Signal Processing 155 (2019) 202–217

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Robust composite binary hypothesis testing via measure-transformed quasi score test

Koby Todros

Ben-Gurion University of the Negev, Israel

ARTICLE INFO

Article history:
Received 12 April 2018
Revised 16 August 2018
Accepted 14 September 2018
Available online 21 September 2018

Keywords:
Composite binary hypothesis testing
Detection theory
Probability measure transform
Robust statistics

ABSTRACT

This paper deals with the problem of composite binary hypothesis testing when an accurate parametric probability model is not available. Under this framework, a robust generalization of the Gaussian quasi score test (GQST) is developed. The proposed generalization, called measure-transformed (MT) GQST assumes a Gaussian probability model after applying a transform to the probability measure (distribution) of the data. The considered measure-transformation is structured by a non-negative data weighting function, called MT-function. By proper selection of the MT-function, we show that, unlike the GQST, the proposed MT-GQST can gain resilience against heavy-tailed noise outliers, leading to significant mitigation of the model mismatch effect (introduced by the normality assumption), and yet, have the implementation advantages of the standard GQST (arising from the convenient Gaussian model). The proposed MT-GQST is applied for testing the vector parameters of linear and nonlinear multivariate data models. Simulation examples illustrate its advantages as compared to the GQST and other robust detectors.

1. Introduction

Composite binary hypothesis testing deals with the problem of testing between two hypotheses that involve unknown parameters [1–3]. This problem is encountered in many statistical signal processing applications, such as signal detection, signal classification, model mismatch detection and spectrum sensing. Rao's score test [1,3,4] is a well established tool for composite binary hypothesis testing, whose test-statistic is based on the score-function defined as the gradient of the log-likelihood function w.r.t. the vector parameter. The main advantage of this test over the generalized likelihood ratio test (GLRT) [1,3] and Wald's test [1,3,5], is that it does not involve maximum likelihood estimation under the alternative hypothesis, and therefore, may have a significantly easier implementation. However, similarly to these tests it assumes complete knowledge of the likelihood function. In many practical scenarios the likelihood function is unknown, or alternatively, does not possess a closed form expression. In these cases, alternatives to the score test, that require only partial statistical information, become highly relevant.

The Gaussian quasi score test (GQST) [6–10] is a popular alternative of this kind that assumes normally distributed observations, and thus, utilizes only first and second-order statistical moments. The GQST belongs to the wide classes of M-tests [11,12] and tests that are based on the generalized method of moments [13,14]. Its test-statistic is obtained by replacing the score-function with a Gaussian quasi score-function (GQSF), 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 popularity of the GQST is attributed to its simple implementation and tractable performance analysis that arise from the convenient Gaussian model. Furthermore, it has an appealing consistency property under some mild regularity conditions [7]. However, despite these advantages, the GQST may be sensitive to large deviations from the normality assumption that can lead to poor decision performance. These deviations can occur, e.g., in the case of non-Gaussian heavy-tailed noise that produces outliers.

A straight-forward approach to overcome this limitation is to apply a non-Gaussian quasi score test (NGQST) [15–19] that assumes a more complex distributional model, e.g., elliptical, at the possible expense of increased implementation complexity, cumbersome performance analysis, and degraded performance under nominal Gaussian data. For example, by assuming Laplace distributed observations the NGQST for DC signal detection in additive i.i.d. noise is the well established sign detector [3,17]. The sign detector is more resilient against heavy-tailed noise outliers as compared to the GQST. However, it has relatively inferior performance when the noise is Gaussian [17,20].

In this paper, we develop a robust generalization of the GQST. The proposed generalization, called measure-transformed (MT)

E-mail address: todros@bgu.ac.il

https://doi.org/10.1016/j.sigpro.2018.09.018

GQST assumes a convenient Gaussian probability model after applying a transform to the probability measure of the data. The considered measure-transformation, also applied in [21–25], is structured by a non-negative function, called MT-function, that weights the data points. In practice, 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 parametric measure-transformed mean vector and covariance matrix. By appropriate selection of the MT-function we show that, unlike the GQST, the proposed MT-GQST 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 regularity conditions. The asymptotic distribution of the test-statistic is shown to be central chi-squared under the null hypothesis, and non-central chi-squared under a sequence of local alternatives. Interestingly, the non-centrality parameter, that controls the asymptotic local power, is increasing with the inverse asymptotic error-covariance of the measure-transformed Gaussian quasi maximum likelihood estimator (MT-GQMLE) [23]. We analyze the robustness of the test to outliers using a generalized version of the second-order influence function [26] of the test-statistic. Unlike the standard influence function, this generalized version involves multiple outliers. Sufficient conditions on the MT-function that guarantee outlier resilience are derived. In the paper we also show that minimization of the spectral norm of the asymptotic error-covariance of the MT-GQMLE amounts to maximization of a worst-case asymptotic local power at a fixed test-size. Hence, 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. An approximate iterative solution for this minimization problem is developed that is based on the steepest-descend method [27].

The MT-GQST is applied for testing the vector parameters of linear and nonlinear multivariate data models in the presence of spherical noise. The MT-function is selected within classes of pseudo-Gaussian and Gaussian shaped 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 is non-Gaussian and heavy-tailed, we show that the MT-GQST outperforms the non-robust GQST and other robust detectors, and significantly reduces the performance gap towards the omniscient score test that, unlike the MT-GQST, requires complete knowledge of the parametric distribution.

The basic idea behind the proposed MT-GQST was first presented in the conference paper [25]. The contribution of the present paper relative to [25] includes: (1) detailed derivation of the MT-GQST, (2) an enhanced outliers contamination model that involves multiple outliers, (3) a steepest-descend based procedure for optimizing the measure-transformation parameters, (4) more complete simulation studies, and (5) rigorous proofs of the propositions and theorem stating the properties of the test.

The paper is organized as follows. In Sections 2, we formulate the considered composite binary hypothesis testing problem and review the GQST. In Section 3, the considered probability measure transformation is reviewed. In Section 4 we develop the proposed MT-GQST. Numerical examples are given in Section 5. Section 6 summarizes the main points of this contribution. Proofs for the theorem and propositions stated in the manuscript are provided in the Appendix.

2. Preliminaries

In this section, we state the considered composite binary hypothesis testing problem. We proceed by reviewing the GQST. We show that the GQST is based on the principle that the Gaussian quasi score-function has a zero expectation under the null hypothesis. This principle will be used in Section 4 to develop the proposed MT-GQST.

2.1. Preliminary definitions and assumptions

We define the measure space $\left(\mathcal{X}, \mathcal{S}_{\mathcal{X}}, P_{\mathbf{X};\boldsymbol{\theta}}\right)$, 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 $\boldsymbol{\Theta} \subseteq \mathbb{R}^m$. It is assumed that $P_{\mathbf{X};\boldsymbol{\theta}}$ is absolutely continuous w.r.t. a dominating σ-finite measure ρ on $\mathcal{S}_{\mathcal{X}}$, such that the Radon-Nykonym derivative [28]

$$f\left(\mathbf{x};\boldsymbol{\theta}\right) \triangleq dP_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x})/d\rho(\mathbf{x}), \tag{1}$$

exists for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The function $f\left(\mathbf{x};\boldsymbol{\theta}\right)$ is called the likelihood function of $\boldsymbol{\theta}$ observed by the vector $\mathbf{x} \in \mathcal{X}$. Let $g : \mathcal{X} \to \mathbb{C}$ denote an integrable scalar function on $\mathcal{X}$. The expectation of $g(\mathbf{X})$ under $P_{\mathbf{X};\boldsymbol{\theta}}$ is defined as

$$\mathrm{E}\left[g(\mathbf{X}); P_{\mathbf{X};\boldsymbol{\theta}}\right] \triangleq \int_{\mathcal{X}} g(\mathbf{x}) dP_{\mathbf{X};\boldsymbol{\theta}}(\mathbf{x}). \tag{2}$$

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 specified by

$$\hat{P}_{\mathbf{X}}(A) = \frac{1}{N}\sum_{n=1}^{N} \delta_{\mathbf{X}_n}(A), \tag{3}$$

where $A \in \mathcal{S}_{\mathcal{X}}$, and $\delta_{\mathbf{X}_n}(\cdot)$ is the Dirac probability measure at $\mathbf{X}_n$ [28].

2.2. Problem statement

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

$$H_0 : \ \boldsymbol{\theta} = \boldsymbol{\theta}_0 \qquad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \tag{4}$$

where $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ is known. We consider the case where the underlying parametric family $\{P_{\mathbf{X};\boldsymbol{\theta}} : \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ is unknown and only partial statistical information is available. The GQST, reviewed in the following subsection, assumes that first and second-order partial statistical information is available through the standard mean vector and the covariance matrix of $P_{\mathbf{X};\boldsymbol{\theta}}$ that are assumed to be known functions of $\boldsymbol{\theta}$. The proposed MT-GQST that will be developed in Section 4 exploits the mean vector and the covariance matrix under a transformed version of $P_{\mathbf{X};\boldsymbol{\theta}}$. Finally, we note that the hypothesis testing problem (4) does not involve nuisance parameters. The case of nuisance parameters will be treated in a followup paper.

2.3. The Gaussian quasi score test

Define the GQSF:

$$\boldsymbol{\psi}(\mathbf{X};\boldsymbol{\theta}) \triangleq \nabla_{\boldsymbol{\theta}} \log\phi(\mathbf{X};\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta})), \tag{5}$$

where $\phi(\cdot;\ \cdot,\ \cdot)$ is a proper complex Gaussian probability density function and it is assumed that $\boldsymbol{\mu}(\boldsymbol{\theta}) \triangleq \mathrm{E}[\mathbf{X}; P_{\mathbf{X};\boldsymbol{\theta}}]$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}) \triangleq \mathrm{E}[\mathbf{X}\mathbf{X}^H; P_{\mathbf{X};\boldsymbol{\theta}}] - \boldsymbol{\mu}(\boldsymbol{\theta})\boldsymbol{\mu}^H(\boldsymbol{\theta})$ are differentiable. Analytical expression for the vector function $\boldsymbol{\psi}(\cdot;\cdot)$ can be directly obtained from relation (15.60) in [29]. By this relation and some elementary trace identities it follows that $\mathrm{E}\left[\boldsymbol{\psi}\left(\mathbf{X};\boldsymbol{\theta}\right); P_{\mathbf{X};\boldsymbol{\theta}}\right] = \mathbf{0}\ \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$, and hence,

$$\boldsymbol{\eta}(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \triangleq \mathrm{E}\left[\boldsymbol{\psi}(\mathbf{X};\boldsymbol{\theta}_0); P_{\mathbf{X};\boldsymbol{\theta}}\right] = \mathbf{0} \quad \text{for} \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0. \tag{6}$$

Therefore, when $\boldsymbol{\eta}(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \neq \mathbf{0}$ for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ an empirical estimate of $\boldsymbol{\eta}(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ can be used for testing between H_0 and H_1. Hence,