MEASURE TRANSFORMED QUASI LIKELIHOOD RATIO TEST

Koby Todros

Ben-Gurion University of the Negev

Alfred O. Hero

University of Michigan

ABSTRACT

In this paper, a generalization of the Gaussian quasi likelihood ratio test (GQLRT) for simple hypotheses is developed. The proposed generalization, called measure-transformed GQLRT (MT-GQLRT), applies GQLRT after transformation of the probability measure of the data. By judicious choice of the transform we show that, unlike the GQLRT, the proposed test can gain sensitivity to higher-order statistical information and resilience to outliers leading to significant mitigation of the model mismatch effect on the decision performance. Under some mild regularity conditions we show that the MT-GQLRT is consistent and its corresponding test statistic is asymptotically normal. A data driven procedure for optimal selection of the measure transformation parameters is developed that maximizes an empirical estimate of the asymptotic power given a fixed empirical asymptotic size. The MT-GQLRT is applied to signal classification in a simulation example that illustrates its sensitivity to higher-order statistical information and resilience to outliers.

***Index Terms*—** Higher-order statistics, hypothesis testing, probability measure transform, robust classification, signal classification.

1. INTRODUCTION

Classical simple binary hypothesis testing deals with t~~he decision problem between~~ two hypotheses based on a sequence of multivariate samples from an underlying probability distribution that ~~belongs to a pair set of~~ probability measures~~ with corresponding singleton parameter spaces~~ [1]. When the probability distributions under each hypothesis are correctly specified the likelihood ratio test (LRT), which is the most powerful test for a given ~~size~~ [2], can be implemented ~~that utilizes complete statistical information.~~ In many practical scenarios the probability distributions are unknown, and therefore, one must resort to~~ other~~ suboptimal tests~~ that require partial statistical information~~.

A popular~~ test of this kind~~ is the Gaussian quasi LRT (GQLRT) [3]-[8] that assumes Gaussian distributions under each hypothesis. The GQLRT operates by selecting the Gaussian probability model that best fits the data. When the observations are i.i.d. this selection is carried out by comparing the empirical Kullback-Leibler divergences [9] between the underlying probability distribution and the assumed normal probability measures. The GQLRT has gained popularity due to its implementation simplicity, ~~performance analysis traceability,~~ and its ~~insightful~~ geometrical interpretations. Despite the model mismatch, introduced by the normality assumption, the GQLRT has the appealing property of consistency when the mean vectors and covariance matrices are correctly specified and identifiable over the considered hypotheses [6]. However, in some circumstances, such as for certain types of non-Gaussian data, large deviation from normality can inflict poor decision performance. This can occur when the first and second-order statistical moments are weakly identifiable over the considered hypotheses, or in the case of heavy-tailed data when the non-robust sample mean and covariance provide poor estimates in the presence of outliers.

In this paper, a generalization of the GQLRT is proposed that applies GQLRT after ~~transformation of the probability distribution of the data.~~ Under the proposed generalization new tests can be obtained that ~~can gain sensitivity to~~ higher-order statistical information~~, resilience~~ to outliers, and yet have the computational and implementation advantages of the GQLRT. This generalization, called measure-transformed GQLRT (MT-GQLRT), is inspired by the measure transformation approach that was recently applied to canonical correlation analysis [10], [11], independent component analysis (ICA) [12], multiple signal classification (MUSIC) [13], [14] and parameter estimation [15].

The proposed transform is structured by a non-negative function, called the MT-function, and maps the probability distribution into a set of new probability measures on the observation space. By modifying the MT-function, classes of measure transformations can be obtained that have different useful properties. Under the proposed transform we define the measure-transformed (MT) mean vector and covariance matrix, derive their strongly consistent estimates, and study their sensitivity to higher-order statistical information and resilience to outliers.

Similarly to the GQLRT, the proposed MT-GQLRT compares the empirical Kullback-Leibler divergences between the transformed ~~probability, distribution~~ of the data and two normal probability measures that are characterized by the MT-mean vector and MT-covariance matrix under each hypothesis. Under some mild regularity conditions we show that the proposed test is consistent and its corresponding test statistic is asymptotically normal. Furthermore, given two training sequences from the probability distribution under each hypothesis, a data-driven procedure for optimal selection of the MT-function within some parametric class of functions is developed that maximizes an empirical estimate of the asymptotic power given a fixed empirical asymptotic size.

We illustrate the MT-GQLRT for the problem of signal classification in the presence of heavy-tailed spherically contoured noise [16] that produces outliers. By specifying the MT-function within the family of zero-centered Gaussian functions parameterized by a scale parameter, we show that the MT-GQLRT outperforms the non-robust GQLRT and attains classification performance that are significantly closer to those obtained by the LRT that, unlike the MT-GQLRT, requires complete knowledge of the likelihood function under each hypothesis.

The paper is organized as follows. In Section 2, ~~a~~ transformation on the probability distribution of the data is ~~developed~~. In Section 3, we use ~~this~~ transformation to construct the MT-GQLRT. The proposed test is applied to a signal classification problem in Section 4. In Section 5, the main points of this contribution are summarized. Proofs for the propositions and corollaries stated throughout the paper will be provided in the full length journal version.

2. PROBABILITY MEASURE TRANSFORM

In this section, we develop a transform on the probability measure of a random vector. Under the proposed transform, we define the measure-transformed mean vector and covariance matrix, derive their strongly consistent estimates, and establish their sensitivity to higher-order statistical information and resilience to outliers. These quantities will be used in the following section to construct the proposed measure-transformed GQLRT.

2.1. Probability measure transformation

We define the measure space $(\mathcal{X}, \mathcal{S}_{\mathcal{X}}, P_{\mathbf{X};\boldsymbol{\theta}})$, where $\mathcal{X} \subseteq \mathbb{C}^p$ is the observation space of a random vector $\mathbf{X}$, $\mathcal{S}_{\mathcal{X}}$ is a σ-algebra over $\mathcal{X}$ and $P_{\mathbf{X};\boldsymbol{\theta}}$ is an unknown probability measure on $\mathcal{S}_{\mathcal{X}}$ parameterized by a vector parameter $\boldsymbol{\theta}$ that belongs to a ~~pair~~ set $\boldsymbol{\Theta} \triangleq \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1\}$.

Definition 1. *Given a non-negative function $u : \mathbb{C}^p \to \mathbb{R}_+$ satisfying*

$$0 < \mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] < \infty, \tag{1}$$

where $\mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] \triangleq \int_{\mathcal{X}} u\left(\mathbf{x}\right) dP_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right)$ and $\mathbf{x} \in \mathcal{X}$, a transform on $P_{\mathbf{X};\boldsymbol{\theta}}$ is defined via the relation:

$$Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\left(A\right) \triangleq \mathrm{T}_u\left[P_{\mathbf{X};\boldsymbol{\theta}}\right]\left(A\right) = \int_A \varphi_u\left(\mathbf{x};\boldsymbol{\theta}\right) dP_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right), \tag{2}$$

where $A \in \mathcal{S}_{\mathcal{X}}$ and $\varphi_u\left(\mathbf{x};\boldsymbol{\theta}\right) \triangleq u\left(\mathbf{x}\right)/\mathrm{E}\left[u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right]$. The function $u\left(\cdot\right)$ is called the MT-function.

Proposition 1 (Properties of the transform)**.** *Let $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ be defined by relation (2). Then 1) $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$. 2) $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ is absolutely continuous w.r.t. $P_{\mathbf{X};\boldsymbol{\theta}}$, with Radon-Nikodym derivative [17]:*

$$dQ^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right)/dP_{\mathbf{X};\boldsymbol{\theta}}\left(\mathbf{x}\right) = \varphi_u\left(\mathbf{x};\boldsymbol{\theta}\right). \tag{3}$$

The probability measure $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ is said to be generated by the MT-function $u\left(\cdot\right)$. By modifying $u\left(\cdot\right)$, such that the condition (1) is satisfied, virtually any probability measure on $\mathcal{S}_{\mathcal{X}}$ can be obtained.

2.2. The MT-mean and MT-covariance

According to (3) the mean vector and covariance matrix of $\mathbf{X}$ under $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ are given by:

$$\boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}} \triangleq \mathrm{E}\left[\mathbf{X}\varphi_u\left(\mathbf{X};\boldsymbol{\theta}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] \tag{4}$$

and

$$\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}} \triangleq \mathrm{E}\left[\mathbf{X}\mathbf{X}^H\varphi_u\left(\mathbf{X};\boldsymbol{\theta}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] - \boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\boldsymbol{\mu}^{(u)H}_{\mathbf{X};\boldsymbol{\theta}}, \tag{5}$$

respectively. Equations (4) and (5) imply that $\boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ are weighted mean and covariance of $\mathbf{X}$ under $P_{\mathbf{X};\boldsymbol{\theta}}$, with the weighting function $\varphi_u\left(\cdot;\cdot\right)$ defined below (2). Hence, they can be estimated using only samples from the distribution $P_{\mathbf{X};\boldsymbol{\theta}}$. By modifying the MT-function $u\left(\cdot\right)$, such that the condition (1) is satisfied, the MT-mean and MT-covariance under $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ are modified. In particular, by choosing $u\left(\cdot\right)$ to be any non-zero constant valued function we have $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}} = P_{\mathbf{X};\boldsymbol{\theta}}$, for which the standard mean vector $\boldsymbol{\mu}_{\mathbf{X};\boldsymbol{\theta}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X};\boldsymbol{\theta}}$ are obtained.

Given a sequence of N i.i.d. samples from $P_{\mathbf{X};\boldsymbol{\theta}}$ the estimators of $\boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ are defined as:

$$\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} \triangleq \sum_{n=1}^{N} \mathbf{X}_n \hat{\varphi}_u\left(\mathbf{X}_n\right) \tag{6}$$

and

$$\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}} \triangleq \sum_{n=1}^{N} \mathbf{X}_n\mathbf{X}_n^H \hat{\varphi}_u\left(\mathbf{X}_n\right) - \hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}}\hat{\boldsymbol{\mu}}^{(u)H}_{\mathbf{X}}, \tag{7}$$

respectively, where $\hat{\varphi}_u\left(\mathbf{X}_n\right) \triangleq u\left(\mathbf{X}_n\right)/\sum_{n=1}^{N} u\left(\mathbf{X}_n\right)$. According to Proposition 2 in [13], if $\mathrm{E}\left[\|\mathbf{X}\|_2^2\, u\left(\mathbf{X}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] < \infty$ then $\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} \xrightarrow[N\to\infty]{\text{w.p. 1}} \boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}} \xrightarrow[N\to\infty]{\text{w.p. 1}} \boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$, where "$\xrightarrow{\text{w.p. 1}}$" denotes convergence with probability (w.p.) 1 [18].

2.3. Robustness to outliers

Robustness of the empirical MT-covariance (7) to outliers was studied in [13] using its influence function [19] which describes the effect on the estimator of an infinitesimal contamination at some point $\mathbf{y} \in \mathbb{C}^p$. An estimator is said to be B-robust if its influence function is bounded [19]. In [13] we have shown that if the MT-function $u(\mathbf{y})$ and the product $u(\mathbf{y})\|\mathbf{y}\|_2^2$ are bounded over $\mathbb{C}^p$ then the influence function of the empirical MT-covariance is bounded. Similarly, it can be shown that under the same conditions the influence function of the empirical MT-mean (6) is bounded.

2.4. Sensitivity to higher-order statistical information

Notice that for any non-constant analytic MT-function $u\left(\cdot\right)$, which has a convergent Taylor series expansion, the MT-mean (4) and the MT-covariance (5) involve higher-order statistical moments of $P_{\mathbf{X};\boldsymbol{\theta}}$. In particular, by choosing $u\left(\mathbf{x};\boldsymbol{t}\right) \triangleq \exp\left(\mathrm{Re}\left\{\boldsymbol{t}^H\mathbf{x}\right\}\right)$, $\boldsymbol{t} \in \mathbb{C}^p$, the resulting exponential MT-mean and MT-covariance are the gradient and Hessian of the cumulant generating function (up to some scaling factors) that have been used for parameter estimation, ICA and channel identification in [20]-[27]. Moreover, by choosing $u\left(\mathbf{x};\boldsymbol{t},\tau\right) \triangleq \exp\left(-\left\|\mathbf{x}-\boldsymbol{t}\right\|^2/\tau^2\right)$, $\tau \in \mathbb{R}_{++}$, we obtain the Gaussian MT-mean and MT-covariance that have been used for non-linear correlation analysis, ICA, robust MUSIC and parameter estimation in [10]-[15].

3. MEASURE-TRANSFORMED GAUSSIAN QUASI LIKELIHOOD RATIO TEST

In this section we use the measure transformation (2) to construct a test between the null and alternative simple hypotheses $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$ based on a sequence of samples $\mathbf{X}_n$, $n = 1, \ldots, N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$. Regularity conditions for asymptotic normality of the corrsponding test statistic are derived. When these conditions are satisfied we show that the proposed test is consistent and derive its asymptotic size and power. Optimal selection of the MT-function $u\left(\cdot\right)$ out of some parametric class of functions is also discussed.

3.1. The MT-GQLRT

Let $\Phi^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ denote a complex circular Gaussian probability distribution [28] that is characterized by the MT-mean $\boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and MT-covariance $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$. The proposed MT-GQLRT compares the empirical Kullback-Leibler divergence (KLD) between $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and $\Phi^{(u)}_{\mathbf{X};\boldsymbol{\theta}_0}$ to the empirical KLD between $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and $\Phi^{(u)}_{\mathbf{X};\boldsymbol{\theta}_1}$. The KLD between $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and $\Phi^{(u)}_{\mathbf{X};\boldsymbol{\theta}_k}$, $k \in \{0,1\}$ is defined as [9]:

$$D_{\mathrm{KL}}\left[Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}||\Phi^{(u)}_{\mathbf{X};\boldsymbol{\theta}_k}\right] \triangleq \mathrm{E}\left[\log\frac{q^{(u)}\left(\mathbf{X};\boldsymbol{\theta}\right)}{\phi^{(u)}\left(\mathbf{X};\boldsymbol{\theta}_k\right)}; Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\right], \tag{8}$$

where $q^{(u)}\left(\mathbf{x};\boldsymbol{\theta}\right)$ and $\phi^{(u)}\left(\mathbf{x};\boldsymbol{\theta}_k\right)$ are the density functions of $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ and $\Phi^{(u)}_{\mathbf{X};\boldsymbol{\theta}_k}$, respectively. According to (3), $D_{\mathrm{KL}}\left[Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}||\Phi^{(u)}_{\mathbf{X};\boldsymbol{\theta}_k}\right]$ can be estimated using only samples from $P_{\mathbf{X};\boldsymbol{\theta}}$. Hence, similarly to (6) and (7), an empirical estimate of (8) given a sequence of samples $\mathbf{X}_n$, $n = 1, \ldots, N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$ is defined as:

$$\hat{D}_{\mathrm{KL}}\left[Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}||\Phi^{(u)}_{\mathbf{X};\boldsymbol{\theta}_k}\right] \triangleq \sum_{n=1}^{N} \hat{\varphi}_u\left(\mathbf{X}_n\right) \log \frac{q^{(u)}\left(\mathbf{X}_n;\boldsymbol{\theta}\right)}{\phi^{(u)}\left(\mathbf{X}_n;\boldsymbol{\theta}_k\right)}, \qquad (9)$$

where $\hat{\varphi}_u(\cdot)$ is defined below (7). Hence, ~~the proposed~~ test statistic ~~is given by~~

$$\begin{aligned} T_u &\triangleq \hat{D}_{\mathrm{KL}}\left[Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}||\Phi^{(u)}_{\mathbf{X};\boldsymbol{\theta}_0}\right] - \hat{D}_{\mathrm{KL}}\left[Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}||\Phi^{(u)}_{\mathbf{X};\boldsymbol{\theta}_1}\right] \qquad (10) \\ &= \sum_{n=1}^{N} \hat{\varphi}_u\left(\mathbf{X}_n\right)\psi_u\left(\mathbf{X}_n;\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right) \\ &= \left(D_{\mathrm{LD}}\left[\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}}||\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_0}\right] + \left\|\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} - \boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_0}\right\|^2_{\left(\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_0}\right)^{-1}}\right) \\ &- \left(D_{\mathrm{LD}}\left[\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}}||\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_1}\right] + \left\|\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}} - \boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_1}\right\|^2_{\left(\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_1}\right)^{-1}}\right), \end{aligned}$$

where

$$\psi_u\left(\mathbf{X};\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right) \triangleq \log \frac{\phi^{(u)}\left(\mathbf{X}_n;\boldsymbol{\theta}_1\right)}{\phi^{(u)}\left(\mathbf{X}_n;\boldsymbol{\theta}_0\right)},$$

$D_{\mathrm{LD}}\left[\mathbf{A}||\mathbf{B}\right] \triangleq \mathrm{tr}\left[\mathbf{A}\mathbf{B}^{-1}\right] - \log\det\left[\mathbf{A}\mathbf{B}^{-1}\right] - p$ is the log-determinant divergence [29] between positive definite matrices $\mathbf{A},\mathbf{B} \in \mathbb{C}^{p\times p}$ and $\|\mathbf{a}\|_{\mathbf{C}} \triangleq \sqrt{\mathbf{a}^H\mathbf{C}\mathbf{a}}$ denotes the weighted Euclidian norm of a vector $\mathbf{a} \in \mathbb{C}^p$ with positive-definite weighting matrix $\mathbf{C} \in \mathbb{C}^{p\times p}$. The decision rule based on the test statistic (10) is given by

$$T_u \underset{H_0}{\overset{H_1}{\gtrless}} t, \qquad (11)$$

where $t \in \mathbb{R}$ denotes a threshold value. Notice that for any non-zero constant MT-function, $u(\cdot)$, $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}} = P_{\mathbf{X};\boldsymbol{\theta}}$ and the standard GQLRT is obtained which only involves first and second-order moments.

3.2. Asymptotic performance analysis

Here, we study the asymptotic performance of the proposed test (11). For simplicity, we assume a sequence of i.i.d. samples $\mathbf{X}_n$, $n = 1, \ldots, N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$.

Proposition 2 (Asymptotic normality)**.** *Assume that the following conditions are satisfied: 1)* $\boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_0} \neq \boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_1}$ *or* $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_0} \neq \boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_1}$. *2)* $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_0}$ *and* $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}_1}$ *are non-singular. 3)* $\mathrm{E}\left[u^2\left(\mathbf{X}\right);P_{\mathbf{X};\boldsymbol{\theta}}\right]$ *and* $\mathrm{E}\left[\|\mathbf{X}\|_2^4 u^2\left(\mathbf{X}\right);P_{\mathbf{X};\boldsymbol{\theta}}\right]$ *are finite for* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ *and* $\boldsymbol{\theta} = \boldsymbol{\theta}_1$. *Then,*

$$T_u \xrightarrow[N\to\infty]{D} \mathcal{N}\left(\eta^{(u)}_{\boldsymbol{\theta}},\lambda^{(u)}_{\boldsymbol{\theta}}\right) \quad \forall\boldsymbol{\theta} \in \boldsymbol{\Theta},$$

where "$\xrightarrow{D}$*" denotes convergence in distribution [18], the mean* $\eta^{(u)}_{\boldsymbol{\theta}} \triangleq \mathrm{E}\left[\varphi_u\left(\mathbf{X};\boldsymbol{\theta}\right)\psi_u\left(\mathbf{X};\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right);P_{\mathbf{X};\boldsymbol{\theta}}\right]$, *and the variance* $\lambda^{(u)}_{\boldsymbol{\theta}} \triangleq N^{-1}\mathrm{Var}\left[\varphi_u\left(\mathbf{X};\boldsymbol{\theta}\right)\psi_u\left(\mathbf{X};\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right);P_{\mathbf{X};\boldsymbol{\theta}}\right]$.

Corollary 1 (Asymptotic size and power)**.** *Assume that the conditions stated in Proposition 2 are satisfied. The asymptotic size and power of the test (11) are given by:*

$$\alpha_u \triangleq Q\left(\frac{t-\eta^{(u)}_{\boldsymbol{\theta}_0}}{\sqrt{\lambda^{(u)}_{\boldsymbol{\theta}_0}}}\right) \quad \text{and} \quad \beta_u \triangleq Q\left(\frac{t-\eta^{(u)}_{\boldsymbol{\theta}_1}}{\sqrt{\lambda^{(u)}_{\boldsymbol{\theta}_1}}}\right), \qquad (12)$$

respectively, where $Q(\cdot)$ *denotes the tail probability of the standard normal distribution.*

Corollary 2 (Consistency)**.** *Assume that the conditions in Proposition 2 are satisfied. Then, for any fixed asymptotic size the asymptotic power of the test (11) satisfies* $\beta_u \to 1$ *as* $N \to \infty$.

In the following Proposition, strongly consistent estimates of the asymptotic size and power (12) ~~are~~ constructed based two i.i.d. sequences from $P_{\mathbf{X};\boldsymbol{\theta}_0}$ and $P_{\mathbf{X};\boldsymbol{\theta}_1}$. These quantities will be used in the sequel for optimal selection of the MT-function.

Proposition 3 (Empirical asymptotic size and power)**.** *Let* $\mathbf{X}^{(k)}_n$, $n = 1, \ldots, N_k$, $k = 0, 1$ *denote sequences of i.i.d. samples from* $P_{\mathbf{X};\boldsymbol{\theta}_0}$ *and* $P_{\mathbf{X};\boldsymbol{\theta}_1}$, *respectively. Define the empirical asymptotic size and power:*

$$\hat{\alpha}_u \triangleq Q\left(\frac{t-\hat{\eta}^{(u)}_{\boldsymbol{\theta}_0}}{\sqrt{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_0}}}\right) \quad \text{and} \quad \hat{\beta}_u \triangleq Q\left(\frac{t-\hat{\eta}^{(u)}_{\boldsymbol{\theta}_1}}{\sqrt{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_1}}}\right), \qquad (13)$$

respectively, where $\hat{\eta}^{(u)}_{\boldsymbol{\theta}_k} \triangleq \sum_{n=1}^{N_k}\hat{\varphi}_u\left(\mathbf{X}^{(k)}_n\right)\psi_u\left(\mathbf{X}^{(k)}_n;\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right)$ *and* $\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_k} \triangleq \frac{N_k}{N}\sum_{n=1}^{N_k}\hat{\varphi}^2_u\left(\mathbf{X}^{(k)}_n\right)\psi^2_u\left(\mathbf{X}^{(k)}_n;\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right) - \frac{1}{N}\left(\hat{\eta}^{(u)}_{\boldsymbol{\theta}_k}\right)^2$. *Assume that* $\mathrm{E}\left[u^2\left(\mathbf{X}\right);P_{\mathbf{X};\boldsymbol{\theta}}\right]$ *and* $\mathrm{E}\left[\|\mathbf{X}\|^4 u^2\left(\mathbf{X}\right);P_{\mathbf{X};\boldsymbol{\theta}}\right]$ *are finite for any* $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. *Then,* $\hat{\alpha}_u \xrightarrow[N_0\to\infty]{w.p.1} \alpha_u$ *and* $\hat{\beta}_u \xrightarrow[N_1\to\infty]{w.p.1} \beta_u$.

3.3. Optimal selection of the MT-function

We propose to specify the MT-function within some parametric family $\{u(\mathbf{X};\boldsymbol{\omega}), \boldsymbol{\omega} \in \boldsymbol{\Omega} \subseteq \mathbb{C}^r\}$ that satisfies the conditions stated in Definition 1 and Proposition 2. An optimal choice of the MT-function parameter $\boldsymbol{\omega}$ would be ~~this~~ that maximizes the empirical asymptotic power (13) at a fixed empirical asymptotic size $\hat{\alpha}_u = \alpha$, i.e., we maximize the following objective function:

$$\hat{\beta}^{(\alpha)}_u(\boldsymbol{\omega}) = Q\left(\frac{\hat{\eta}^{(u)}_{\boldsymbol{\theta}_0}(\boldsymbol{\omega}) - \hat{\eta}^{(u)}_{\boldsymbol{\theta}_1}(\boldsymbol{\omega}) + \sqrt{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_0}(\boldsymbol{\omega})}Q^{-1}(\alpha)}{\sqrt{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_1}(\boldsymbol{\omega})}}\right). \qquad (14)$$

4. EXAMPLE

We consider a signal classification problem that is stated as the problem of testing the following simple hypotheses:

$$\begin{aligned} H_0 &: \quad \mathbf{X}_n = \mathbf{S}_n\left(\boldsymbol{\theta}_0\right) + \mathbf{W}_n, \quad n = 1, \ldots, N, \qquad (15) \\ H_1 &: \quad \mathbf{X}_n = \mathbf{S}_n\left(\boldsymbol{\theta}_1\right) + \mathbf{W}_n, \quad n = 1, \ldots, N, \end{aligned}$$

where $\mathbf{X}_n \in \mathbb{C}^p$ is an observation vector, $\mathbf{S}_n(\boldsymbol{\theta}) \triangleq S_n\boldsymbol{\theta}$ is a latent random vector, $S_n \in \mathbb{C}$ is a first-order stationary random signal that is symmetrically distributed about the origin, $\boldsymbol{\theta} \in \mathbb{C}^p$ is a deterministic unit norm vector and $\mathbf{W}_n \in \mathbb{C}^p$ is a first-order stationary additive noise that is statistically independent of S_n. We assume that the noise component is spherically contoured with stochastic representation [16]:

$$\mathbf{W}_n = \nu_n\mathbf{Z}_n, \qquad (16)$$

where $\nu_n \in \mathbb{R}_{++}$ is a first-order stationary process and $\mathbf{Z}_n \in \mathbb{C}^p$ is a proper-complex wide-sense stationary Gaussian process with zero-mean and scaled unit covariance $\sigma^2_{\mathbf{Z}}\mathbf{I}$. The processes ν_n and $\mathbf{Z}_n$ are assumed to be statistically independent.

In order to gain robustness against outliers, as well as sensitivity to higher-order moments, we specify the MT-function in the zero-centred Gaussian family of functions parametrized by a width parameter ω, i.e.,

$$u\left(\mathbf{x};\omega\right)=\exp\left(-\|\mathbf{x}\|^{2}/\omega^{2}\right),\quad \omega\in\mathbb{R}_{++}. \tag{17}$$

Notice that the MT-function (17) satisfies the B-robustness conditions stated in Subsection 2.3. Using (4), (5) and (15)-(17) it can be shown that the MT-mean and MT-covariance under $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ are:

$$\boldsymbol{\mu}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\left(\omega\right)=\mathbf{0} \tag{18}$$

and

$$\boldsymbol{\Sigma}^{(u)}_{\mathbf{X};\boldsymbol{\theta}}\left(\omega\right)=r_{S}\left(\omega\right)\boldsymbol{\theta}\boldsymbol{\theta}^{H}+r_{W}\left(\omega\right)\mathbf{I}, \tag{19}$$

respectively, where $r_S\left(\omega\right)$ and $r_W\left(\omega\right)$ are some strictly positive functions of ω. Hence, by substituting (17)-(19) into (10) followed by normalization by the observation-independent factor $c\left(\omega\right)\triangleq\frac{r_S(\omega)}{r_W(\omega)(r_S(\omega)+r_W(\omega))}$, the MT-GQLRT (11) simplifies to

$$T'_u\triangleq T_u/c(\omega)=\boldsymbol{\theta}_1^H\hat{\mathbf{C}}^{(u)}_{\mathbf{X}}\left(\omega\right)\boldsymbol{\theta}_1-\boldsymbol{\theta}_0^H\hat{\mathbf{C}}^{(u)}_{\mathbf{X}}\left(\omega\right)\boldsymbol{\theta}_0\underset{H_0}{\overset{H_1}{\gtrless}}t',$$

where $\hat{\mathbf{C}}^{(u)}_{\mathbf{X}}\left(\omega\right)\triangleq\hat{\boldsymbol{\Sigma}}^{(u)}_{\mathbf{X}}\left(\omega\right)+\hat{\boldsymbol{\mu}}^{(u)}_{\mathbf{X}}\left(\omega\right)\hat{\boldsymbol{\mu}}^{(u)H}_{\mathbf{X}}\left(\omega\right)$ and $t'\triangleq t/c(\omega)$.

Under the considered settings, it can be shown that the conditions stated in Proposition 2 are satisfied. The resulting asymptotic power (12) at a given asymptotic size $\alpha_u=\alpha$ takes the form:

$$\beta^{(\alpha)}_{u}\left(\omega\right)=Q\left(-\sqrt{2N\left(\frac{G(\omega)}{1-|\boldsymbol{\theta}_0^H\boldsymbol{\theta}_1|^2}-\frac{1}{2}\right)^{-1}}+Q^{-1}\left(\alpha\right)\right), \tag{20}$$

where $G\left(\omega\right)\triangleq\frac{\mathrm{E}\left[g\left(S,\sqrt{2}\bar{\nu},\omega\right)h\left(\sqrt{2}S,\sqrt{2}\bar{\nu},\omega\right);P_{S,\nu}\right]}{\mathrm{E}^2\left[|S|^2h(S,\bar{\nu},\omega);P_{S,\nu}\right]}$, $\bar{\nu}\triangleq\nu\sigma_{\mathbf{Z}}$, $g\left(S,\nu,\omega\right)\triangleq\left(\frac{\omega^2|S|^2}{\omega^2+\nu^2}\right)^2+3\nu^2\left(\frac{\omega^2|S|^2}{\omega^2+\nu^2}\right)+\nu^4$, and $h\left(S,\nu,\omega\right)\triangleq\left(\frac{\omega^2}{\nu^2+\omega^2}\right)^{p+2}\exp\left(-\frac{|S|^2}{\nu^2+\omega^2}\right)$. Furthermore, its empirical estimate (14) is given by

$$\hat{\beta}^{(\alpha)}_{u}\left(\omega\right)=Q\left(\frac{\tilde{\eta}^{(u)}_{\boldsymbol{\theta}_0}\left(\omega\right)-\tilde{\eta}^{(u)}_{\boldsymbol{\theta}_1}\left(\omega\right)+\sqrt{\tilde{\lambda}^{(u)}_{\boldsymbol{\theta}_0}\left(\omega\right)}Q^{-1}\left(\alpha\right)}{\sqrt{\tilde{\lambda}^{(u)}_{\boldsymbol{\theta}_1}\left(\omega\right)}}\right), \tag{21}$$

where $\tilde{\eta}^{(u)}_{\boldsymbol{\theta}_k}(\omega)\triangleq\frac{\hat{\eta}^{(u)}_{\boldsymbol{\theta}_k}(\omega)}{c(\omega)}=\sum_{n=1}^{N_k}\hat{\varphi}_u\left(\mathbf{X}^{(k)}_n;\omega\right)\xi\left(\mathbf{X}^{(k)}_n;\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right)$, $\tilde{\lambda}^{(u)}_{\boldsymbol{\theta}_k}(\omega)\triangleq\frac{\hat{\lambda}^{(u)}_{\boldsymbol{\theta}_k}(\omega)}{c^2(\omega)}=\frac{N_k}{N}\sum_{n=1}^{N_k}\hat{\varphi}^2_u\left(\mathbf{X}^{(k)}_n;\omega\right)\xi^2\left(\mathbf{X}^{(k)}_n;\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right)-\frac{1}{N}\left(\tilde{\eta}^{(u)}_{\boldsymbol{\theta}_k}\right)^2$, $k=0,1$, and $\xi\left(\mathbf{X};\boldsymbol{\theta}_0,\boldsymbol{\theta}_1\right)\triangleq\left|\boldsymbol{\theta}_1^H\mathbf{X}\right|^2-\left|\boldsymbol{\theta}_0^H\mathbf{X}\right|^2$.

In the following simulation examples, the vectors $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ were set to $\boldsymbol{\theta}_k\triangleq\frac{1}{\sqrt{p}}\left[1,e^{-i\pi\sin(\vartheta_k)},\ldots,e^{-i\pi(p-1)\sin(\vartheta_k)}\right]^T$, $k=0,1$, were $\vartheta_0=0$, $\vartheta_1=\pi/3$ and $p=4$. We consider a BPSK signal with variance σ^2_S and an ϵ-contaminated Gaussian noise model [16] under which the texture component ν in (16) is a binary random variable satisfying $\nu=1$ w.p. $1-\epsilon$ and $\nu=\delta$ w.p. ϵ. The parameters ϵ and δ that control the heaviness of the noise tails were set to 0.2 and 10, respectively. We define the signal-to-noise-ratio (SNR) as $\mathrm{SNR}\triangleq10\log_{10}\sigma^2_S/\left[\sigma^2_{\mathbf{Z}}\left((1-\epsilon)+\epsilon\delta^2\right)\right]$. In all simulation examples the sample size was set to $N=10^3$. The empirical asymptotic power (21) was obtained using two i.i.d. training sequences from $P_{\mathbf{X};\boldsymbol{\theta}_0}$ and $P_{\mathbf{X};\boldsymbol{\theta}_1}$ containing $N_0=N_1=10^5$ samples.

In the first example, we compared the asymptotic power (20) to its empirical estimate (21) at test size equal to 0.05 as a function of ω for $\mathrm{SNR}=-10$ [dB]. Observing Fig. 1, one sees that due to the consistency of (21) the compared quantities are very close.

In the second example, we compared the empirical, asymptotic (20) and empirical asymptotic (21) power of the MT-GQLRT to the empirical powers of the GQLRT and the LRT for test size equal to 0.05. The optimal Gaussian MT-function parameter $\omega_{\rm opt}$ was obtained by maximizing (21) over $\Omega=[1,100]$. The empirical power curves were obtained using 10^4 Monte-Carlo simulations. The SNR is used to index the performances as depicted in Fig. 2. One sees that the MT-GQLRT outperforms the non-robust GQLRT and for most examined SNR values performs similarly to the LRT that, unlike the MT-GQLRT, requires complete knowledge of the likelihood function under each hypothesis.

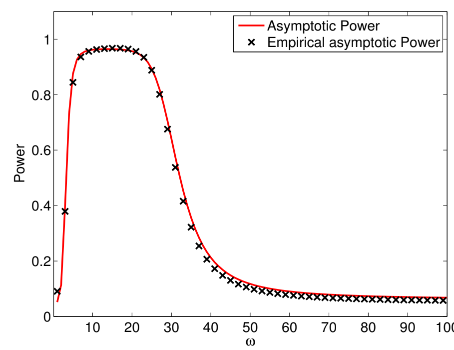

Fig. 1. Asymptotic power (20) and its empirical estimate (21) at test size $\alpha=0.05$ versus the width parameter ω of the MT-function (17).

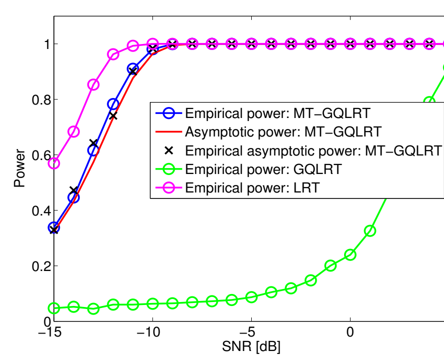

Fig. 2. The empirical, asymptotic (20) and empirical asymptotic (21) powers of the MT-GQLRT as compared to the empirical powers of the GQLRT and LRT at test size $\alpha=0.05$.

5. CONCLUSION

In this paper a new test, called MT-GQLRT, was ~~developed~~ that applies Gaussian LRT after transformation of the probability distribution of the data. By specifying the MT-function in the Gaussian family, the proposed test was applied to signal classification in non-Gaussian noise. Exploration of other MT-functions may result in additional tests in this class that have different useful properties.

6. REFERENCES

[1] L. L. Scharf, *Statistical signal processing,* Addison-Wesley, 1991.

[2] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences,* vol. 231, pp. 289-337, 1933.

[3] W. K. Härdle, V. Spokoiny, V. Panov, and W. Wang, *Basics of Modern Mathematical Statistics,* Springer, 2013.

[4] M. Taniguchi and Y. Kakizawa, *Asymptotic theory of statistical inference for time series,* Springer, 2012.

[5] J. Bai, "Likelihood ratio tests for multiple structural changes," *Journal of Econometrics,* vol. 91, no. 2, pp. 299-323, 1999.

[6] K. Sakiyama and M. Taniguchi, "Discriminant analysis for locally stationary processes," *Journal of Multivariate Analysis,* vol. 90, no. 2, pp. 282-300, 2004.

[7] I. A. Canay, "EL inference for partially identified models: large deviations optimality and bootstrap validity," *Journal of Econometrics,* vol. 156, no. 2, pp. 408-425, 2010.

[8] M. C. Beaulieu, J. M. Dufour and L. Khalaf, "Finite sample multivariate tests of asset pricing models with coskewness," *Computational Statistics & Data Analysis,* vol. 53, no. 6, pp. 2008-2021, 2009.

[9] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics,* vol. 22, pp. 79-86, 1951.

[10] K. Todros and A. O. Hero, "On measure-transformed canonical correlation analysis," *IEEE Transactions on Signal Processing,* vol. 60, no. 9, pp. 4570-4585, 2012.

[11] K. Todros and A. O. Hero, "Measure transformed canonical correlation analysis with application to financial data," *Proceedings of SAM 2012,* pp. 361-364, 2012.

[12] K. Todros and A. O. Hero, "Measure independent component analysis," available online as arXiv:1302.0730.

[13] K. Todros and A. O. Hero, "Robust Multiple Signal Classification via Probability Measure Transformation," *IEEE Transactions on Signal Processing,* vol. 63, no. 5, 2015.

[14] K. Todros and A. O. Hero, "Robust Measure transformed MUSIC for DOA estimation," *Proceeding of ICASSP 2014,* pp. 4190-4194, 2014.

[15] K. Todros and A. O. Hero, "Measure-transformed quasi maximum likelihood estimation with application to source localization," *Proceedings of ICASSP 2015,* pp. 3462 - 3466.

[16] E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor, "Complex elliptically symmetric distributions: survey, new results and applications," *IEEE Transactions on Signal Processing,* vol. 60, no. 1, pp. 5597-5625, 2012.

[17] G. B. Folland, *Real Analysis*. John Wiley and Sons, 1984.

[18] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory,* Springer-Verlag, 2006.

[19] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel, *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011.

[20] A. Slapak and A. Yeredor, "Charrelation and Charm: generic statistics incorporating higher-order information," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5089-5106, 2012.

[21] A. Slapak and A. Yeredor, "Charrelation-based estimation of the parameters of non-Gaussian autoregressive processes," *Proceedings of SSP 2012*, pp. 448-451.

[22] A. Yeredor, "blind source separation via the second characteristic function," *Signal Processing,* vol. 80, pp. 897-902, 2000.

[23] A. Slapak and A. Yeredor, ""Weighting for more": enhancing characteristic-function based ICA with asymptotically optimal weighting," *Signal Processing,* vol. 91, no. 8, pp. 2016-2027, 2011.

[24] A. Yeredor, "Blind channel estimation using first and second derivatives of the characteristic function," *IEEE Signal Processing Letters,* vol. 9, pp. 100-103, 2002.

[25] A. Yeredor, "Substituting the cumulants in the super-exponential blind equalization algorithm," *Proceedings of the ICASSP 2008*.

[26] E. Edinger and A. Yeredor, "Blind MIMO identification using the second characteristic function," *IEEE Transactions on Signal Processing,* vol. 53, pp. 4067-4079, 2005.

[27] A. Yeredor, "MUSIC using off-origin Hessians of the second characteristic function," *Proceedings of SAM 2006*, pp. 176-180.

[28] P. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: the theory of improper and noncircular signals.* p. 39, Cambridge University Press, 2010.

[29] I. S. Dhillon and J. A. Tropp, "Matrix nearness problems with Bregman divergences," *SIAM Journal on Matrix Analysis and Applications,* vol. 29, no. 4, pp. 1120-1146, 2007.