$ in part a) of Fig. 3.

(a)

(b)

Fig. 3.   Least favorable density $g_0^{\mathrm{L}}(y)$ for a tolerance $\epsilon = 0.1$ and a) SNR =0dB, b) SNR = 10dB.

The three segments of the density described earlier are clearly in evidence in this plot. Note however that as the SNR increases, the middle segment shrinks. For example, the least-favorable density for a SNR value of 10dB is shown in part b) of Fig. 3. Although the KL tolerance $\epsilon = 0.1$

is the same as in part a), the deviation of g_0^{L} away from f_0 is much smaller than for a SNR value of 0dB. Note also that g_0^{L} is not symmetric about -1 since a fraction of the probability mass has been transferred from the left tail to the right tail in the direction of the location parameter 1 of the competing hypothesis H_1. Similarly the least favorable distribution $g_1^{\mathrm{L}}(y) = g_0^{\mathrm{L}}(-y)$ transfers a portion of its probability mass from its right tail to its left tail. In terms of model (2.8), this means that the least favorable densities of the noise N are different under H_0 and H_1, since one tilts rightward while the other tilts leftward. In contrast, [11] requires that the least-favorable noise should be the same under both hypotheses. For the above example with $\epsilon = 0.1$ and 10dB SNR, the least favorable noise density is plotted in the SouthWest corner of Figure 3 of [11]. It is symmetric and thus differs from the least-favorable densities obtained here.

Finally, for $\epsilon = 0.01$ and 0.1, and for SNR values between 0 and 15dB, the worst-case performance of the robust test δ_{R} given by (3.25) is compared in Fig. 4 with the probability of error $P_E = Q(\mathrm{SNR}^{1/2})$ of the maximum likelihood detector for nominal densities (4.1). As indicated by the figure, the loss of performance is rather spectacular. Of course, since this performance represents a worst case situation, it is not truly indicative of the degradation incurred for more benign choices of densities g_j in $\mathcal{F}_j$ with $j = 0, 1$.

Fig. 4. Comparison of the worst case probability of error of test δ_{R} for $\epsilon = 0.01$ and $\epsilon = 0.1$ with the ML probability of error for the nominal model.

Example 2: Consider model (2.8) where under H_1 N admits the asymmetric Laplace density $f_L(n)$ given by (3.12) with $b > a$ and under H_0, N admits the flipped density $f_L(-n)$. Then the densities

$$f_1(y) = f_L(y - 1) \quad , \quad f_0(y) = f_L(-(y + 1))$$

satisfy the symmetry condition (3.11), and as indicated by (3.13), the likelihood ratio $L(y)$ is

monotone increasing. In this case, the half-way density

$$
\begin{aligned}
f_{1/2}(y) &= \frac{f_0^{1/2}(y) f_1^{1/2}(y)}{Z(1/2)} \\
&= \begin{cases} c \exp(-b)/Z(1/2) & -1 \le y \le 1 \\ c \exp\left(-\frac{a+b}{2}|y| + \frac{a-b}{2}\right)/Z(1/2) & |y| \ge 1 \end{cases}
\end{aligned}
$$

with

$$
Z(1/2) = 2c \exp(-b)\left(1 + \frac{2}{a+b}\right)
$$

is constant for $-1 \le y \le 1$ and has a symmetrized exponential decay rate for its two tails. For $y_U < 1$, the parametrization (3.16) of the least favorable density g_0^{L} indicates that over segments $(-\infty, -y_U)$ and (y_U, ∞) it is proportional to f_0, but over $[-y_U, y_U]$ it is constant since $f_{1/2}$ is constant.

To illustrate this feature the nominal and least favorable densities are plotted in Fig. 5 for $a = 2$, $b = 4$, and $\epsilon = 0.1$. For this choice of parameters $y_U = 0.3640$.

Fig. 5. Nominal asymmetric Laplace density f_0 and least favorable density g_0^{L} for $a = 2$, $b = 4$ and tolerance $\epsilon = 0.1$.

V. CONCLUSION

A minimax hypothesis testing procedure has been derived for a binary hypothesis testing problem where the actual observation density under each hypothesis is required to be within a fixed KL ball centered about the nominal density. The robust test applies a nonlinear transformation which flattens the nominal LR in the vicinity of $L = 1$. The least-favorable densities include three segments where, quite interestingly, the middle segment is formed by a section of the density located mid-way on the geodesic linking the nominal densities under the two hypotheses.

The results were derived under a motonicity condition for the LR as well as a symmetry condition for the two hypotheses. While the first condition is benign and appears in Huber's work [2]–[4], it would be desirable to remove the symmetry condition (3.11), since this would open the way to the study of more general robust signal detection problems of the type discussed in [1].

REFERENCES

[1] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: a survey", *Proc. IEEE*, vol. 73, pp. 433–482, March 1985.

[2] P. J. Huber, *Robust Statistics*. New York: J. Wiley, 1981.

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

[4] P. J. Huber, "Robust confidence limits," *Z. Wahrcheinlichkeitstheorie verw. Gebiete*, vol. 10, pp. 269–278, 1968.

[5] G. J. McLachlan and Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.

[6] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*. Providence, RI: American Math. Society, 2000.

[7] R. K. Boel, M. R. James and I. R. Petersen, "Robustess and risk-sensitivity filtering", *IEEE Trans. Automat. Control*, vol. 47, pp. 451–461, March 2002.

[8] B. C. Levy and R. Nikoukhah, "Robust least-squares estimation with a relative entropy constraint", *IEEE Trans. Information theory*, vol. 50, pp. 89–104, Jan. 2004.

[9] L. P. Hansen and T. J. Sargent, *Robustness*. Prinnceton, NJ: Princeton University Press, 2008.

[10] A. L. McKellips and S. Verdu, "Worst case additive noise for binary-input channels and zero-threshold detection under constraints of power and divergence," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1256–1264, July 1997.

[11] A. L. McKellips and S. Verdu, "Maximin performance of binary-input channels with uncertain noise distributions", *IEEE Trans. Inform. Theory*, vol. 44, pp. 947–972, May 1998.

[12] A. G. Dabak and D. H. Johnson, "Geometrically based robust detection", in *Proc. Conf. Information Sciences and Systems*, Baltimore, MD, The Johns Hopkins Univ, March 1993, pp. 73–77.

[13] J.-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*. New York: J. Wiley, 1984.

[14] D. P. Bertsekas and A. Nedic and A. E. Ozdaglar, *Convex Analysis and Optimization*. Belmont, MA: Athena Scientific, 2003.

[15] N. N Cenkov, *Statistical Decision Rules and Optimal Inference* (Translations of Mathrematical Monographs, vol. 53), Providence, RI: American Math. Society, 1980