Measure-Transformed Gaussian Quasi Score Test

Koby Todros
Ben-Gurion University of the Negev

Abstract—**In this paper, we develop a robust generalization of the Gaussian quasi score test (GQST) for composite binary hypothesis testing. The proposed test, called measure-transformed GQST (MT-GQST), is based on a transformation applied to the probability distribution of the data. The considered transform is structured by a non-negative function, called MT-function, that weights the data points. By appropriate selection of the MT-function we show that, unlike the GQST, the proposed MT-GQST incorporates higher-order moments and can gain robustness to outliers. The MT-GQST is applied for testing the parameter of a non-linear model. Simulation example illustrates its advantages as compared to the standard GQST and other robust detectors.**

I. Introduction

The score test [1], also known as Rao's score test, or the Lagrangian multiplier test, is a well established technique for composite binary hypothesis testing [1]. Its test-statistic is based on the score-function, defined as the gradient of the log-likelihood function w.r.t. the vector parameter. Unlike the generalized likelihood ratio test (GLRT) and Wald's test [1], it does not necessitate the maximum likelihood estimate under the alternative hypothesis, and therefore, may be significantly easier to compute. However, similarly to Wald's test and the GLRT, it requires knowledge of the likelihood function. In many practical scenarios the likelihood function is unknown, and thus, alternatives to the score test that require only partial statistical information become attractive.

One alternative that utilizes only first and second-order statistical moments is called the Gaussian quasi score test (GQST) [2]-[5], which assumes normally distributed observations. The GQST, that belongs to the class of M-tests [6], [7], is obtained by replacing the score-function with a Gaussian quasi score-function (GQSF). The GQSF is defined as the gradient (w.r.t. the vector parameter) of a Gaussian log-likelihood function that is characterized by the parametric mean vector and covariance matrix of the underlying distribution. The GQST has simple implementation and tractable performance analysis that arise from the convenient Gaussian model. Furthermore, the GQST is consistent under some mild regularity conditions [3]. However, the GQST is not resilient against large deviations from normality that can occur, e.g., in the case of heavy-tailed noise that generates outliers.

In this paper, we develop a robust generalization of the GQST, called measure-transformed GQST (MT-GQST), that is based on the parametric probability measure-transform [8]. The considered measure-transformation is structured by a non-negative function, called MT-function, that weights the data points. The MT-GQST is obtained by replacing the GQSF with a measure-transformed GQSF multiplied by the MT-function. The measure-transformed GQSF is defined as the gradient of a Gaussian log-likelihood function that is characterized by the measure-transformed mean vector and covariance matrix. By appropriate selection of the MT-function we show that, unlike the GQST, the proposed MT-GQST incorporates higher-order statistical moments and can gain robustness to outliers, while maintaining the implementation advantages of the GQST.

In the paper we show that the MT-GQST is consistent under some mild assumptions. We also show that the asymptotic distribution of the test-statistic is central chi-squared under the null hypothesis, and non-central chi-squared under a sequence of local alternatives, with non-centrality parameter that is increasing with the inverse asymptotic error-covariance of the MT-GQMLE [8]. We analyze the robustness of the test to outliers via the second-order influence function [9] of the test statistic. Selection of the MT-function, within some parametric family, is carried out by minimizing the spectral norm of the empirical asymptotic error-covariance of the MT-GQMLE. We show that this minimization amounts to maximization of an empirical worst-case asymptotic local power at a fixed asymptotic size.

The MT-GQST is applied for testing the vector parameter of a non-linear data model in the presence of spherically invariant noise [10]. The MT-function is selected within the class of spherical Gaussian functions that are centered about the origin and parameterized by a scale parameter. We show that the MT-GQST performs similarly to the GQST for normally distributed noise. When the noise obeys a heavy-tailed K-distribution [10], we show that the MT-GQST outperforms the non-robust GQST and other robust detectors, and significantly reduces the performance gap towards the score test that, unlike the MT-GQST, necessitates complete information of the parametric distribution.

Lastly, we emphasize that unlike the works in [11] and [12] that deal with *simple* binary hypothesis testing, this work deals with *composite* binary hypothesis testing.

The paper is organized as follows. In Sections II, the considered probability measure-transform is reviewed. In Section III, the proposed MT-GQST is derived. A numerical example illustrating the advantages of the MT-GQST is given in Section IV. Section V provides concluding remarks. Proofs for the theorem and propositions stated in the manuscript will be provided in the full journal version.

II. Probability measure transform: Review

This section provides a brief review of the parametric probability measure transform [8]. Based on this transformation, we redefine the parametric measure-transformed mean vector and covariance matrix and show their relation to higher-order

ISBN 978-0-9928626-7-1

statistical moments. These quantities will be used in Section III to construct the proposed test.

A. Preliminaries

Let $(\mathcal{X}, \mathcal{S}_{\mathcal{X}}, P_{\mathbf{X};\boldsymbol{\theta}})$ denote a measure space, where $\mathcal{X} \subseteq \mathbb{C}^p$ is the observation space of a complex random vector $\mathbf{X}$, $\mathcal{S}_{\mathcal{X}}$ is a σ-algebra over $\mathcal{X}$ and $P_{\mathbf{X};\boldsymbol{\theta}}$ is a probability measure on $\mathcal{S}_{\mathcal{X}}$ parameterized by a vector parameter $\boldsymbol{\theta}$ that belongs to an open parameter space $\boldsymbol{\Theta} \subseteq \mathbb{R}^m$. For any integrable scalar function $h : \mathcal{X} \to \mathbb{C}$ the expectation of $h(\mathbf{X})$ under $P_{\mathbf{X};\boldsymbol{\theta}}$ is given by:

$$\mathrm{E}\left[h(\mathbf{X}); P_{\mathbf{X};\boldsymbol{\theta}}\right] \triangleq \int_{\mathcal{X}} h(\mathbf{x})\, dP_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}). \tag{1}$$

The empirical probability measure $\hat{P}_{\mathbf{X}}$ given a sequence of samples $\mathbf{X}_n$, $n = 1, \ldots, N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$ is defined as:

$$\hat{P}_{\mathbf{X}}(A) = \frac{1}{N} \sum\nolimits_{n=1}^{N} \delta_{\mathbf{X}_n}(A), \tag{2}$$

where $A \in \mathcal{S}_{\mathcal{X}}$, and $\delta_{\mathbf{X}_n}(\cdot)$ is the Dirac probability measure at $\mathbf{X}_n$ [13].

B. Probability measure transformation

Definition 1 (Definition of the transform)**.** *Given a non-negative function $u : \mathbb{C}^p \to \mathbb{R}_+$ such that*

$$0 < \mathrm{E}\left[u(\mathbf{X}); P_{\mathbf{X};\boldsymbol{\theta}}\right] < \infty, \tag{3}$$

a transform on $P_{\mathbf{X};\boldsymbol{\theta}}$ is defined as:

$$Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}(A) \triangleq \mathrm{T}_u\left[P_{\mathbf{X};\boldsymbol{\theta}}\right](A) = \int_A \varphi_u(\mathbf{x}; \boldsymbol{\theta})\, dP_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}), \tag{4}$$

where $A \in \mathcal{S}_{\mathcal{X}}$ and

$$\varphi_u(\mathbf{x}; \boldsymbol{\theta}) \triangleq u(\mathbf{x}) / \mathrm{E}\left[u(\mathbf{X}); P_{\mathbf{X};\boldsymbol{\theta}}\right]. \tag{5}$$

The function $u(\cdot)$ is called the MT-function.

By Definition 1, $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ (4) is a probability measure on $\mathcal{S}_{\mathcal{X}}$ that is absolutely continuous w.r.t. $P_{\mathbf{X};\boldsymbol{\theta}}$, with Radon-Nikodym derivative [13]:

$$dQ^{(u)}_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) / dP_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}) = \varphi_u(\mathbf{x}; \boldsymbol{\theta}). \tag{6}$$

The MT-function $u(\cdot)$ is the generating function of the probability measure $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$. By modifying $u(\cdot)$ a wide range of probability measures on $\mathcal{S}_{\mathcal{X}}$ can be obtained.

C. The measure-transformed mean and covariance

By the Radon-Nikodym derivative (6) the mean vector and the covariance matrix of $\mathbf{X}$ under $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}$ take the forms:

$$\boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta}) \triangleq \mathrm{E}\left[\mathbf{X}\varphi_u(\mathbf{X}; \boldsymbol{\theta}); P_{\mathbf{X};\boldsymbol{\theta}}\right] \tag{7}$$

and

$$\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta}) \triangleq \mathrm{E}\left[\mathbf{X}\mathbf{X}^H \varphi_u(\mathbf{X}; \boldsymbol{\theta}); P_{\mathbf{X};\boldsymbol{\theta}}\right] - \boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta}) \boldsymbol{\mu}^{(u)H}_{\mathbf{X}}(\boldsymbol{\theta}), \tag{8}$$

respectively. The quantities (7) and (8) will be called the MT-mean vector and the MT-covariance matrix, respectively. Notice 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 $\varphi_u(\cdot;\cdot)$ defined in (5). Also notice that when $u(\cdot)$ is non-zero and constant valued $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}} = P_{\mathbf{X};\boldsymbol{\theta}}$ and then (7) and (8) coincide with the standard mean vector $\boldsymbol{\mu}_{\mathbf{X}}(\boldsymbol{\theta})$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{X}}(\boldsymbol{\theta})$, respectively. Alternatively, when $u(\cdot)$ is a non-constant analytic function, that has a convergent power series expansion, the MT-mean and MT-covariance incorporate higher-order statistical moments of $P_{\mathbf{X};\boldsymbol{\theta}}$.

III. Derivation of the test

In this section, the proposed MT-GQST is derived. Under some regularity assumptions, the MT-GQST is shown to be consistent. Furthermore, we derive the asymptotic distribution of its test statistic under the null hypothesis and under a sequence of contiguous local alternatives. Robustness of the MT-GQST to outliers is analyzed via its second-order influence function [9]. Finally, selection of the MT-function is discussed.

A. Problem formulation

Given a sequence of samples $\mathbf{X}_1, \ldots, \mathbf{X}_N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$, we consider the following composite hypothesis testing problem:

$$\begin{aligned} H_0 &: \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0 \\ H_1 &: \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \end{aligned} \tag{9}$$

where $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$. We consider the case where the underlying parametric family $\{P_{\mathbf{X};\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ is unknown. Partial statistical information is available through the MT-mean $\boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta})$ and the MT-covariance $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta})$ that are assumed to be known parametrized functions (up to some redundant constants).

B. The MT-GQST

Define the measure-transformed GQSF:

$$\boldsymbol{\psi}_u(\mathbf{x}; \boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{\theta}} \log \phi(\mathbf{x}; \boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta}), \boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta})), \tag{10}$$

where $\phi(\cdot;\cdot,\cdot)$ is a proper complex Gaussian probability density function and it is assumed that $\boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta})$ are differentiable. One can verify that

$$\mathrm{E}[\boldsymbol{\psi}_u(\mathbf{X}; \boldsymbol{\theta}_0); Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}] = \mathbf{0} \;\text{ for }\; \boldsymbol{\theta} = \boldsymbol{\theta}_0. \tag{11}$$

Therefore, since by (3), (5) and (6) any integrable scalar function $h : \mathcal{X} \to \mathbb{C}$ satisfies $\mathrm{E}[h(\mathbf{X}); Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}}] = 0$ if and only if $\mathrm{E}[u(\mathbf{X})h(\mathbf{X}); P_{\mathbf{X};\boldsymbol{\theta}}] = 0$ we conclude that

$$\boldsymbol{\eta}_u(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \triangleq \mathrm{E}[u(\mathbf{X})\boldsymbol{\psi}_u(\mathbf{X}; \boldsymbol{\theta}_0); P_{\mathbf{X};\boldsymbol{\theta}}] = \mathbf{0} \;\text{ for }\; \boldsymbol{\theta} = \boldsymbol{\theta}_0. \tag{12}$$

Thus, when $\boldsymbol{\eta}_u(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \neq \mathbf{0}$ for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ an empirical estimate of $\boldsymbol{\eta}_u(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ can be used for obtaining a consistent test between H_0 and H_1. Hence, given a sequence of samples $\mathbf{X}_1, \ldots, \mathbf{X}_N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$ the MT-GQST for the hypothesis testing problem (9) is defined as:

$$T_u \triangleq N \cdot \hat{\boldsymbol{\eta}}^T_u(\boldsymbol{\theta}_0)\, \hat{\mathbf{G}}^{-1}_u(\boldsymbol{\theta}_0) \hat{\boldsymbol{\eta}}_u(\boldsymbol{\theta}_0) \underset{H_0}{\overset{H_1}{\gtrless}} t, \tag{13}$$

where

$$\hat{\boldsymbol{\eta}}_u(\boldsymbol{\theta}_0) \triangleq \frac{1}{N} \sum_{n=1}^{N} u(\mathbf{X}_n) \boldsymbol{\psi}_u(\mathbf{X}_n; \boldsymbol{\theta}_0) \tag{14}$$

is an empirical estimate of $\boldsymbol{\eta}_u(\boldsymbol{\theta}_0, \boldsymbol{\theta})$,

$$\hat{\mathbf{G}}_u(\boldsymbol{\theta}_0) \triangleq \frac{1}{N}\sum_{n=1}^{N} u^2\left(\mathbf{X}_n\right)\boldsymbol{\psi}_u\left(\mathbf{X}_n;\boldsymbol{\theta}_0\right)\boldsymbol{\psi}_u^T\left(\mathbf{X}_n;\boldsymbol{\theta}_0\right) \quad (15)$$

is the empirical autocorrelation matrix of $u(\mathbf{X})\boldsymbol{\psi}_u(\mathbf{X};\boldsymbol{\theta}_0)$, which is assumed to be non-singular, and $t \in \mathbb{R}_+$ denotes a threshold. Notice that (13) defines a class of tests over the set of MT-functions that obey Definition 1. In particular, when the MT-function $u(\cdot)$ is any non-zero constant function (for which $Q^{(u)}_{\mathbf{X};\boldsymbol{\theta}} = P_{\mathbf{X};\boldsymbol{\theta}}$) the standard GQST, that incorporates only first and second-order statistical moments, is obtained from (13).

C. Asymptotic performance analysis

Here, we analyze the asymptotic performance of the proposed test (13). For simplicity, we assume that a sequence of i.i.d. samples $\mathbf{X}_n$, $n = 1, \ldots, N$ from $P_{\mathbf{X};\boldsymbol{\theta}}$ is available. We begin by stating some regularity conditions that will be used in the sequel:

(A-1) $\boldsymbol{\eta}_u(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \neq \mathbf{0}$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.
(A-2) $\mathbf{H}_u(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \triangleq \mathrm{E}[u^2(\mathbf{X})\boldsymbol{\psi}_u(\mathbf{X};\boldsymbol{\theta}_0)\boldsymbol{\psi}_u^T(\mathbf{X};\boldsymbol{\theta}_0); P_{\mathbf{X};\boldsymbol{\theta}}]$ is bounded and non-singular over $\boldsymbol{\Theta} \times \boldsymbol{\Theta}$.
(A-3) $\boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta})$ are twice continuously differentiable with bounded first and second-order derivatives.
(A-4) $\mathrm{E}[u^4(\mathbf{X}); P_{\mathbf{X};\boldsymbol{\theta}}]$ and $\mathrm{E}[\|\mathbf{X}\|^8 u^4(\mathbf{X}); P_{\mathbf{X};\boldsymbol{\theta}}]$ are bounded.
(A-5) The density of $P_{\mathbf{X};\boldsymbol{\theta}}$ is differentiable in $\boldsymbol{\Theta}$ a.e. over $\mathcal{X}$.
(A-6) The Fisher information matrix is bounded over $\boldsymbol{\Theta}$.

The following proposition states sufficient conditions for consistency of the MT-GQST.

Proposition 1 (Consistency)**.** *Assume that conditions A-1-A-4 are satisfied. Then, for any threshold $t \in \mathbb{R}_+$*

$$\Pr\left[T_u > t\right] \xrightarrow[N\to\infty]{} 1 \ \textit{ under } H_1. \quad (16)$$

Next, we derive the asymptotic distribution of the test-statistic under the null hypothesis and under a sequence of local alternatives.

Proposition 2 (Asymptotic distribution under the null hypothesis)**.** *Assume that conditions A-3 and A-4 are satisfied. Then,*

$$T_u \xrightarrow[N\to\infty]{D} \chi^2_m \ \textit{ under } H_0, \quad (17)$$

where χ^2_m denotes a central chi-squared distribution with m-degrees of freedom, and "$\xrightarrow{D}$" denotes convergence in distribution [13].

Theorem 1 (Asymptotic distribution under local alternatives)**.** *Assume that conditions A-2–A-6 are satisfied. Furthermore, consider a sequence of local alternatives that converges to $\boldsymbol{\theta}_0$ at a rate of $1/\sqrt{N}$. Specifically, consider*

$$H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 + \mathbf{h}/\sqrt{N}, \quad (18)$$

where $\mathbf{h} \in \mathbb{R}^m$ is a non-zero locality parameter. Then,

$$T_u \xrightarrow[N\to\infty]{D} \chi^2_m\left(\lambda_u(\mathbf{h})\right) \ \textit{ under } H_1, \quad (19)$$

where $\chi^2_m\left(\lambda_u(\mathbf{h})\right)$ is a non-central chi-squared distribution with m-degrees of freedom and non-centrality parameter

$$\lambda_u(\mathbf{h}) \triangleq \mathbf{h}^T\mathbf{R}_u^{-1}(\boldsymbol{\theta}_0)\mathbf{h}. \quad (20)$$

The matrix $\mathbf{R}_u(\boldsymbol{\theta})$ is defined as:

$$\mathbf{R}_u(\boldsymbol{\theta}) \triangleq \mathbf{F}_u^{-1}\left(\boldsymbol{\theta}\right)\mathbf{G}_u\left(\boldsymbol{\theta}\right)\mathbf{F}_u^{-1}\left(\boldsymbol{\theta}\right), \quad (21)$$

where

$$\mathbf{G}_u\left(\boldsymbol{\theta}\right) \triangleq \mathrm{E}[u^2\left(\mathbf{X}\right)\boldsymbol{\psi}_u\left(\mathbf{X};\boldsymbol{\theta}\right)\boldsymbol{\psi}_u^T\left(\mathbf{X};\boldsymbol{\theta}\right); P_{\mathbf{X};\boldsymbol{\theta}}], \quad (22)$$

$$\mathbf{F}_u\left(\boldsymbol{\theta}\right) \triangleq -\mathrm{E}\left[u\left(\mathbf{X}\right)\boldsymbol{\Gamma}_u\left(\mathbf{X};\boldsymbol{\theta}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right], \quad (23)$$

$$\boldsymbol{\Gamma}_u\left(\mathbf{x};\boldsymbol{\theta}\right) \triangleq \nabla^2_{\boldsymbol{\theta}}\log\phi(\mathbf{x};\boldsymbol{\mu}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta}), \boldsymbol{\Sigma}^{(u)}_{\mathbf{X}}(\boldsymbol{\theta})), \quad (24)$$

and $\mathbf{F}_u\left(\boldsymbol{\theta}\right)$ is assumed to be a non-singular matrix function.

We note that the matrix $\mathbf{R}_u(\boldsymbol{\theta})$ (21) is exactly the error-covariance of the MT-GQMLE [11]. The following Corollary is a direct consequence of (19), (20), the Rayleigh-Ritz Theorem [14] and the property that the tail probability of the non-central chi-squared distribution is monotonically increasing in the non-centrality parameter [15].

Corollary 1 (Asymptotic local power)**.** *Assume that the conditions stated in Theorem 1 hold. Under the local alternatives (18), the asymptotic power at a fixed asymptotic size α satisfies*

$$\beta^{(\alpha)}_u\left(\mathbf{h}\right) = Q_{\chi^2_m(\lambda_u(\mathbf{h}))}\left(Q^{-1}_{\chi^2_m}\left(\alpha\right)\right), \quad (25)$$

where $Q_{\chi^2_m}(\cdot)$ and $Q_{\chi^2_m(\cdot)}(\cdot)$ denote the tail probabilities of the central and non-central chi-squared distributions, respectively. Furthermore, for any $c > 0$ the worst-case asymptotic power

$$\bar{\beta}^{(\alpha)}_u(c) \triangleq \min_{\mathbf{h}:\|\mathbf{h}\|\geq c} \beta^{(\alpha)}_u(\mathbf{h}) = Q_{\chi^2_m(\gamma_u(c))}\left(Q^{-1}_{\chi^2_m}\left(\alpha\right)\right), \quad (26)$$

where $\gamma_u(c) \triangleq c^2\left\|\mathbf{R}_u(\boldsymbol{\theta}_0)\right\|^{-1}_S$ and $\|\cdot\|_S$ denotes the spectral norm.

D. Robustness study

Here, we analyze the robustness of the test-statistic (13) to outliers via Hampel's second-order influence function evaluated at the null distribution $P_{\mathbf{X};\boldsymbol{\theta}_0}$ [9]. Using (1) the test-statistic in (13) can be written as:

$$T_u = NS_u[\hat{P}_{\mathbf{X}}], \quad (27)$$

where

$$S_u[\hat{P}_{\mathbf{X}}] \triangleq \mathbf{d}_u^T[\hat{P}_{\mathbf{X}}]\mathbf{J}_u[\hat{P}_{\mathbf{X}}]\mathbf{d}_u[\hat{P}_{\mathbf{X}}]$$

is a statistical functional of the empirical probability distribution (2), with $\mathbf{d}_u[\cdot] \triangleq \mathrm{E}[u\left(\mathbf{X}\right)\boldsymbol{\psi}_u\left(\mathbf{X};\boldsymbol{\theta}_0\right);\cdot]$ and $\mathbf{J}_u[\cdot] \triangleq \mathrm{E}[u^2\left(\mathbf{X}\right)\boldsymbol{\psi}_u\left(\mathbf{X};\boldsymbol{\theta}_0\right)\boldsymbol{\psi}_u^T\left(\mathbf{X};\boldsymbol{\theta}_0\right);\cdot]$. Define the ϵ-contaminated probability measure:

$$P_\epsilon \triangleq (1-\epsilon)P_{\mathbf{X};\boldsymbol{\theta}_0} + \epsilon\delta_{\mathbf{y}},$$

where $0 \leq \epsilon \leq 1$, $\mathbf{y} \in \mathbb{C}^p$, and $\delta_{\mathbf{y}}$ is the Dirac probability measure at $\mathbf{y}$. The n^{th}-order influence function of the test-statistic (27) at $P_{\mathbf{x};\boldsymbol{\theta}_0}$ is defined as:

$$\text{IC}^{(n)}\left(\mathbf{y}; P_{\mathbf{x};\boldsymbol{\theta}_0}\right) \triangleq \left.\frac{\partial^n S_u\left[P_\epsilon\right]}{\partial \epsilon^n}\right|_{\epsilon=0}. \tag{28}$$

This function quantifies the bias effect on the test-statistic introduced by an infinitesimal contamination at some point $\mathbf{y}$. Under the considered settings, it can shown that $\text{IC}^{(1)}\left(\mathbf{y}; P_{\mathbf{x};\boldsymbol{\theta}_0}\right) = 0$. Therefore, we study the second-order influence function:

$$\text{IC}^{(2)}\left(\mathbf{y}; P_{\mathbf{x};\boldsymbol{\theta}_0}\right) = 2\|u\left(\mathbf{y}\right)\boldsymbol{\psi}_u\left(\mathbf{y};\boldsymbol{\theta}_0\right)\|^2_{\mathbf{G}_u^{-1}(\boldsymbol{\theta}_0)}, \tag{29}$$

where $\|\mathbf{a}\|_{\mathbf{C}} \triangleq \sqrt{\mathbf{a}^H\mathbf{C}\mathbf{a}}$ denotes a weighted Euclidian norm. The test-statistic is said to be robust if $\text{IC}^{(2)}\left(\mathbf{y}; P_{\mathbf{x};\boldsymbol{\theta}_0}\right)$ is bounded over $\mathbb{C}^p$. The following proposition states a sufficient condition for boundedness of (29).

Proposition 3. *The second-order influence function (29) is bounded if there exists a positive constant C such that $u\left(\mathbf{y}\right) \leq C$ and $u\left(\mathbf{y}\right)\|\mathbf{y}\|^2 \leq C$ for any $\mathbf{y} \in \mathbb{C}^p$.*

E. Selection of the MT-function

According to Propositions 1 and 2, the asymptotic global power and size are invariant to the choice of the MT-function $u(\cdot)$. However, by Corollary 1 one sees that it controls the asymptotic local power through the error-covariance $\mathbf{R}_u(\boldsymbol{\theta}_0)$ (21). Since the tail probability of the non-central chi-squared distribution is monotonically increasing in the non-centrality parameter [15], minimization of the spectral norm $\|\mathbf{R}_u(\boldsymbol{\theta}_0)\|_S$ amounts to maximization of the worst case asymptotic local power $\bar{\beta}_u^{(\alpha)}(c)$ (26) for any fixed c and asymptotic size α. Hence, we propose to choose $u(\cdot)$ that minimizes $\|\hat{\mathbf{R}}_u(\boldsymbol{\theta}_0)\|_S$, where $\hat{\mathbf{R}}_u(\boldsymbol{\theta})$ is an empirical estimate of error-covariance (21):

$$\hat{\mathbf{R}}_u(\boldsymbol{\theta}) \triangleq \hat{\mathbf{F}}_u^{-1}(\boldsymbol{\theta})\hat{\mathbf{G}}_u(\boldsymbol{\theta})\hat{\mathbf{F}}_u^{-1}(\boldsymbol{\theta}), \tag{30}$$

where $\hat{\mathbf{F}}_u\left(\boldsymbol{\theta}\right) \triangleq -N^{-1}\sum_{n=1}^N u\left(\mathbf{X}_n\right)\boldsymbol{\Gamma}_u\left(\mathbf{X}_n;\boldsymbol{\theta}\right)$ is an estimate of (23) and $\hat{\mathbf{G}}_u\left(\boldsymbol{\theta}\right)$ is defined in (15). It can be shown that if conditions A-3−A-6 are satisfied then $\hat{\mathbf{R}}_u(\boldsymbol{\theta}_0) \xrightarrow[N\to\infty]{P} \mathbf{R}_u(\boldsymbol{\theta}_0)$ under the local alternatives (18), where "$\xrightarrow{P}$" denotes convergence in probability [13].

Here, the class of MT-functions is restricted to some parametric family $\{u\left(\mathbf{X};\boldsymbol{\omega}\right), \boldsymbol{\omega} \in \boldsymbol{\Omega} \subseteq \mathbb{C}^r\}$ that satisfies conditions (3) and A-2−A-6. Hence, the optimal MT-function parameter $\boldsymbol{\omega}_{\text{opt}}$ is the minimizer of the spectral norm $\|\hat{\mathbf{R}}_u(\boldsymbol{\theta}_0;\boldsymbol{\omega})\|_S$ that is constructed from (30) by the same data samples comprising the MT-GQST (13).

IV. Numerical Example

We consider the problem of testing the parameter $\boldsymbol{\theta}$ of the following non-linear data model:

$$\mathbf{X}_n = S_n\mathbf{a}\left(\boldsymbol{\theta}\right) + \mathbf{W}_n, \quad n = 1, \ldots, N, \tag{31}$$

where $\{\mathbf{X}_n \in \mathbb{C}^p\}$ is an observation process, $\{S_n \in \mathbb{C}\}$ is an i.i.d. latent random process with unknown zero-mean symmetric distribution, and $\mathbf{a} : \boldsymbol{\Theta} \to \mathbb{C}^p$ is a known non-linear unit-norm twice differentiable vector function. The process $\{\mathbf{W}_n\}$ is an i.i.d. spherically invariant [10] noise process that is statistically independent of $\{S_n\}$ and satisfies the stochastic representation $\mathbf{W}_n = A_n\mathbf{Z}_n$, where $\{A_n \in \mathbb{R}_{++}\}$ is an i.i.d. process with unknown distribution and $\{\mathbf{Z}_n \in \mathbb{C}^p\}$ is a proper-complex i.i.d. Gaussian process with zero-mean and scaled unit covariance $\sigma^2_{\mathbf{Z}}\mathbf{I}$ with unknown scale parameter. The processes $\{A_n\}$ and $\{\mathbf{Z}_n\}$ are assumed to be independent.

We choose the following spherical Gaussian MT-function with a scale parameter $\omega \in \mathbb{R}_{++}$:

$$u\left(\mathbf{x};\omega\right) = \exp(-\|\mathbf{x}\|^2/\omega^2). \tag{32}$$

Notice that $u\left(\mathbf{x};\omega\right)$ satisfies the robustness condition stated in Proposition 3. By (31) and (32) it can be shown that the MT-mean (7) and the MT-covariance (8) are given by:

$$\boldsymbol{\mu}_{\mathbf{x}}^{(u)}\left(\boldsymbol{\theta}\right) = \mathbf{0} \tag{33}$$

and

$$\boldsymbol{\Sigma}_{\mathbf{x}}^{(u)}\left(\boldsymbol{\theta}\right) = c_u\left(\omega\right)\mathbf{a}(\boldsymbol{\theta})\mathbf{a}^H(\boldsymbol{\theta}) + r_u\left(\omega\right)\mathbf{I}, \tag{34}$$

respectively, where $c_u\left(\omega\right)$ and $r_u\left(\omega\right)$ are some strictly positive functions of ω. Thus, by (10), (24), (33) and (34) the vector and matrix functions $\boldsymbol{\psi}_u\left(\mathbf{x};\boldsymbol{\theta}\right)$ and $\boldsymbol{\Gamma}_u\left(\mathbf{x};\boldsymbol{\theta}\right)$ comprising the test-statistic (13) and the empirical estimate of the error-covariance (30) take the following simple forms:

$$\boldsymbol{\psi}_u\left(\mathbf{x};\boldsymbol{\theta}\right) = d_u\left(\omega\right)\nabla_{\boldsymbol{\theta}}\left|\mathbf{x}^H\mathbf{a}\left(\boldsymbol{\theta}\right)\right|^2 \tag{35}$$

and

$$\boldsymbol{\Gamma}_u\left(\mathbf{x};\boldsymbol{\theta}\right) = d_u\left(\omega\right)\nabla^2_{\boldsymbol{\theta}}\left|\mathbf{x}^H\mathbf{a}\left(\boldsymbol{\theta}\right)\right|^2, \tag{36}$$

respectively, where $d_u\left(\omega\right) \triangleq c_u(\omega)/(r_u(\omega)(c_u(\omega) + r_u(\omega)))$. Notice that the resulting test-statistic (13) and the empirical error-covariance estimate (30) are *independent* of $d_u\left(\omega\right)$.

In the simulation example that follows, we compare the detection performance of the MT-GQST (13) to the standard non-robust GQST, the ZMNL-GQST, a robust score-type M-test [6], [7] based on Cauchy's loss function [17], and to the score test which assumes that the likelihood function is known.

The GQST is obtained from the MT-GQST (13) by setting $u(\cdot) = 1$. The ZMNL-GQST is a robust extension of the GQST that operates by applying GQST after outliers suppression, carried out by clipping the amplitude of the observations using a zero-memory non-linear (ZMNL) function. We use the same ZMNL preprocessing approach that has been applied in [16] to suppress outliers in some detection and estimation problems.

The score-type M-test we compared to is obtained from (13) by setting $u(\cdot) = 1$ and replacing $\boldsymbol{\psi}_u(\mathbf{x};\boldsymbol{\theta})$ with $\boldsymbol{\psi}_{\text{C}}(\mathbf{x};\boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{\theta}}\rho_{\text{C}}(r(\mathbf{x};\boldsymbol{\theta}))$, where $\rho_{\text{C}}(r) \triangleq \log(1 + r)$ is Cauchy's loss function, $r(\mathbf{x};\boldsymbol{\theta}) = \|\mathbf{P}^{\perp}_{\mathbf{a}}(\boldsymbol{\theta})\mathbf{x}\|^2/\hat{\sigma}^2$, and $\mathbf{P}^{\perp}_{\mathbf{a}}(\boldsymbol{\theta})$ is the projection matrix onto the null-space of $\mathbf{a}(\boldsymbol{\theta})$. The scale parameter estimate $\hat{\sigma}$ is defined as $\hat{\sigma} \triangleq (p^{-1}\sum_{k=1}^p \hat{\sigma}^2_{X_k})^{1/2}$, where $\hat{\sigma}^2_{X_k} = c^2[(\text{MAD}(\{\text{Re}(X_{k,n})\}_{n=1}^N))^2 + (\text{MAD}(\{\text{Im}(X_{k,n})\}_{n=1}^N))^2]$ is a robust median absolute deviation (MAD) estimate of variance [7]. The constant $c \triangleq 1/\text{erf}^{-1}(3/4)$ ensures consistency of the variance estimate for normally distributed data [7].

ISBN 978-0-9928626-7-1

We considered a BPSK signal with variance σ_S^2. The non-linear vector function $\mathbf{a}(\boldsymbol{\theta})$ we examined represents a steering vector of $p = 8$ elements uniform linear array with quarter wavelength spacing $d = \lambda/4 = 0.25$ [m] corresponding to a near-field narrowband signal with range r and bearing ϑ. Here, the vector parameter $\boldsymbol{\theta} \triangleq [r, \vartheta]^T$. By Fresnel's approximation [18] when $0.62(d^3(p-1)^3/\lambda)^{1/2} < r < 2d^2(p-1)^2/\lambda$

$$[\mathbf{a}(\boldsymbol{\theta})]_k = \exp(j(\omega_{\rm e}k + \phi_{\rm e}k^2 + O(d^2/r^2))),$$

$k = 0, \ldots, p-1$, where $\omega_{\rm e} \triangleq -2\pi d \sin(\vartheta)/\lambda$ and $\phi_{\rm e} \triangleq \pi d^2 \cos^2(\vartheta)/(\lambda r)$ are called "electrical angles" (in practice, the $O(d^2/r^2)$ term is neglected). Two types of noise distributions were examined: 1) Gaussian and 2) heavy-tailed K-distributed noise [10] with shape parameter $\kappa = 0.75$. The sample size was set to $N = 10^3$. The signal-to-noise-ratio (SNR), used to index the detection performance, is defined as $\text{SNR} \triangleq 10\log_{10}\sigma_S^2/\sigma_{\mathbf{z}}^2$. The vector parameter at the null hypothesis H_0 was set to $\boldsymbol{\theta}_0 = [r_0, \vartheta_0]^T$, where $r_0 = 1.5$ [m] and $\vartheta_0 = 0°$. The test size of all compared tests was fixed to $\alpha = 10^{-2}$. We considered a specific local alternative $\boldsymbol{\theta}_1 = [r_1, \vartheta_1]^T$, corresponding to $\mathbf{h} = \sqrt{N}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$ in (18), where $r_1 = r_0 + 0.01$ [m] and $\vartheta_1 = \vartheta_0 + 0.5°$. The optimal width parameter $\omega_{\rm opt}$ of Gaussian MT-function (32) was obtained by minimizing the spectral norm $\|\hat{\mathbf{R}}_u(\boldsymbol{\theta}_0;\omega)\|_S$ of the empirical error-covariance (30) over $\Omega = [1, 30]$. All empirical power curves were obtained via 10^4 Monte-Carlo trials.

Fig. 1 depicts the empirical and asymptotic (25) power curves of the MT-GQST as compared to the empirical power curves of the GQST, ZMNL-GQST, Cauchy's score-type M-test, and the score test. Notice that when the noise is Gaussian, the MT-GQST, GQST, ZMNL-GQST, and Cauchy's score-type M-test attain similar performance. For the K-distributed noise, the MT-GQST outperforms the GQST, ZMNL-GQST, and Cauchy's score-type M-test, and significantly reduces the gap towards the score test, which unlike the MT-GQST, necessitates complete knowledge of the parametric distribution.

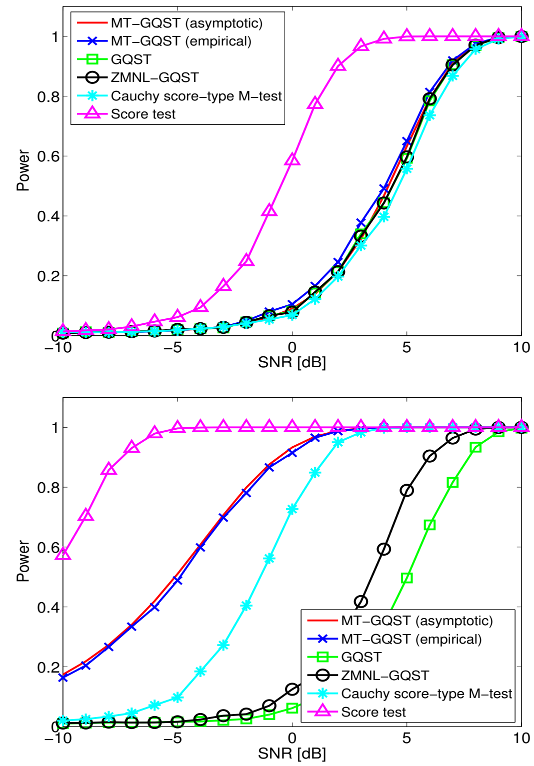

Fig. 1. Testing the parameter of a non-linear model in Gaussian noise **(top)** and K-distributed noise **(bottom)**.

V. Conclusion

In this paper a new robust score-type test was developed based on a transformation of the probability distribution of the data. The proposed test was applied for testing the parameter of a non-linear observation model in Gaussian and non-Gaussian heavy-tailed noise. Simulation study demonstrates significant performance improvement as compared to the non-robust GQST and other robust detectors.

References

[1] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses.* Springer Texts in Statistics, 2005.
[2] H. White, "Maximum likelihood estimation of misspecified models," *Econometrica: Journal of the Econometric Society,* pp. 1-25, 1982.
[3] H. White, *Estimation, inference and specification analysis,* Cambridge university press, 1996.
[4] G. Fiorentini and E. Sentana, "Tests for serial dependence in static, non-Gaussian factor models," *Unobserved components and time series econometrics,* pp. 118-163, Oxford University Press, 2015.
[5] T. Bollerslev and J. M. Wooldridge, "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric reviews,* vol. 11, no. 2, pp. 143-172, 1992.
[6] F. Peracchi, "Robust M-tests," *Econometric Theory,* vol. 7, no. 1, pp. 69-84, 1991.
[7] P. J. Huber, *Robust statistics,* Springer, 2011.
[8] K. Todros and A. O. Hero, "Measure-transformed quasi maximum likelihood estimation," *IEEE Transactions on Signal Processing,* vol. 65, no. 3, pp. 748-763, 2017.
[9] 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.
[10] 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.
[11] K. Todros and A. O. Hero, "Measure-transformed quasi likelihood ratio test," *Proceedings of ICASSP 2016,* pp. 4259-4263.
[12] N. Halay, K. Todros and A. O. Hero, "Measure-transformed quasi likelihood ratio test for Bayesian binary hypothesis testing," *Proceedings of SSP 2016*, pp. 1-5.
[13] K. B. Athreya and S. N. Lahiri, *Measure theory and probability theory,* Springer-Verlag, 2006.
[14] R. Horn and C. R. Johnson, *Matrix analysis.* Cambridge University Press, 2012.
[15] A. Aubel and W. Gawronski, "Analytic properties of non-central distributions," *Applied Mathematics & Computation,* vol. 141, pp. 3-12, 2003.
[16] A. Swami and B. M. Sadler, "On some detection and estimation problems in heavy-tailed noise," *Signal Processing,* vol. 82, no. 12, pp. 1829-1846, 2002.
[17] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics-theory and Methods,* vol. 6, no. 9, pp. 813-827, 1977.
[18] A. L. Swindlehurst and T. Kailath, "Passive direction of arrival and range estimation for near-field sources," *IEEE Spec. Est. and Mod. Workshop,* pp. 123-128, 1988.