ROBUST HYPOTHESIS TESTING WITH SQUARED HELLINGER DISTANCE

Gökhan Gül and Abdelhak M. Zoubir

Signal Processing Group, Technische Universitat Darmstadt
Merckstrasse 25, Darmstadt 64283, Germany.
{ggul,zoubir}@spg.tu-darmstadt.de

ABSTRACT

We extend an earlier work of the same authors, which proposes a minimax robust hypothesis testing strategy between two composite hypotheses based on a squared Hellinger distance. We show that without any further restrictions the former four non-linear equations in four parameters, that have to be solved to design the robust test, can be reduced to two equations in two parameters. Additionally, we show that the same equations can be combined into a single equation if the nominal probability density functions satisfy the symmetry condition. The parameters controlling the degree of robustness are bounded from above depending on the nominal distributions and shown to be determined via solving a polynomial equation of degree two. Experiments justify the benefits of the proposed contributions.

Index Terms— Detection, hypothesis testing, robustness

1. INTRODUCTION

In many practical applications we are concerned with detecting the presence or absence of an event. For example in radar, radio signals are transmitted to detect (if any) aircrafts, ships, motor vehicles or guided missiles. Unless the disturbance effects, which might corrupt the received signal are fully known or can be determined a-priori, for which detection theory would be useless, we consider signals of random nature, which are modeled by statistical hypotheses. To build such a model, the main assumption is that the signals under each hypothesis follow a certain probability distribution with possibly some unknown parameters that are to be estimated. However, such an assumption is rather optimistic and usually there are some secondary physical effects that go under modeled [1].

In such cases, a more realistic approach is to model the probability distributions to belong to a class of distributions $\mathcal{F}$ based on some distance measure D. It is assumed that every probability distribution which is at least ϵ close to the nominal distribution G w.r.t. D is an element of $\mathcal{F}$. The choice of the unknown parameter ϵ depends on the degree of contamination and it varies with the application. In this model, we have as many classes as hypotheses, e.g., for binary hypotheses we have two classes.

The ultimate task of the designer is to determine a single probability distribution from each class and a decision rule such that some predefined performance measure is met. A well known performance measure is the bounded error probability. In other words, a decision rule is chosen which minimizes the error probability for the worst case probability distributions from each class. In this way, a certain level of detection can always be maintained. This type of optimization is called minimax detection and the corresponding worst case distributions are called least favorable distributions (LFD)s.

For some applications like pattern recognition, image or speech classification, we are interested in a statistical test that performs well on average. However, for safety oriented applications such as forest fire detection, seismology, radar or sonar, one is interested in maximizing the worst case performance since the consequences of an incorrect decision might be severe.

In this research field, one of the earliest works goes back to Huber, who proposed a robust version of the probability ratio test [2]. In his paper, he proved the existence of LFDs for the ϵ-contamination class of distributions and showed that the resulting test was obtained by clipping the nominal likelihood ratios. Later, the same concept was extended to a wider class of distributions that involves the ϵ-contamination class as a special case [3], [4]. The models proposed by Huber are very suitable for modeling the outliers, however, not all uncertainties are due to outliers, some of them result from a modeling mismatch.

In order to deal with the uncertainties due to modeling mismatch, it was Levy who proposed a minimax robust hypothesis testing approach that is based on a relative entropy distance [5]. This approach has two basic drawbacks; first it works only when the probability distributions satisfy the symmetry condition, and second it might be difficult to determine the robustness parameter ϵ since its range is $[0,\infty]$. In [6], we proposed an alternative approach based on a squared Hellinger distance which requires no symmetry assumption. Additionally, the robustness parameter ϵ scales in $[0,1]$, which makes it easier to determine. However, as mentioned in the conclusion of [6], it is quite difficult and time consuming to solve four non-linear equations in four parameters to determine the LFDs and the robust decision rule.

In this paper, we show that the former four non-linear equations in four parameters can be reduced to two non-linear equations in two parameters without loss of generality. If additionally the probability distributions satisfy the symmetry condition, we show that a single equation with a single parameter suffices to determine the LFDs. Finally, we propose a method to find the maximum of the robustness parameters above which a robust test can not be designed.
The organization of this paper is as follows. In the following section, an overview of our previous work is outlined. In Section 3, we show how to reduce the number of equations to be solved to obtain the LFDs. In Section 4, we derive an equation to determine the maximum of the robustness parameters. In Section 5, we provide some simulations and finally in Section 6 we conclude the paper.

2. OVERVIEW OF THE PREVIOUS WORK

In this section, we briefly summarize [6] in order to provide a basis for the follow-up work that will be presented in the next sections. For detailed derivations the reader is referred to [6].

2.1. Problem Definition

Let $(\Omega, \mathscr{A}, P_i)$ be a probability space, f_i be a nominal probability density function and $Y : \Omega \mapsto \mathbb{R}$ be a random variable which follows a continuous probability density function g_i, belonging to the set

$$\mathcal{F}_i = \{g_i : S(g_i, f_i) \leq \epsilon_i\}, \tag{1}$$

based on a squared Hellinger distance

$$S(g_i, f_i) = H^2(g_i, f_i) = \frac{1}{2}\int_{\mathbb{R}} \left(\sqrt{g_i(y)} - \sqrt{f_i(y)}\right)^2 \mathrm{d}y \tag{2}$$

when the hypothesis $\mathcal{H}_i$, $i = 0, 1$, is true. A randomized decision rule $u(y)$ assigns each y to $\mathcal{H}_1$ with probability $\delta(y)$ and to $\mathcal{H}_0$ with probability $1 - \delta(y)$, where $\delta \in \Delta$, with $\Delta = C^0(\mathbb{R}, [0, 1])$, resulting in false alarm probability,

$$P_E^1(\delta, f_0) = \int_{\mathbb{R}} \delta(y) g_0(y) dy \tag{3}$$

miss detection probability,

$$P_E^2(\delta, f_1) = \int_{\mathbb{R}} (1 - \delta(y)) g_1(y) dy \tag{4}$$

and overall error probability,

$$P_E(\delta, f_0, f_1) = P(\mathcal{H}_0) P_E^1(\delta, f_0) + P(\mathcal{H}_1) P_E^2(\delta, f_1). \tag{5}$$

The designer aims at solving the minimax optimization problem

$$\hat{\delta}, (\hat{g}_0, \hat{g}_1) = \arg\min_{\delta \in \Delta} \max_{(g_0, g_1) \in \mathcal{F}_0 \times \mathcal{F}_1} P_E(\delta, g_0, g_1) \tag{6}$$

so that a certain level of performance can be guaranteed.

2.2. Least Favorable Distributions and the Decision Rule

The minimax optimization problem defined by (6) can be solved using two coupled Lagrangians with (positive) parameters λ_0, μ_0 and λ_1, μ_1. Let $L = f_1/f_0$, and $y_l, y_u \in \mathbb{R}$ be the lower and upper thresholds of the robust test which are shown to be given by

$$y_l = L^{-1}\left[\left(\frac{\frac{1}{2} + \frac{\mu_1 - 1}{\lambda_1}}{\frac{1}{2} + \frac{\mu_0}{\lambda_0}}\right)^2\right], y_u = L^{-1}\left[\left(\frac{\frac{1}{2} + \frac{\mu_1}{\lambda_1}}{\frac{1}{2} + \frac{\mu_0 - 1}{\lambda_0}}\right)^2\right] \tag{7}$$

respectively. Furthermore, let the decision regions be $R_1 = \{y : y < y_l\}$, $R_2 = \{y : y_l < y < y_u\}$, and $R_3 = \{y : y > y_u\}$, then, by [6] we have the robust decision rule

$$\hat{\delta}(y) = \begin{cases} 0, & y \in R_1 \\ \frac{2\mu_0\lambda_1\sqrt{L(y)} + \lambda_0\left(2 - 2\mu_1 + \lambda_1(-1 + \sqrt{L(y)})\right)}{2\left(\lambda_0 + \lambda_1\sqrt{L(y)}\right)} & y \in R_2 \\ 1, & y \in R_3 \end{cases} \tag{8}$$

and the least favorable densities,

$$\hat{g}_0(y) = \begin{cases} \frac{1}{4\left(\frac{1}{2} + \frac{\mu_0}{\lambda_0}\right)^2} f_0(y), & y \in R_1 \\ \frac{\left(\lambda_0\sqrt{f_0(y)} + \lambda_1\sqrt{f_1(y)}\right)^2}{(\lambda_0 + \lambda_1 + 2(\mu_0 + \mu_1 - 1))^2}, & y \in R_2 \\ \frac{1}{4\left(\frac{1}{2} + \frac{\mu_0 - 1}{\lambda_0}\right)^2} f_0(y), & y \in R_3 \end{cases} \tag{9}$$

and

$$\hat{g}_1(y) = \begin{cases} \frac{1}{4\left(\frac{1}{2} + \frac{\mu_1 - 1}{\lambda_1}\right)^2} f_1(y), & y \in R_1 \\ \frac{\left(\lambda_0\sqrt{f_0(y)} + \lambda_1\sqrt{f_1(y)}\right)^2}{(\lambda_0 + \lambda_1 + 2(\mu_0 + \mu_1 - 1))^2}, & y \in R_2 \\ \frac{1}{4\left(\frac{1}{2} + \frac{\mu_1}{\lambda_1}\right)^2} f_1(y), & y \in R_3 \end{cases} \cdot \tag{10}$$

in four parameters.

2.3. Non-linear Equations

In order to determine the unknown parameters, one needs to solve the set of four non-linear equations

$$c_1 \int_{-\infty}^{y_l} f_0(y)\mathrm{d}y + \int_{y_l}^{y_u} \Phi(y)\mathrm{d}y + c_2 \int_{y_u}^{\infty} f_0(y)\mathrm{d}y = 1$$

$$c_3 \int_{-\infty}^{y_l} f_1(y)\mathrm{d}y + \int_{y_l}^{y_u} \Phi(y)\mathrm{d}y + c_4 \int_{y_u}^{\infty} f_1(y)\mathrm{d}y = 1$$

$$\sqrt{c_1} \int_{-\infty}^{y_l} f_0(y)\mathrm{d}y + \int_{y_l}^{y_u} \sqrt{\Phi(y) f_0(y)}\mathrm{d}y + \sqrt{c_2} \int_{y_u}^{\infty} f_0(y)\mathrm{d}y = 1 - \epsilon_0$$

$$\sqrt{c_3} \int_{-\infty}^{y_l} f_1(y)\mathrm{d}y + \int_{y_l}^{y_u} \sqrt{\Phi(y) f_1(y)}\mathrm{d}y + \sqrt{c_4} \int_{y_u}^{\infty} f_1(y)\mathrm{d}y = 1 - \epsilon_1 \tag{11}$$

where

$$c_1=\frac{1}{4\left(\frac{1}{2}+\frac{\mu_0}{\lambda_0}\right)^2},\ c_2=\frac{1}{4\left(\frac{1}{2}+\frac{\mu_0-1}{\lambda_0}\right)^2}$$

$$c_3=\frac{1}{4\left(\frac{1}{2}+\frac{\mu_1-1}{\lambda_1}\right)^2},\ c_4=\frac{1}{4\left(\frac{1}{2}+\frac{\mu_1}{\lambda_1}\right)^2}$$

and

$$\Phi(y)=\frac{\left(\lambda_0\sqrt{f_0(y)}+\lambda_1\sqrt{f_1(y)}\right)^2}{\left(\lambda_0+\lambda_1+2\left(\mu_0+\mu_1-1\right)\right)^2},$$

in four unknowns, which is quite challenging. It is therefore desirable to reduce the number of equations as well as the number of parameters, if possible.

3. REDUCING THE SYSTEM OF EQUATIONS

In this section, we show how the set of non-linear equations in (11) can be reduced with and without regarding additional constrains. Note that the first two equations are required to make sure that $\hat{g}_0$ and $\hat{g}_1$ are density functions, i.e., they integrate to one and the other two equations are required to guarantee that $\hat{g}_0\in\mathcal{F}_0$ and $\hat{g}_1\in\mathcal{F}_1$.

3.1. Without any Additional Constraints

In its current form, we have implicitly four unknowns μ_0,μ_1, λ_0,λ_1 in four equations. After some manipulations we intend to obtain only two equations in two parameters, y_l and y_u. First, observe that $L(y_u)=c_2/c_4$ and $L(y_l)=c_1/c_3$. Then, we can express the equations either in terms of c_1/c_3 or c_2/c_4. We choose the former one and divide the first two equations by c_3 and the latter two by $\sqrt{c_3}$. Substituting the first equation in the third and the second equation in the fourth via $1/\sqrt{c_3}$ we get

$$\begin{aligned}&\sqrt{\frac{c_1}{c_3}}\int_{-\infty}^{y_l}f_0(y)\mathrm{d}y+\int_{y_l}^{y_u}\sqrt{\frac{\Phi(y)f_0(y)}{c_3}}\mathrm{d}y+\sqrt{\frac{c_2}{c_3}}\int_{y_u}^{\infty}f_0(y)\mathrm{d}y\\&=(1-\epsilon_0)\sqrt{\frac{c_1}{c_3}\int_{-\infty}^{y_l}f_0(y)\mathrm{d}y+\int_{y_l}^{y_u}\frac{\Phi(y)}{c_3}\mathrm{d}y+\frac{c_2}{c_3}\int_{y_u}^{\infty}f_0(y)\mathrm{d}y}\end{aligned}\tag{12}$$

and

$$\begin{aligned}&\int_{-\infty}^{y_l}f_1(y)\mathrm{d}y+\int_{y_l}^{y_u}\sqrt{\frac{\Phi(y)f_1(y)}{c_3}}\mathrm{d}y+\sqrt{\frac{c_4}{c_3}}\int_{y_u}^{\infty}f_1(y)\mathrm{d}y\\&=(1-\epsilon_1)\sqrt{\int_{-\infty}^{y_l}f_1(y)\mathrm{d}y+\int_{y_l}^{y_u}\frac{\Phi(y)}{c_3}\mathrm{d}y+\frac{c_4}{c_3}\int_{y_u}^{\infty}f_1(y)\mathrm{d}y}\end{aligned}\tag{13}$$

Once again considering the substitution $c_2:=c_4L(y_u)$ and $c_1/c_3:=L(y_l)$, one sees that (12) and (13) are functions of y_l, y_u, $\Phi(y)/c_3$, and c_4/c_3. In the following, we manipulate $\Phi(y)/c_3$ such that it can be written in terms of c_4/c_3, y_l and y_u. We first divide both the nominator and denominator of $\Phi(y)$ by λ_0^2, use the identity $1/\sqrt{c_1}=1+2\mu_0/\lambda_0$ and get

$$\Phi(y)=\frac{\left(\sqrt{f_0(y)}+\frac{\lambda_1}{\lambda_0}\sqrt{f_1(y)}\right)^2}{\left(\frac{1}{\sqrt{c_1}}+\frac{\lambda_1}{\lambda_0}+\frac{2\mu_1}{\lambda_0}-\frac{2}{\lambda_0}\right)^2}\tag{14}$$

In the next step, we multiply both the nominator and denominator by $(\lambda_0/\lambda_1)^2$ and use the identity $1/\sqrt{c_3}=|1+(2\mu_1-2)/\lambda_1|$ and get

$$\Phi(y)=\frac{\left(\lambda_0\sqrt{f_0(y)}+\lambda_1\sqrt{f_1(y)}\right)^2}{\lambda_1^2\left(\frac{\lambda_0}{\lambda_1}\frac{1}{\sqrt{c_1}}\pm\frac{1}{\sqrt{c_3}}\right)^2}.\tag{15}$$

Diving (15) by c_3 leads to

$$\frac{\Phi(y)}{c_3}=\frac{\left(\lambda_0\sqrt{f_0(y)}+\lambda_1\sqrt{f_1(y)}\right)^2}{\left(\lambda_0\sqrt{\frac{1}{L(y_l)}}\pm\lambda_1\right)^2}.\tag{16}$$

Note that we can also write

$$\lambda_0=\frac{2\sqrt{c_1c_2}}{\sqrt{c_2}\mp\sqrt{c_1}},\quad \lambda_1=\frac{2\sqrt{c_3c_4}}{\sqrt{c_3}\mp\sqrt{c_4}}.\tag{17}$$

In (16) and the first equation of (17), the upper sign is chosen if $\mu_1+\lambda_1/2>1$ and for (17) second equation, if $\mu_0+\lambda_0/2>1$. When $\mu_i+\lambda_i/2>1$, $i=0,1$ is true, inserting λ_0 and λ_1 from (17) into (16) and using $c_1:=c_3L(y_l)$ and $c_2:=c_4L(y_u)$ we get

$$\begin{aligned}\frac{\Phi(y)}{c_3}=&\left(\frac{\sqrt{L(y_l)}\left(\sqrt{L(y_u)f_0(y)}-\sqrt{f_1(y)}\right)}{\sqrt{L(y_u)}-\sqrt{L(y_l)}}\right.\\&\left.+\frac{\sqrt{c_4/c_3}\sqrt{L(y_u)}(\sqrt{f_1(y)}-\sqrt{f_0(y)L(y_l)})}{\sqrt{L(y_u)}-\sqrt{L(y_l)}}\right)^2\end{aligned}\tag{18}$$

Equation (18) is only a function of the unknowns y_l, y_u and c_4/c_3 as mentioned before. The choice of the upper and lower signs in equations (16) and (17) can only change the signs in (18), i.e., the equation remains the same except for the sign changes inside it. Inserting (18) into (12) and (13), we eventually have two equations in three parameters, y_l, y_u and c_4/c_3. Both equations are quadratic in $x=\sqrt{c_4/c_3}$ and therefore the solution of the equations, (12) and (13), with respect to x yields two roots for each, i.e., two equations of type $x_1=h_1(y_l,y_u)$ and $x_2=h_2(y_l,y_u)$ for (12) and two equations of type $x_3=h_3(y_l,y_u)$ and $x_4=h_4(y_l,y_u)$ for (13) for some known functions h_i, $i=1,\dots,4$. In the final step one obtains two equations in two parameters with $x_1=x_3$ and $x_2=x_4$ or with $x_1=x_4$ and $x_2=x_3$. The equations are lengthy therefore they will not be reproduced here.

3.2. When the Densities are Symmetric

In this section, we assume that the nominal probability density functions satisfy the symmetry condition $f_0(y) = f_1(-y)$ and the robustness parameters are set to $\epsilon = \epsilon_0 = \epsilon_1$. As a consequence, $L(-y) = 1/L(y)$, $\hat{\delta}(-y) = 1 - \hat{\delta}(y)$ and $\hat{L}(-y) = 1/\hat{L}(y)$ hold [5]. This implies $\mu = \mu_0 = \mu_1$, $\lambda = \lambda_0 = \lambda_1$, $c_1 = c_4$, $c_2 = c_3$ and $y_l = -y_u$. As a result we obtain

$$\Phi(y) = \frac{\left(\sqrt{f_0(y)} + \sqrt{f_1(y)}\right)^2}{\left(\frac{1}{\sqrt{c_3}} + \frac{1}{\sqrt{c_4}}\right)^2}. \tag{19}$$

With these simplifications, there are implicitly two unknowns μ and λ and it suffices to consider two equations from (11), the first and the third or the second and the fourth. We select the second and the fourth equations, and use the substitution $c_3 := c_4 L(y_u)$. Eventually, we obtain

$$\begin{aligned}\frac{1}{c_4} &= L(y_u)\int_{-\infty}^{-y_u} f_1(y)\mathrm{d}y \\ &+ k(y_u)\int_{-y_u}^{y_u}\left(\sqrt{f_0(y)} + \sqrt{f_1(y)}\right)^2 \mathrm{d}y + \int_{y_u}^{\infty} f_1(y)\mathrm{d}y\end{aligned} \tag{20}$$

and

$$\begin{aligned}\frac{1-\epsilon}{\sqrt{c_4}} &= L(y_u)\int_{-\infty}^{-y_u} f_1(y)\mathrm{d}y \\ &+ \sqrt{k(y_u)}\int_{-y_u}^{y_u}\sqrt{f_0(y)f_1(y)} + f_1(y)\mathrm{d}y + \int_{y_u}^{\infty} f_1(y)\mathrm{d}y\end{aligned} \tag{21}$$

with $k(y_u) = L(y_u)/(\sqrt{L(y_u)}+1)^2$. Squaring both sides of (21) and inserting (20) into the outcome through c_4, we get

$$\begin{aligned}&L(y_u)\int_{-\infty}^{-y_u} f_1(y)\mathrm{d}y + k(y_u)\int_{-y_u}^{y_u}\left(\sqrt{f_0(y)} + \sqrt{f_1(y)}\right)^2 \mathrm{d}y \\ &+ \int_{y_u}^{\infty} f_1(y)\mathrm{d}y - \frac{1}{(1-\epsilon)^2}\left(\sqrt{L(y_u)}\int_{-\infty}^{-y_u} f_1(y)\mathrm{d}y + \sqrt{k(y_u)}\cdot\right. \\ &\left.\int_{-y_u}^{y_u} f_1(y) + \sqrt{f_0(y)f_1(y)}\mathrm{d}y + \int_{y_u}^{\infty} f_1(y)\mathrm{d}y\right)^2 = 0\end{aligned} \tag{22}$$

4. MAXIMUM ROBUSTNESS PARAMETERS

In this section, we derive the maximum of the robustness parameters above which a robust test can not be designed. This is equivalent to saying that the LFDs from each class become identical.

4.1. General Case

First, we observe that the LFDs, (9) and (10), can totally overlap if R_1 and R_3 are some empty sets. This is achieved when $y_l \to -\infty$ and $y_u \to \infty$. Another possibility is that R_1 and/or R_3 are non-empty sets and f_0 and f_1 are some scaled versions of each other in R_1 and/or R_3. In this case, y_l and y_u will be finite. Referring to (11), this corresponds to $c_1 f_0(y) = c_3 f_1(y)$ and $c_2 f_0(y) = c_4 f_1(y)$ for all $y \in R_1$ and/or $y \in R_1$. As a result the first two equations in (11) will be the same and the latter two will differ only in (ϵ_0, ϵ_1) and the integrals over (y_l, y_u). Therefore, we will not loose much generality by considering R_1 and R_3 to be empty sets.
When $y_l \to -\infty$ and $y_u \to \infty$, the first two equations in (11) reduce to

$$\int_{-\infty}^{\infty} \frac{\left(\lambda_0\sqrt{f_0(y)} + \lambda_1\sqrt{f_1(y)}\right)^2}{(\lambda_0 + \lambda_1 + 2(\mu_0 + \mu_1 - 1))^2} = 1 \tag{23}$$

and the latter two equations reduce to

$$\int_{-\infty}^{\infty}\left|\frac{\lambda_0 f_0(y) + \lambda_1\sqrt{f_0(y)f_1(y)}}{\lambda_0 + \lambda_1 + 2(\mu_0 + \mu_1 - 1)}\right| = 1 - \epsilon_0 \tag{24}$$

and

$$\int_{-\infty}^{\infty}\left|\frac{\lambda_0\sqrt{f_0(y)f_1(y)} + \lambda_1 f_1(y)}{\lambda_0 + \lambda_1 + 2(\mu_0 + \mu_1 - 1)}\right| = 1 - \epsilon_1. \tag{25}$$

The nominal likelihood ratio $L = f_1/f_0$ is assumed to be monotone both in [6] and [5] due to mathematical tractability. Without loss of generality, we further assume that $\lim_{y\to\infty} L(y) = \infty$ and $\lim_{y\to-\infty} L(y) = 0$, which is true for many pairs of nominal distributions on $\mathbb{R}$, such as mean shifted Gaussian distributions. If this assumption is not true one needs to solve (7) w.r.t. the limits on L. With this assumption, from (7) we obtain

$$\mu_0 = \frac{-\lambda_0}{2} + 1, \qquad \mu_1 = \frac{-\lambda_1}{2} + 1. \tag{26}$$

Inserting (26) into (23), (24), and (25) we get

$$\lambda_0^2 + \lambda_1^2 + 2\lambda_0\lambda_1\int_{-\infty}^{\infty}\sqrt{f_0(y)f_1(y)}\mathrm{d}y = 4, \tag{27}$$

$$\lambda_0 + \lambda_1\int_{-\infty}^{\infty}\sqrt{f_0(y)f_1(y)}\mathrm{d}y = 2(1-\epsilon_0), \tag{28}$$

and

$$\lambda_1 + \lambda_0\int_{-\infty}^{\infty}\sqrt{f_0(y)f_1(y)}\mathrm{d}y = 2(1-\epsilon_1), \tag{29}$$

respectively. Let $a = \int_{-\infty}^{\infty}\sqrt{f_0(y)f_1(y)}\mathrm{d}y$, $x_0 = (1-\epsilon_0)$ and $x_1 = (1-\epsilon_1)$. Then, if we substitute (28) into (29) through λ_0, we obtain

$$\lambda_1 = \frac{2x_1 - 2x_0 a}{1 - a^2}. \tag{30}$$

In a similar way, we substitute (28) into (27) and get

$$\lambda_1^2 - \lambda_0^2 + 4\lambda_0 x_0 = 4 \tag{31}$$

and substitute (29) into (27) and get

$$\lambda_0^2 - \lambda_1^2 + 4\lambda_1 x_1 = 4. \tag{32}$$

Summing (31) with (32) we obtain

$$\lambda_0 = \frac{2 - \lambda_1 x_1}{x_0}. \tag{33}$$

In the final step, we first substitute (33) into (32) through λ_0 and then substitute (30) to the outcome through λ_1. Using some algebra and noting that a is finite, this leads to

$$\left(1+a^2\left(-1+{x_0}^2\right)-{x_1}^2\right)\left(1-a^2-{x_0}^2+2ax_0x_1-{x_1}^2\right)=0. \tag{34}$$

In fact, $0 < a \leq 1$ from the integral of a positive function and Cauchy-Schwarz inequality. According to (34), whenever the designer decides for a certain parameter, say ϵ_0, the corresponding maximum ϵ_1 can be determined via solving two polynomial equations of order two. Having a closer look at (34) suggests that the first two roots w.r.t. the first (left) multiplicand and one other root w.r.t. the second (right) multiplicand cannot be valid candidates. Because for all $\epsilon_0 = \epsilon_1 \in [0,1]$, the roots are $a = \pm 1$. Hence, one can obtain the valid root as

$$a = (1-\epsilon_1)(1-\epsilon_0) - \sqrt{(-2+\epsilon_0)\epsilon_0(-2+\epsilon_1)\epsilon_1}, \tag{35}$$

which is symmetric in ϵ_0 and ϵ_1 as expected.

4.2. Equal Parameters ($\epsilon_0 = \epsilon_1$)

If the robustness parameters are equal, we achieve the maximum robustness with

$$\epsilon_{\max} = 1 - \sqrt{\frac{1+a}{2}}. \tag{36}$$

This can either be obtained by letting $\epsilon_0 = \epsilon_1$ in (35) or directly from (22) using the limiting conditions. Notice that over all $a \in [0,1]$, $\epsilon_{\max}$ is bounded by $1-\sqrt{2}/2$. For $\epsilon_{\max} = 0.104$, we get $a \approx 0.606$. This result is in agreement with the asymptotic of the squared Hellinger distance in [6, Fig. 3.].

5. SIMULATIONS

The contributions w.r.t. Section 3 are theoretical. Basically, without any further constraints, one needs to solve four non-linear equations in four parameters. This has a complexity of order $\mathcal{O}(k^4)$, where k indicates the running time. We reduced this complexity to $\mathcal{O}(k^2)$. When the symmetry condition is satisfied and the robustness parameters are equal, $\epsilon_0 = \epsilon_1$, we reduced the complexity from $\mathcal{O}(k^2)$ to $\mathcal{O}(k)$. In the reduced search space, as expected, the equations are more complicated but the benefits are obvious.

With respect to the contributions in Section 4, we simulate (35). More in details, for all a in $(0,1]$ we determine all possible pairs of (ϵ_0, ϵ_1) for which LFDs are the same. i.e., the hypothesis overlap. Figure 1 illustrates a 3D plot of this experiment. One can see that the non-zero values of a map to a rotated, cropped cone like surface on the $(\epsilon_0, \epsilon_1, a)$ space. Except for the intersection curve of the cone-like surface with the $a = 0$ plane, the rest of the points $(\epsilon_0, \epsilon_1, a = 0)$ are undefined, meaning that such points do not exist, i.e., for those points a minimax robust test cannot be designed.

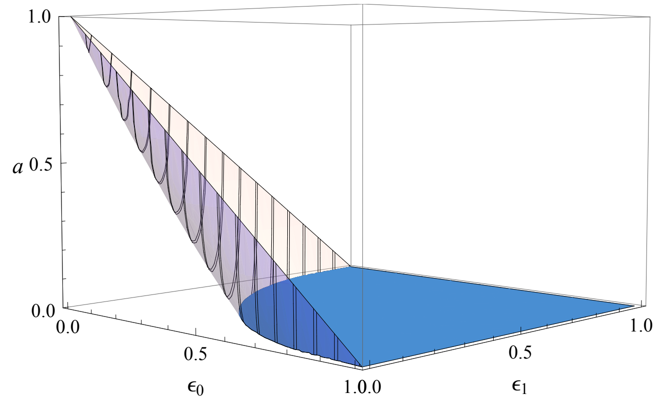

Fig. 1. All allowable pairs of maximum robustness parameters, (ϵ_0, ϵ_1), w.r.t. all distances $a \in (0,1]$.

6. CONCLUSIONS

In this paper, we have extended a previous work of the same authors in two folds. First, we have reduced the number of non-linear equations to be solved for the design of robust tests from four to two in the general case. We have shown that the equations can further be reduced to a single equation if the symmetry assumption between the nominal density functions hold. Second, we have derived the maximum achievable robustness parameters for the general case as well as when the robustness parameters are the same. In the simulations, we have noted that with the proposed formulations the computational complexity was reduced from $\mathcal{O}(k^4)$ to $\mathcal{O}(k^2)$ in the general case and from $\mathcal{O}(k^2)$ to $\mathcal{O}(k)$ for the special case. We have also shown that the maximum robustness parameters map to a cone like shape.

7. ACKNOWLEDGMENT

This work was supported by the LOEWE Priority Program Cocoon (http://www.cocoon.tu-darmstadt.de).

REFERENCES

[1] Bernard C. Levy, *Principles of Signal Detection and Parameter Estimation*, Springer Publishing Company, Incorporated, 1 edition, 2008.

[2] P. J. Huber, “A robust version of the probability ratio test,” *Ann. Math. Statist.*, vol. 36, pp. 1753–1758, 1965.

[3] P. J. Huber and V. Strassen, “Robust confidence limits,” *Z. Wahrcheinlichkeitstheorie verw. Gebiete*, vol. 10, pp. 269278, 1968.

[4] P. J. Huber and V. Strassen, “Minimax tests and the Neyman-Pearson lemma for capacities,” *Ann. Statistics*, vol. 1, pp. 251–263, 1973.

[5] Bernard C. Levy, “Robust hypothesis testing with a relative entropy tolerance.,” *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 413–421, 2009.

[6] G. Gül and A. M. Zoubir, “Robust hypothesis testing for modeling errors,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 5514 – 5518.