



HAL
open science

Random distortion testing with linear measurements

Dominique Pastor, François-Xavier Socheleau

► **To cite this version:**

Dominique Pastor, François-Xavier Socheleau. Random distortion testing with linear measurements. Signal Processing, 2018, 145, pp.116 - 126. 10.1016/j.sigpro.2017.11.017 . hal-01736179

HAL Id: hal-01736179

<https://hal.science/hal-01736179>

Submitted on 30 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Random Distortion Testing with Linear Measurements

Dominique Pastor, Francois-Xavier Socheleau

IMT Atlantique, Lab-STICC, UBL, Technopole Brest-Iroise CS83818, Brest 29238, France

Abstract

This paper addresses the problem of testing whether, after linear transformations and possible dimensionality reductions, a random matrix of interest Θ deviates significantly from some matrix model θ_0 , when Θ is observed in additive independent Gaussian noise with known covariance matrix. In contrast to standard likelihood theory, the probability distribution of Θ is assumed to be unknown. This problem generalizes the Random Distortion Testing (RDT) problem addressed in a former paper. Although the notions of size and power can be extended so as to deal with this generalized problem, no Uniformly Most Powerful (UMP) test exists for it. We can however exhibit a relevant subclass of tests and prove the existence of an optimal test within this class, that is, a test with specified size and maximal constant conditional power. As a consequence, this test is also UMP among invariant tests. The method fits within a wide range of signal processing scenarios. It is here specifically applied to sequential detection, subspace detection and random distortion testing after space-time compressive sensing. It is also used to extend the GLRT optimality properties for testing a waveform amplitude in noise.

Keywords: Hypothesis testing, random distortion testing, Mahalanobis norm, UMPI, sequential detection, subspace detection, compressive sensing, GLRT.

1. Introduction

In many applications, a sensor captures a multi-dimensional observation often modeled as a real random vector \mathbf{Y} that depends on some vector of

Email addresses: dominique.pastor@imt-atlantique.fr (Dominique Pastor),
fx.socheleau@imt-atlantique.fr (Francois-Xavier Socheleau)

interest Θ , possibly multi-dimensional as well. A basic problem is then to test whether this vector of interest, hereafter called the signal, deviates significantly from a given deterministic model θ_0 or not.

A first and usual approach involves posing this problem as a two-sided testing problem, where Θ is assumed to be deterministic and satisfy either the null hypothesis $\Theta = \theta_0$ or the alternative one $\Theta \neq \theta_0$. This approach can however be unsatisfactory in practice [1]. Indeed, such a binary hypothesis testing problem is an idealization of reality. Actually, Θ can hardly equate θ_0 , because θ_0 is merely a model and, as such, cannot take into account all the environmental fluctuations that may intervene, without control and regardless of measurement noise. It follows that Θ should better be modeled as a random vector that exhibits more or less big deviations with respect to θ_0 . In other terms, Θ should better be regarded as a *random distorted* version of θ_0 . In this respect, one may think of casting the problem into the Bayesian framework, which requires an *a priori* model for the probability distribution of Θ . Unfortunately, prior knowledge of the possible distribution of this distortion is hardly possible in many situations where the uncertainty on the model is so high that standard likelihood theory, including the Generalized Likelihood Ratio Test [2], the Rao score test [3, Sec. 3, p. 53] and the Wald test [4, p. 478, Theorem VIII], can poorly cope with the possible model mismatches.

Whence the proposition made in [1] to test whether Θ deviates significantly or not from θ_0 , the distortion being evaluated via an appropriate distortion measure. By so proceeding, the null hypothesis concerns only the small deviations of poor interest, whereas the alternative hypothesis encompasses large enough deviations, whose detection raises the user's interest. In [1], this problem is called Random Distortion Testing (RDT) and addressed under the additive and signal-independent Gaussian noise model, according to which the observation is modeled as:

$$\mathbf{Y} = \Theta + \mathbf{X}, \tag{1}$$

where Θ has unknown distribution and is independent of $\mathbf{X} \sim \mathcal{N}(0, \mathbf{C})$ with positive definite covariance matrix \mathbf{C} . The measure of distortion is then evaluated via the Mahalanobis norm [5] induced by \mathbf{C} . The problem then becomes that of testing whether the Mahalanobis norm of $\Theta - \theta_0$ is below or above some non-negative real tolerance τ . Because the distribution of Θ is not assumed to be known, standard likelihood theory does not apply to RDT. The solution established by [1, Theorem 2] is then a semi-parametric approach that basically combines the advantages of both parametric and non-parametric approaches: on the one hand, it guarantees robustness against

model mismatches and, on the other hand, it provides statistical optimality similar to Neyman-Pearson's [6].

It turns out that in many applications, sensors capture time-series of multi-dimensional observations. In these cases and under the Gaussian noise assumption, the model is still that of (1), where \mathbf{Y} , Θ and \mathbf{X} are now random matrices and independence of the signal and noise is defined by the independence of the vectorized versions of Θ and \mathbf{X} .

The present paper then addresses the problem of testing whether, after some linear transformations and some possible dimensionality reductions, the signal of interest Θ deviates significantly from some model θ_0 or not when we are given \mathbf{Y} . More precisely, suppose that the random matrix Θ is $N \times M$. Let \mathbf{A} and \mathbf{B} be two matrices that have full rank and suppose that \mathbf{A} is $n \times N$ with $n \leq N$ and that \mathbf{B} is $m \times M$ with $m \leq M$. Our purpose is then to test whether $\mathbf{A}\Theta\mathbf{B}^T$ deviates from a known $n \times m$ matrix θ_0 . It is actually an extension of the RDT problem of [1] and, as such, is hereafter called RDT with linear measurements (RDT ℓm). The RDT ℓm problem is thus very comprehensive by nature and actually fits within a wide range of signal processing applications. For instance, \mathbf{A} and/or \mathbf{B} can be projection matrices for either subspace detectors or (2D) compressive sensing, DFT matrices (in their composite real-valued forms), beamforming vectors, averaging matrices, time-varying filters etc. Specific examples will be detailed below.

The solution proposed in this paper to the RDT ℓm problem is encapsulated in our main result, namely Theorem 1. This result establishes the existence of an optimal agnostic test for RDT ℓm . By agnostic test, it is meant a test whose output does not depend whatsoever on the distribution of Θ . This test is hereafter called RDT ℓm test.

With respect to the outline above, the rest of the paper is divided in five sections, completed by several appendices. The next section analyzes the problem mathematically, which will lead to the introduction of definitions that extend those of [1] and are used to establish our theoretical results. These ones are then stated in Section 3. In this section, we begin with Theorem 1 before giving some corollaries of it. In particular, the optimality properties stated in Theorem 1 induces the optimality of the RDT ℓm test among invariant tests. Section 4 tackles several applications of Theorem 1 in statistical signal processing. First, it is shown that our present results encompass all those established in [1, 7, 8]. Second, a specific focus is given to an application of our approach to space-time compressive sensing, where experimental results illustrate the behavior of the RDT ℓm test. The generalized likelihood ratio test (GLRT) for testing a waveform amplitude in noise is also proved to be a RDT ℓm test.

Notation: All the vectors considered below are real and columns ones. The set of all natural numbers $1, 2, \dots$ is denoted by \mathbb{N} . Given a natural number d , the vector space of all d -dimensional real vectors is denoted as \mathbb{R}^d where, as usual, \mathbb{R} is the set of all real values.

All random vectors and variables encountered below are assumed to be defined on the same probability space $(\Omega, \mathcal{B}, \mathbb{P})$. For any natural number d , the set of all d -dimensional real random vectors defined on (Ω, \mathcal{B}) is hereafter denoted by $\mathcal{M}(\Omega, \mathbb{R}^d)$. Given $\mathbf{X} \in \mathcal{M}(\Omega, \mathbb{R}^d)$ and a Borel set B of \mathbb{R}^d , the event $\mathbf{X}^{-1}(B) = \{\omega \in \Omega : \mathbf{X}(\omega) \in B\}$ is hereafter denoted $[\mathbf{X} \in B]$. The conditional probability of $A \in \mathcal{B}$ given $[\mathbf{X} \in B]$ (resp. $[\mathbf{X} = \mathbf{x}]$ with $\mathbf{x} \in \mathbb{R}^d$) is denoted by $\mathbb{P}[A | \mathbf{X} \in B]$ (resp. $\mathbb{P}[A | \mathbf{X} = \mathbf{x}]$). The probability distribution of \mathbf{X} will be denoted as $\mathbb{P}_{\mathbf{X}}$ so that $\mathbb{P}_{\mathbf{X}}(B) = \mathbb{P}[\mathbf{X} \in B]$ for any Borel set B of \mathbb{R}^d . A domain of a random variable U is any measurable subset \mathcal{D} of \mathbb{R} such that $U \in \mathcal{D}$ (a-s), so that $\mathbb{P}[U \in \mathcal{D}] = \mathbb{P}_U(\mathcal{D}) = 1$.

Given two natural numbers k and ℓ , we denote the set of all real matrices with k rows and ℓ columns by the slight abuse of notation $\mathbb{R}^{k \times \ell}$. The set of all real random matrices with k rows and ℓ columns defined on (Ω, \mathcal{B}) is then denoted by $\mathcal{M}(\Omega, \mathbb{R}^{k \times \ell})$. The superscript T denotes transposition. The Kronecker product is denoted by \otimes . The *vectorization* of a given matrix \mathbf{A} , which converts \mathbf{A} into a column vector, is denoted by $\text{vec}(\mathbf{A})$. The dimension of $\text{vec}(\mathbf{A})$ will not be specified because it will be clear from the context. The inverse of the vectorization from $\mathbb{R}^{k \times \ell}$ to $\mathbb{R}^{k\ell}$ will be denoted vec^{-1} .

Let \mathbf{C} stand for some positive definite $d \times d$ covariance matrix. The Mahalanobis norm $\nu_{\mathbf{C}} : \mathbb{R}^d \rightarrow [0, \infty)$ defined in \mathbb{R}^d with respect to \mathbf{C} is hereafter called the \mathbf{C} -Mahalanobis norm. To any $\mathbf{y} \in \mathbb{R}^d$, it assigns the positive real number $\nu_{\mathbf{C}}(\mathbf{y}) = \sqrt{\mathbf{y}^{\text{T}} \mathbf{C}^{-1} \mathbf{y}}$. The Euclidean norm in \mathbb{R}^d is, as usual, denoted by $\|\bullet\|_2$ and assigns to any $\mathbf{y} \in \mathbb{R}^d$ the non negative real number $\|\mathbf{y}\| = \sqrt{\sum_{n=1}^d y_n^2}$ with $\mathbf{y} = (y_1, y_2, \dots, y_d)^{\text{T}}$. In other words, $\|\bullet\|_2$ is the \mathbf{I}_d -Mahalanobis norm, where \mathbf{I}_d henceforth denotes the $d \times d$ identity matrix.

Given $d \in \mathbb{N}$, a d -dimensional test is defined as a (measurable) map of \mathbb{R}^d into $\{0, 1\}$. Similarly, for $n, m \in \mathbb{N}$, a $n \times m$ -dimensional test is defined as a (measurable) map of $\mathbb{R}^{n \times m}$ into $\{0, 1\}$. The generalized Marcum function $Q_{d/2}$ is defined for any pair $(\rho, \eta) \in [0, \infty) \times [0, \infty)$ by $Q_{d/2}(\rho, \eta) = 1 - \mathbb{F}_{\chi_d^2(\rho^2)}(\eta^2)$, where $\mathbb{F}_{\chi_d^2(\rho^2)}$ is the cumulative distribution function of the non-central χ^2 distribution $\chi_d^2(\rho^2)$ with d degrees of freedom and non-central parameter ρ^2 . Given $\gamma \in (0, 1)$ and $\rho \geq 0$, $\lambda_d(\rho, \gamma) \geq 0$ denotes the unique solution in η to $Q_{d/2}(\rho, \eta) = \gamma$.

Finally, the indicator function of a subset A is denoted as $\mathbb{1}_A$ and \circ

designates, as usual, the composition operator.

2. Problem statement

Let \mathbf{Y} , Θ , \mathbf{X} be three elements of $\mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ where Θ and \mathbf{X} are independent in the sense that $\text{vec}(\Theta)$ and $\text{vec}(\mathbf{X})$ are independent. Throughout the rest of this paper, Θ and \mathbf{X} denote the signal and independent noise, respectively. To lighten the text, notation and formulas, this usual assumption of independence between Θ and \mathbf{X} will not be systematically recalled in the sequel. As mentioned above, Θ is the (multi-dimensional) signal of interest with unknown distribution, \mathbf{X} is the noise matrix satisfying $\text{vec}(\mathbf{X}) \sim \mathcal{N}(0, \Sigma)$, with known $NM \times NM$ positive definite covariance matrix Σ . We then consider the additive-noise model, where the observation matrix is $\mathbf{Y} = \Theta + \mathbf{X}$.

To model possible linear transforms and dimensionality reductions, we consider two matrices $\mathbf{A} \in \mathbb{R}^{n \times N}$ and $\mathbf{B} \in \mathbb{R}^{m \times M}$ where $n \leq N$ and $m \leq M$. We assume that \mathbf{A} and \mathbf{B} have full rank for technical reasons that will be made clear later on. As specified in the introduction, our purpose is then to test whether $\mathbf{A}\Theta\mathbf{B}^T$ deviates from a known $\theta_0 \in \mathbb{R}^{n \times m}$ by using an agnostic test satisfying a suitable optimality criterion.

To design our test, we need a measure of deviation between $\mathbf{A}\Theta\mathbf{B}^T$ and θ_0 that can be tested optimally and agnostically. Since norms are standard means to measure deviations between two vectors, it is thus natural to look for a suitable norm, hereafter denoted as $\nu : \mathbb{R}^{n \times m} \rightarrow [0, \infty)$, that meets the requirements of our test. We thus want to test the null hypothesis $\mathcal{H}_0(\Theta)$ that $\nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0)$ is lesser than some specified value $\tau \in [0, \infty)$ against the alternative hypothesis $\mathcal{H}_1(\Theta)$ that $\nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0)$ is larger than τ . Formally, the RDT ℓm problem can therefore be summarized as follows:

$$[\text{RDT}\ell m] : \left\{ \begin{array}{l} \text{Observation: } \mathbf{Y} = \Theta + \mathbf{X} \left\{ \begin{array}{l} \Theta, \mathbf{X} \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M}), \\ \text{vec}(\mathbf{X}) \sim \mathcal{N}(0, \Sigma), \\ \Theta \text{ and } \mathbf{X} \text{ independent,} \end{array} \right. \\ \mathcal{H}_0(\Theta) : \nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0) \leq \tau, \\ \mathcal{H}_1(\Theta) : \nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0) > \tau. \end{array} \right. \quad (2)$$

This problem does not concern a parameterized family of distributions for the observation, but the huge class of all the random matrices $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ with unknown distribution, whereas usual approaches in statistical signal processing generally assume deterministic unknown parameters or suppose that random parameters to test have known prior. In

this respect, the null and alternative hypotheses are then actual probabilistic events, in contrast to standard theory in statistical inference. Basically, testing $\mathcal{H}_0(\Theta)$ against $\mathcal{H}_1(\Theta)$ when the observation is \mathbf{Y} , means that we want to decide whether $\nu(\mathbf{A}\Theta(\omega)\mathbf{B}^T - \boldsymbol{\theta}_0) \leq \tau$ or not when we are given $\mathbf{Y}(\omega) = \Theta(\omega) + \mathbf{X}(\omega)$ with $\omega \in \Omega$.

The issue is then to exhibit a test $\mathfrak{T}^* : \mathbb{R}^{N \times M} \rightarrow \{0, 1\}$, optimal with respect to a certain criterion to be defined, to test $\mathcal{H}_0(\Theta)$ against $\mathcal{H}_1(\Theta)$. In the sequel, we put:

$$\begin{cases} \mathbb{P}[\mathcal{H}_0(\Theta)] \stackrel{\text{def}}{=} \mathbb{P}[\nu(\mathbf{A}\Theta\mathbf{B}^T - \boldsymbol{\theta}_0) \leq \tau] \\ \mathbb{P}[\mathcal{H}_1(\Theta)] \stackrel{\text{def}}{=} \mathbb{P}[\nu(\mathbf{A}\Theta\mathbf{B}^T - \boldsymbol{\theta}_0) > \tau] \end{cases}$$

Similarly, given $F \in \mathcal{B}$, let us write:

$$\begin{cases} \mathbb{P}[F | \mathcal{H}_0(\Theta)] \stackrel{\text{def}}{=} \mathbb{P}[F | \nu(\mathbf{A}\Theta\mathbf{B}^T - \boldsymbol{\theta}_0) \leq \tau] \\ \mathbb{P}[F | \mathcal{H}_1(\Theta)] \stackrel{\text{def}}{=} \mathbb{P}[F | \nu(\mathbf{A}\Theta\mathbf{B}^T - \boldsymbol{\theta}_0) > \tau] \end{cases}$$

To tackle the $\text{RDT}\ell m$ problem, we can easily transpose standard terminology in statistical inference by:

(i) Defining the $[\text{RDT}\ell m]$ -size of a test $\mathfrak{T} : \mathbb{R}^{N \times M} \rightarrow \{0, 1\}$ as:

$$\alpha_{\mathfrak{T}}^{[\text{RDT}\ell m]} = \sup_{\Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M}) : \mathbb{P}[\mathcal{H}_0(\Theta)] \neq 0} \mathbb{P}[\mathfrak{T}(\mathbf{Y}) = 1 | \mathcal{H}_0(\Theta)] \quad (3)$$

and saying that \mathfrak{T} has level γ if $\alpha_{\mathfrak{T}} \leq \gamma$.

(ii) Defining the $[\text{RDT}\ell m]$ -power of \mathfrak{T} as the map that assigns, to every $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ with $\mathbb{P}[\mathcal{H}_1(\Theta)] \neq 0$, the value:

$$\beta_{\mathfrak{T}}^{[\text{RDT}\ell m]}(\Theta) = \mathbb{P}[\mathfrak{T}(\mathbf{Y}) = 1 | \mathcal{H}_1(\Theta)]. \quad (4)$$

In a detection context, that is, if $\tau = 0$, $\alpha_{\mathfrak{T}}^{[\text{RDT}\ell m]}$ is the probability of false alarm and $\beta_{\mathfrak{T}}^{[\text{RDT}\ell m]}(\Theta)$ the probability of detection when the signal is Θ .

(iii) Seeking a test $\mathfrak{T}^{\text{UMP}}$ that would be Uniformly Most Powerful (UMP) with level γ in the sense that $\alpha_{\mathfrak{T}^*} \leq \gamma$ and, given $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ such that $\mathbb{P}[\mathcal{H}_1(\Theta)] \neq 0$

$$\beta_{\mathfrak{T}^{\text{UMP}}}^{[\text{RDT}\ell m]}(\Theta) \geq \beta_{\mathfrak{T}}^{[\text{RDT}\ell m]}(\Theta) \quad (5)$$

for any other test \mathfrak{T} with level γ .

Unfortunately, as shown in Appendix A, no such UMP test exists for the $\text{RDT}\ell m$ problem. We thus propose to restrict attention to the class of $[\text{RDT}\ell m]$ -coherent tests defined as follows.

Definition 1. An $N \times M$ -dimensional test $\mathfrak{T} : \mathbb{R}^{N \times M} \rightarrow \{0, 1\}$ is said to be [RDT ℓm]-coherent if:

[RDT ℓm -invariance] For any \mathbf{y} and \mathbf{y}' in $\mathbb{R}^{N \times M}$ such that $\mathbf{A}\mathbf{y}\mathbf{B}^T = \mathbf{A}\mathbf{y}'\mathbf{B}^T$, $\mathfrak{T}(\mathbf{y}) = \mathfrak{T}(\mathbf{y}')$;

[RDT ℓm -constant conditional power]: Given $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ independent of $\mathbf{X} \sim \mathcal{N}(0, \Sigma)$, there exists a domain \mathcal{D} of $\nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0)$ such that, for every $\rho \in \mathcal{D} \cap (\tau, \infty)$:

$$\mathbb{P}[\mathfrak{T}(\mathbf{Y}) = 1 \mid \nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0) = \rho] \text{ is independent of } \mathbb{P}_{\nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0)}. \quad (6)$$

The class of all [RDT ℓm]-coherent tests with level γ is henceforth denoted by $\mathcal{C}_\gamma^{\text{[RDT}\ell m\text{]}}$. \square

[RDT ℓm -constant conditional power] can also be defined by writing that, for any given $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ independent of $\mathbf{X} \sim \mathcal{N}(0, \Sigma)$, \mathfrak{T} satisfies (6) for $\mathbb{P}_{\nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0)}$ -almost every $\rho > \tau$, which avoids stipulating any domain \mathcal{D} . For readiness sake, this compact terminology will be avoided in the main core of the paper.

Basically, Definition 1 is a formalization of the rationale that a test for the RDT ℓm problem (2) must return the same result in the well-identified situations specified by [RDT ℓm -invariance] and [RDT ℓm -constant conditional power]. In this respect, the relevance of $\mathcal{C}_\gamma^{\text{[RDT}\ell m\text{]}}$ for RDT ℓm can be emphasized as follows.

1) [RDT ℓm -invariance] basically takes into account that the initial observation \mathbf{Y} is transformed into $\mathbf{A}\mathbf{Y}\mathbf{B}^T$, so that the test should not but return the same value for any \mathbf{y} and \mathbf{y}' in $\mathbb{R}^{N \times M}$ such that $\mathbf{A}\mathbf{y}\mathbf{B}^T = \mathbf{A}\mathbf{y}'\mathbf{B}^T$.

2) Regarding [RDT ℓm -constant conditional power], consider $\rho \in \mathcal{D} \cap (\tau, \infty)$ and $\omega \in \Omega$ such that $\nu(\mathbf{A}\Theta(\omega)\mathbf{B}^T - \theta_0) = \rho$. Because of the randomness of the signal and noise, no test \mathfrak{T} is guaranteed to take the correct decision to reject $\mathcal{H}_0(\Theta)$. We can however seek tests for which the conditional $\mathbb{P}[\mathfrak{T}(\mathbf{Y}) = 1 \mid \nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0) = \rho]$ is independent of Θ . This conditional basically plays the role of the power handled in standard statistical inference. To better grasp the meaning of this notion of constant conditional power for RDT ℓm , consider the particular case where $\mathbb{P}_\Theta = p\delta_\theta + (1-p)\delta_{\theta'}$ with $p \in [0, 1]$ and where δ_θ and $\delta_{\theta'}$ are the respective Dirac measures on two elements $\theta \neq \theta'$ of $\mathbb{R}^{N \times M}$ such that $\nu(\mathbf{A}\theta\mathbf{B}^T - \theta_0) = \rho$ and $\nu(\mathbf{A}\theta'\mathbf{B}^T - \theta_0) \neq \rho$. We then have $\mathbb{P}[\mathfrak{T}(\mathbf{Y}) = 1 \mid \nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0) = \rho] = \mathbb{P}[\mathfrak{T}(\mathbf{Y}) = 1]$. If \mathfrak{T} has constant conditional power given $\nu(\mathbf{A}\Theta\mathbf{B}^T - \theta_0) = \rho$, it guarantees a constant value $\mathbb{P}[\mathfrak{T}(\mathbf{Y}) = 1]$ whenever $\mathbf{Y} = \theta$, which is rather natural.

The set $\mathcal{C}_\gamma^{\text{[RDT}\ell m\text{]}}$ can be partially ordered as follows. Given two elements

\mathfrak{T} and \mathfrak{T}' of $\mathcal{C}_\gamma^{[\text{RDT}\ell m]}$, set $\mathfrak{T} \stackrel{[\text{RDT}\ell m]}{\preceq} \mathfrak{T}'$ if, given any $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$:

(i) The two tests \mathfrak{T} and \mathfrak{T}' satisfy [RDT ℓm -constant conditional power], which means the existence of a domain \mathcal{D} of $\nu(\mathbf{A}\Theta\mathbf{B}^\text{T} - \theta_0)$ such that, for every $\rho \in \mathcal{D} \cap (\tau, \infty)$, (6) holds true on \mathcal{D} for both \mathfrak{T} and \mathfrak{T}' ;

(ii) For any given $\rho \in \mathcal{D} \cap (\tau, \infty)$,

$$\begin{aligned} & \mathbb{P}[\mathfrak{T}(\mathbf{Y}) = 1 \mid \nu(\mathbf{A}\Theta\mathbf{B}^\text{T} - \theta_0) = \rho] \\ & \leq \mathbb{P}[\mathfrak{T}'(\mathbf{Y}) = 1 \mid \nu(\mathbf{A}\Theta\mathbf{B}^\text{T} - \theta_0) = \rho]. \end{aligned}$$

The question is then whether we can always exhibit and calculate a maximal element \mathfrak{T}^* in $\mathcal{C}_\gamma^{[\text{RDT}\ell m]}$. If such a maximal element $\mathfrak{T}^* \in \mathcal{C}_\gamma^{[\text{RDT}\ell m]}$ exists, we call it an *optimal [RDT ℓm]-coherent test* and it must satisfy:

[Maximality]: For any $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ and any $\mathfrak{T} \in \mathcal{C}_\gamma^{[\text{RDT}\ell m]}$,

$$\mathfrak{T} \stackrel{[\text{RDT}\ell m]}{\preceq} \mathfrak{T}^*.$$

At this stage, we do not know whether such an optimal [RDT ℓm]-coherent test actually exists for any given norm ν . We can however exhibit a norm ν for which the existence and a close form of this optimal test can be given. This is our main theoretical result stated in the next section.

3. Theoretical results

This section begins by stating the main theoretical result of this paper. After some comments on this result, we provide a series of applications of it.

3.1. The existence of an optimal [RDT ℓm]-coherent test

Theorem 1. *Let $\nu : \mathbb{R}^{n \times m} \rightarrow [0, \infty)$ be the norm on $\mathbb{R}^{n \times m}$ defined for each $\mathbf{z} \in \mathbb{R}^{n \times m}$ by setting $\nu(\mathbf{z}) = \nu_{\mathbf{C}}(\text{vec}(\mathbf{z})) = \sqrt{\text{vec}(\mathbf{z})^\text{T} \mathbf{C}^{-1} \text{vec}(\mathbf{z})}$, with $\mathbf{C} = (\mathbf{B} \otimes \mathbf{A}) \Sigma (\mathbf{B} \otimes \mathbf{A})^\text{T}$. Given $\gamma \in (0, 1)$ and $\tau \geq 0$, the test \mathfrak{T}^* defined for each $\mathbf{y} \in \mathbb{R}^{N \times M}$ by:*

$$\mathfrak{T}^*(\mathbf{y}) = \begin{cases} 1 & \text{if } \nu(\mathbf{A}\mathbf{y}\mathbf{B}^\text{T} - \theta_0) > \lambda_{nm}(\tau, \gamma) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where $\lambda_{nm}(\tau, \gamma)$ is the unique solution in η to $Q_{nm/2}(\tau, \eta) = \gamma$, is an optimal [RDT ℓm]-coherent test with size γ and, for any $\rho \in (0, \infty)$:

$$\mathbb{P}[\mathfrak{T}^*(\mathbf{Y}) = 1 \mid \nu(\mathbf{A}\Theta\mathbf{B}^\text{T} - \theta_0) = \rho] = Q_{nm/2}(\rho, \lambda_{nm}(\tau, \gamma)). \quad (8)$$

Proof. See Appendix B. □

The sufficiency specified by Definition 1 restricts our attention to a subclass of tests that act after transformation of the initial observation. The theorem above tells us that, although this restriction entails an unavoidable information loss, some optimality formulated in the initial observation domain can anyway be guaranteed.

Now, to better understand the decision-making performed by the RDT ℓm test \mathfrak{T}^* and its use of the Mahalanobis norm, note that

$$\mathfrak{T}^*(\mathbf{y}) = \mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}(\text{vec}(\mathbf{A} \mathbf{y} \mathbf{B}^T)) \quad (9)$$

for any $\mathbf{y} \in \mathbb{R}^{N \times M}$, where $\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}$ is defined for every $\mathbf{z} \in \mathbb{R}^{nm}$ by:

$$\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}(\mathbf{z}) = \begin{cases} 1 & \text{if } \nu_C(\mathbf{z} - \text{vec}(\boldsymbol{\theta}_0)) > \lambda_{nm}(\tau, \gamma) \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Eq. (9) shows that the decision is carried out in 3 steps: first, the observation is transformed; second, the outcome of the transformation is vectorized; third, the outcome of the vectorization is the statistics used to take the decision via $\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}$, which is based on the \mathbf{C} -Mahalanobis norm. It can then be easily guessed that the properties of $\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}$ stated by [1, Theorem 2] and resulting from those of the \mathbf{C} -Mahalanobis norm will play a crucial role in establishing the optimality of \mathfrak{T}^* .

It is also worth noticing that there exists a form without “vectorization” for \mathfrak{T}^* . This point is discussed in Appendix C and concerns the particular case where $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M \stackrel{iid}{\sim} \mathcal{N}(0, \mathbf{K})$ with positive definite covariance matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$.

3.2. Power and unbiasedness

Let $\boldsymbol{\Theta}$ be any element of $\mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ such that $\mathbb{P}[\mathcal{H}_1(\boldsymbol{\Theta})] \neq 0$. It follows from (B.10), (B.13) and (B.16) of Appendix B that $\mathbb{P}[\nu_C(\tilde{\boldsymbol{\Theta}} - \text{vec}(\boldsymbol{\theta}_0)) > \tau] \neq 0$ and:

$$\begin{aligned} \beta_{\mathfrak{T}^*}^{[\text{RDT}\ell m]}(\boldsymbol{\Theta}) &= \mathbb{P}[\mathfrak{T}^*(\mathbf{Y}) = 1 \mid \mathcal{H}_1(\boldsymbol{\Theta})] \\ &= \mathbb{P}[\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}(\tilde{\boldsymbol{\Theta}} + \tilde{\mathbf{X}}) = 1 \mid \nu_C(\tilde{\boldsymbol{\Theta}} - \text{vec}(\boldsymbol{\theta}_0)) > \tau] \end{aligned}$$

where $\tilde{\boldsymbol{\Theta}} = \text{vec}(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T)$ and $\tilde{\mathbf{X}} = \text{vec}(\mathbf{A} \mathbf{X} \mathbf{B}^T)$. It then follows from [1, Theorem 2 (ii)] that:

$$\beta_{\mathfrak{T}^*}^{[\text{RDT}\ell m]}(\boldsymbol{\Theta}) \geq Q_{nm/2}(\tau^+, \lambda_{nm}(\tau, \gamma)), \quad (11)$$

where $\tau^+ \in [\tau, \infty)$ is the supremum of the set of all those real values $t \geq \tau$ such that $\mathbb{P}[\tau < \nu(\mathbf{A}\Theta\mathbf{B}^T - \boldsymbol{\theta}_0) \leq t] = 0$:

$$\tau^+ = \sup \{t \geq \tau : \mathbb{P}[\tau < \nu(\mathbf{A}\Theta\mathbf{B}^T - \boldsymbol{\theta}_0) \leq t] = 0\}. \quad (12)$$

According to (11), \mathfrak{T}^* is also unbiased for RDT ℓm in that $\beta_{\mathfrak{T}^*}^{[\text{RDT}\ell m]}(\Theta) \geq \gamma$.

3.3. Invariance under group action and optimality

When no UMP test exists, it is common to restrict the attention to tests invariant to sets of transformations for which the problem is itself invariant [9, 10, 11]. The purpose is then to find a UMP test among this class of tests. In the problem under consideration, the notion of [RDT ℓm]-coherence is more immediate and natural than the notion of invariance. Since invariance is quite standard in the literature, and also for the sake of completeness, a group that leaves the problem invariant is nevertheless presented in this section. \mathfrak{T}^* is also shown to be UMP invariant with respect to this group.

There exists a group \mathcal{G} whose orbits are the surfaces:

$$\Gamma_\rho = \{\mathbf{y} \in \mathbb{R}^{N \times M} : \nu(\mathbf{A}\mathbf{y}\mathbf{B}^T - \boldsymbol{\theta}_0) = \rho\}, \rho \geq 0. \quad (13)$$

This group can be calculated by looking for transformations g such that, for all $\mathbf{y} \in \mathbb{R}^{N \times M}$:

$$\nu(\mathbf{A}g(\mathbf{y})\mathbf{B}^T - \boldsymbol{\theta}_0) = \nu(\mathbf{A}\mathbf{y}\mathbf{B}^T - \boldsymbol{\theta}_0), \quad (14)$$

or equivalently:

$$\|\mathbf{Q} \text{vec}(\mathbf{A}g(\mathbf{y})\mathbf{B}^T - \boldsymbol{\theta}_0)\|_2 = \|\mathbf{Q} \text{vec}(\mathbf{A}\mathbf{y}\mathbf{B}^T - \boldsymbol{\theta}_0)\|_2,$$

where $\mathbf{Q} = \boldsymbol{\Delta}^{-1/2}\mathbf{U}^T$, $\boldsymbol{\Delta} = \text{diag}(\xi_1, \dots, \xi_{nm})$ is the diagonal matrix whose diagonal elements ξ_1, \dots, ξ_{nm} are the eigenvalues of \mathbf{C} , $\boldsymbol{\Delta} = \text{diag}(\xi^{-1/2}, \dots, \xi_{nm}^{-1/2})$ and $\mathbf{U} \in O(nm)$ is such that $\mathbf{C} = \mathbf{U}\boldsymbol{\Delta}\mathbf{U}^T$ is the eigen-vector decomposition of \mathbf{C} . We thus have: $\mathbf{Q}^T\mathbf{Q} = \mathbf{C}^{-1}$.

Since the $\|\bullet\|_2$ -norm is invariant under unitary transforms, some routine algebra then leads to choose \mathcal{G} as the set of transforms $g_{\mathbf{R}}$ defined for every $\mathbf{y} \in \mathbb{R}^{N \times M}$ by:

$$g_{\mathbf{R}}(\mathbf{y}) = \text{vec}^{-1} \left((\mathbf{H}^+ - \mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}) \text{vec}(\boldsymbol{\theta}) + (\mathbf{Q}^{-1}\mathbf{R}\mathbf{Q}\mathbf{H} + \mathbf{I}_{MN} - \mathbf{H}^+\mathbf{H}) \text{vec}(\mathbf{y}) \right) \quad (15)$$

where $\mathbf{H} = \mathbf{B} \otimes \mathbf{A}$, $\mathbf{R} \in O(nm)$ and $\mathbf{H}^+ = \boldsymbol{\Sigma}\mathbf{H}^T(\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^T)^{-1}$ is a right pseudo-inverse of \mathbf{H} in that $\mathbf{H}\mathbf{H}^+ = \mathbf{I}_{nm}$. It then turns out that for any $\mathbf{R} \in$

$O(nm)$, (i) $\mathcal{H}_0(g_{\mathbf{R}}(\boldsymbol{\Theta})) = \mathcal{H}_0(\boldsymbol{\Theta})$ and $\mathcal{H}_1(g_{\mathbf{R}}(\boldsymbol{\Theta})) = \mathcal{H}_1(\boldsymbol{\Theta})$; (ii) $g_{\mathbf{R}}(\mathbf{Y}) = g_{\mathbf{R}}(\boldsymbol{\Theta}) + \mathbf{X}'$ where \mathbf{X}' satisfies $\text{vec}(\mathbf{X}') \sim \mathcal{N}(0, \boldsymbol{\Sigma})$. In the sense specified by (i) and (ii), the RDT ℓ m problem (2) is invariant under the action of \mathcal{G} .

Definition 2.

(i) \mathfrak{T} is said to be \mathcal{G} -invariant if $\mathfrak{T}(g(\mathbf{y})) = \mathfrak{T}(\mathbf{y})$ for any $g \in \mathcal{G}$ and any $\mathbf{y} \in \mathbb{R}^{N \times M}$.

(ii) \mathfrak{T} is said to be \mathcal{G} -UMPI with level (resp. size) $\gamma \in (0, 1)$ if $\mathfrak{T} \in \mathcal{C}_\gamma$ (resp. $\alpha_{\mathfrak{T}} = \gamma$) and

$$\beta_{\mathfrak{T}}(\boldsymbol{\Theta}) \geq \beta_{\mathfrak{T}'}(\boldsymbol{\Theta})$$

for any \mathcal{G} -invariant test $\mathfrak{T}' \in \mathcal{C}_\gamma$.

Theorem 2. *Everything being as above, \mathfrak{T}^* is \mathcal{G} -UMPI with size γ .*

Proof. See Appendix D. □

4. Applications

We begin by noticing that [1, Theorem 2], basically derives from Theorem 1 for $n = N = d$, $m = M = 1$, $\mathbf{A} = \mathbf{I}_N$ and $\mathbf{B} = 1$. Now, we can address several applications of the theoretical results established in the section above. Some of these applications extend standard results and others open new theoretical and practical prospects.

4.1. Optimality of GLRT for testing the amplitude level of a waveform in noise

In binary hypothesis testing, the probability density functions (PDFs) under the null and alternative hypothesis can be incompletely specified. This often happens in practice. In such cases, likelihood ratio tests such as Neyman-Pearson's do not apply. The Generalized Likelihood Ratio Test (GLRT) is then an alternative to such tests. It accommodates the unknown parameters of the PDFs by simply replacing these ones by their maximum likelihood estimates. Basically, the GLRT is not constructed so as to satisfy some optimality criterion. However, it often performs well, even in comparison to clairvoyant detectors [12, Chapter 6, Section 6.4.2], [13, 14].

Suppose that $\mathbf{Y} = \boldsymbol{\Theta} + \mathbf{X}$ where, as above, $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^d)$ and $\mathbf{X} \sim \mathcal{N}(0, \mathbf{I}_d)$. In addition, we assume that $\boldsymbol{\Theta} = \xi \mathbf{s}$ where the steering vector \mathbf{s} is known and such that $\|\mathbf{s}\|_2 = 1$, whereas the amplitude level $\xi \in \mathbb{R}$ is unknown.

Consider the composite hypothesis testing problem:

$$\left\{ \begin{array}{l} \text{Observation: } \mathbf{Y} = \xi \mathbf{s} + \mathbf{X} \left\{ \begin{array}{l} \Theta = \xi \mathbf{s}, \\ \mathbf{X} \sim \mathcal{N}(0, \mathbf{I}_d), \\ \Theta \text{ and } \mathbf{X} \text{ independent,} \end{array} \right. \\ \text{Null hyp. } (h_0): \xi = 0, \\ \text{Alternative hyp. } (h_1): \xi \neq 0. \end{array} \right. \quad (16)$$

For this testing problem, which can serve to pose the problem of detecting a radar target in noise, the GLRT test with size γ is hereafter denoted by Λ_G and is given by [12, Example 7.4.1, p. 255]:

$$\Lambda_G(\mathbf{Y}) = \begin{cases} 1 & \text{if } |\mathbf{s}^T \mathbf{Y}| > \eta \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

where η is the unique solution in t to the equation:

$$\mathbb{P}[|W| > \eta] = \gamma \text{ with } W \sim \mathcal{N}(0, 1). \quad (18)$$

The testing problem (16) turns out to be the particular RDT ℓm problem (2) with $N = d$, $M = m = n = 1$, $\mathbf{A} = \mathbf{s}^T$, $\mathbf{B} = 1$, $\mathbf{\Sigma} = \mathbf{C} = \mathbf{I}_d$, $\boldsymbol{\theta}_0 = 0$, $\nu_{\mathbf{C}}(\bullet) = |\bullet|$ and $\tau = 0$. Thereby, the assumption that ξ is deterministic unknown is even useless in (16). The RDT ℓm test \mathfrak{T}^* of Theorem 1 applied to \mathbf{Y} is then given by

$$\mathfrak{T}^*(\mathbf{Y}) = \mathfrak{T}_{\lambda_1(0, \gamma)}(|\mathbf{s}^T \mathbf{Y}|) = \begin{cases} 1 & \text{if } |\mathbf{s}^T \mathbf{Y}| > \lambda_1(0, \gamma) \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

Since $\lambda_1(0, \gamma)$ is the unique solution in η to the equation $Q_{1/2}(0, \eta) = \gamma$ and $Q_{1/2}(0, \eta) = \mathbb{P}[|W| > \eta]$ with $W \sim \mathcal{N}(0, 1)$ again, we conclude that the GLRT test (17) is exactly the RDT ℓm test (19). All the optimality properties of the RDT ℓm test stated in Section 3 thus transfer to the GLRT test. This finding extends [10] and may open new prospects in connection with GLRT theory, as enhanced in the conclusion of the present paper.

4.2. Detection of a change in the empirical mean of a process

Change detection is of great importance in many applications such as quality control, monitoring, tracking, fault detection or statistical process control [15, 16, 17].

Let $(\mathbf{Y})_{n \in \mathbb{N}}$ be a real N -dimensional discrete random process that we assume to be stationary in mean, in the sense that there exists $\boldsymbol{\theta} \in \mathbb{R}^d$ such

that $\mathbb{E}[\mathbf{Y}_n] = \boldsymbol{\theta}$ for each $n \in \mathbb{N}$. The change-in-mean detection problem is standardly posed [16] [17] as the binary hypothesis testing problem with null hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ and alternative hypothesis $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$ where $\boldsymbol{\theta}_0 \neq \boldsymbol{\theta}_1$ are two known elements of \mathbb{R}^N . In practice, it seems unrealistic to assume prior knowledge on the process distribution, especially if the alternative hypothesis is aimed at modeling the behavior of the process when this one is out of control. Therefore, prior knowledge of $\boldsymbol{\theta}_1$ is questionable and composite testing with alternative hypothesis $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ is certainly preferable. In addition, assuming iid observations for the observed process, either under H_0 or H_1 , is not acceptable in many statistical signal processing applications. Thence the following approach, which is an extension of [8] to the change-in-mean detection problem.

As the Shewhart chart [15][16, Sec. 2.2.1][17], we split the observation process into blocks of M samples each. Given the M samples $\mathbf{Y}_{(k-1)M+1}, \dots, \mathbf{Y}_{kM}$ of the process in the k th block, we form the random matrix $\mathbf{Y} = (\mathbf{Y}_{(k-1)M+1}, \dots, \mathbf{Y}_{kM}) \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$. We then assume that $\mathbf{Y} = \boldsymbol{\Theta} + \mathbf{X}$, where $\boldsymbol{\Theta}$ and \mathbf{X} are independent elements of $\mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ with $\mathbf{X} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$. As above, $\boldsymbol{\Sigma}$ is assumed to be positive definite. Instead of testing H_0 against H_1 in each successive block until H_0 is rejected, we propound to test whether the empirical mean of $\boldsymbol{\Theta}$ in each successive block deviates significantly from the nominal model $\boldsymbol{\theta}_0$ or not, until a block exhibits a significantly big drift between the empirical mean of $\boldsymbol{\Theta}$ and $\boldsymbol{\theta}_0$.

In such a formulation, $\boldsymbol{\Theta}$ is a signal fluctuating around $\boldsymbol{\theta}_0$. It is possibly non-stationary and its probability distribution is unknown, especially in case the process is out of control. Denoting the empirical mean of a given finite sequence x_1, x_2, \dots, x_M of N -dimensional real vectors by $\langle x \rangle_M$ so that $\langle x \rangle_M = (1/M) \sum_{m=1}^M x_m$, the rationale is then that $\langle \boldsymbol{\Theta} \rangle_M$ should not significantly deviate from $\boldsymbol{\theta}_0$ in the nominal situation. Therefore, following [8], the change-in-mean detection problem can be posed as the problem of testing whether $\nu_{\mathcal{C}}(\langle \boldsymbol{\Theta} \rangle_M - \boldsymbol{\theta}_0)$ is below or above tolerance $\tau \in [0, \infty)$, when we observe \mathbf{Y} . The choice of the Mahalanobis norm is motivated by the same reasons as above and the role of τ is to make the distinction between small and large deviations. The testing problem resulting from the foregoing

analysis is then summarized by:

$$\left\{ \begin{array}{l} \textbf{Observation: } \mathbf{Y} = \mathbf{\Theta} + \mathbf{X} \left\{ \begin{array}{l} \mathbf{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M}), \\ \text{vec}(\mathbf{X}) \sim \mathcal{N}(0, \mathbf{\Sigma}), \\ \mathbf{\Theta} \text{ and } \mathbf{X} \text{ independent,} \end{array} \right. \\ \textbf{Null hyp. } (\mathcal{H}_0(\mathbf{\Theta})): \nu_C(\langle \mathbf{\Theta} \rangle_M - \boldsymbol{\theta}_0) \leq \tau, \\ \textbf{Alternative hyp. } (\mathcal{H}_1(\mathbf{\Theta})): \nu_C(\langle \mathbf{\Theta} \rangle_M - \boldsymbol{\theta}_0) > \tau. \end{array} \right.$$

This problem is a particular case of (2) with $n = N$, $\mathbf{A} = \mathbf{I}_N$, $m = 1$ and $\mathbf{B} = \frac{1}{M} \underbrace{(1, \dots, 1)}_{M \text{ terms}}$. In this case,

$$\text{vec}(\mathbf{A} \mathbf{Y} \mathbf{B}^T) = \langle \mathbf{Y} \rangle_M \quad \text{and} \quad \text{vec}(\boldsymbol{\theta}_0) = \boldsymbol{\theta}_0.$$

Therefore, the RDT ℓm test provided by Theorem 1 and applied to \mathbf{Y} is $\mathfrak{T}^*(\mathbf{Y}) = \mathfrak{T}_{\lambda_N(\tau, \gamma)}(\langle \mathbf{Y} \rangle_M)$. The results given in [8] concern the particular case where $\mathbf{\Sigma} = \mathbf{I}_M \otimes \mathbf{K}$ and \mathbf{K} is positive definite, so that Appendix C directly applies. Note that this result can be extended to the problem of mean-change detection after compression by choosing an arbitrary $n \times N$, $n < N$, matrix \mathbf{A} .

4.3. Subspace detection

Since the seminal work of Scharf and Friedlander [9], subspace detectors have been widely used in various applications such as radar [18], functional MRI [19], passive acoustic monitoring [20], hyperspectral target detection [21] and many more.

Let $\mathbf{Y}, \mathbf{\Theta}, \mathbf{X}$ be three elements of $\mathcal{M}(\Omega, \mathbb{R}^{N \times 1})$ such that $\mathbf{Y} = \mathbf{\Theta} + \mathbf{X} \in \mathcal{M}(\Omega, \mathbb{R}^{N \times 1})$ where $\mathbf{\Theta}$ has unknown distribution, $\mathbf{X} \sim \mathcal{N}(0, \mathbf{\Sigma})$ is independent with $\mathbf{\Theta}$ and $\mathbf{\Sigma}$ is positive definite. We endow \mathbb{R}^N with the Mahalanobis norm $\nu_{\mathbf{\Sigma}}(\bullet)$ associated with $\mathbf{\Sigma}$ so that $(\mathbb{R}^N, \nu_{\mathbf{\Sigma}}(\bullet))$ is an Hilbert space. Let $\mathbf{\Psi}$ denote an $N \times n$ matrix that spans a rank- n subspace with $n \leq N$. The projection matrix associated with $\mathbf{\Psi}$ in the Hilbert space $(\mathbb{R}^N, \nu_{\mathbf{\Sigma}}(\bullet))$ is

$$\mathbf{P}_{\mathbf{\Psi}} = \mathbf{\Psi} (\mathbf{\Psi}^T \mathbf{\Sigma}^{-1} \mathbf{\Psi})^{-1} \mathbf{\Psi}^T \mathbf{\Sigma}^{-1}.$$

Given a tolerance $\tau \geq 0$, we address the problem of testing whether $\nu_{\mathbf{\Sigma}}(\mathbf{P}_{\mathbf{\Psi}} \mathbf{\Theta}(\omega)) \geq \tau$ or not, when we are given $\mathbf{Y}(\omega)$ for some $\omega \in \Omega$ and

when the probability distribution of Θ is unknown. The problem is summarized as follows:

$$\left\{ \begin{array}{l} \text{Observation: } Y = \Theta + X \left\{ \begin{array}{l} \Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times 1}), \\ X \sim \mathcal{N}(0, \Sigma), \\ \Theta \text{ and } X \text{ independent} \end{array} \right. \\ \text{Null hyp. } (\mathcal{H}_0(\Theta)): \nu_{\Sigma}(P_{\Psi}\Theta) \leq \tau, \\ \text{Alternative hyp. } (\mathcal{H}_1(\Theta)): \nu_{\Sigma}(P_{\Psi}\Theta) > \tau. \end{array} \right. \quad (20)$$

This problem actually extends the one considered in [7], which is the particular case of (20) when $\Sigma = \sigma^2 \mathbf{I}_N$. As such, Problem (20) is not formulated as a particular case of (2) because the Mahalanobis norm in (20) is computed with respect to Σ and not with respect to the noise covariance matrix after projection. However, by applying simple transforms, we show that (20) is actually a particular case of (2). More specifically, we have the following results.

Since $\Psi^T \Sigma^{-1} \Psi \in \mathcal{M}(\Omega, \mathbb{R}^{n \times n})$ is symmetric and positive definite, this matrix can be written in the form

$$\Psi^T \Sigma^{-1} \Psi = U \Delta U^T, \quad (21)$$

where U belongs to the orthogonal group of \mathbb{R}^n and Δ is diagonal with elements equal to the eigenvalues of $\Psi^T \Sigma^{-1} \Psi$. Define, $U_{\Psi} = \Sigma^{-1} \Psi U \Delta^{-1/2}$. For all $x \in \mathbb{R}^N$, the following relation is satisfied

$$\begin{aligned} \nu_{\Sigma}(P_{\Psi}x) &= \sqrt{(P_{\Psi}x)^T \Sigma^{-1} P_{\Psi}x} \\ &= \sqrt{x^T \Sigma^{-1} \Psi U \Delta^{-1/2} \Delta^{-1/2} U^T \Psi^T \Sigma^{-1} x} \\ &= \|\Delta^{-1/2} U^T \Psi^T \Sigma^{-1} x\|_2 \\ &= \|U_{\Psi}^T x\|_2. \end{aligned} \quad (22)$$

It then turns out that $\widetilde{X} = U_{\Psi}^T X \sim \mathcal{N}(0, \mathbf{I}_N)$. The testing problem (20) can thus be reformulated in the form:

$$\left\{ \begin{array}{l} \text{Observation: } Y = \Theta + X \left\{ \begin{array}{l} \Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times 1}), \\ X \sim \mathcal{N}(0, \Sigma), \\ \Theta \text{ and } X \text{ independent} \end{array} \right. \\ \text{Null hyp. } (\mathcal{H}_0(\Theta)): \|U_{\Psi}^T \Theta\|_2 \leq \tau, \\ \text{Alternative hyp. } (\mathcal{H}_1(\Theta)): \|U_{\Psi}^T \Theta\|_2 > \tau. \end{array} \right. \quad (23)$$

We see that the testing problem (23) is equivalent to (2) with $\boldsymbol{\theta}_0 = 0$, $m = M = 1$, $\mathbf{A} = \mathbf{U}_{\Psi}^T$, $\mathbf{B} = 1$ and $\mathbf{C} = \mathbf{U}_{\Psi}^T \boldsymbol{\Sigma} \mathbf{U}_{\Psi}$. Since $\mathbf{U}_{\Psi}^T \boldsymbol{\Sigma} \mathbf{U}_{\Psi} = \mathbf{I}_n$, it follows that $\nu_{\mathbf{C}}(\mathbf{U}_{\Psi}^T \mathbf{Y}) = \nu_{\boldsymbol{\Sigma}}(\mathbf{P}_{\Psi} \mathbf{Y})$. Therefore, the optimal test advised by Theorem 1 and applied to \mathbf{Y} is

$$\mathfrak{T}^*(\mathbf{Y}) = \mathfrak{T}_{\lambda_N(\tau, \gamma)}(\mathbf{U}_{\Psi}^T \mathbf{Y}).$$

We retrieve [7, Eq. (3)] when $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$. The results given above thus generalize [9].

4.4. Space-time compressive sensing

We here consider a compressive sensing based array processing scenario where \mathbf{Y} contains time series of multiple sensors. In this case, \mathbf{A} and \mathbf{B} can represent time and space compressed measurement matrices, respectively [22]. Such matrices are usually random and contain sub-Gaussian i.i.d. entries [23]. If one wants to apply a random distortion test directly in the compressed domain (without recovering the original signal), two natural questions to ask are: how to set the tolerance in the compressed domain if a given tolerance is specified in the original domain and what is the performance loss induced by running the test after compression instead of running it without compression? More formally, the original problem of interest is expressed as

$$\left\{ \begin{array}{l} \text{Observation: } \mathbf{Y} = \boldsymbol{\Theta} + \mathbf{X} \text{ with } \left\{ \begin{array}{l} \boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M}), \\ \text{vec}(\mathbf{X}) \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \\ \boldsymbol{\Theta} \text{ and } \mathbf{X} \text{ independent,} \end{array} \right. \\ \text{Null hyp. } (\mathcal{H}_0(\boldsymbol{\Theta})) : \nu_{\boldsymbol{\Sigma}}(\text{vec}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0^{\text{NC}})) \leq \tau_1, \\ \text{Alternative hyp. } (\mathcal{H}_1(\boldsymbol{\Theta})) : \nu_{\boldsymbol{\Sigma}}(\text{vec}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0^{\text{NC}})) > \tau_1, \end{array} \right. \quad (24)$$

and after compressive measurements it becomes

$$\left\{ \begin{array}{l} \text{Observation: } \mathbf{Z} = \mathbf{A} \mathbf{Y} \mathbf{B}^T \\ \text{Null hyp. } (\mathcal{H}_0(\boldsymbol{\Theta})) : \nu(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T - \boldsymbol{\theta}_0) \leq \tau_2, \\ \text{Alternative hyp. } (\mathcal{H}_1(\boldsymbol{\Theta})) : \nu(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T - \boldsymbol{\theta}_0) > \tau_2 \end{array} \right. \quad (25)$$

where $\boldsymbol{\theta}_0 = \mathbf{A} \boldsymbol{\theta}_0^{\text{NC}} \mathbf{B}^T$.

For a specified size γ and based on the results of Sec. 3, the optimal decision for problem (24) is

$$\mathfrak{T}_1^*(\mathbf{Y}) = \begin{cases} 1 & \text{if } \nu_{\boldsymbol{\Sigma}}(\text{vec}(\mathbf{Y} - \boldsymbol{\theta}_0^{\text{NC}})) > \lambda_{NM}(\tau_1, \gamma) \\ 0 & \text{otherwise} \end{cases} \quad (26)$$

and for problem (25)

$$\mathfrak{I}_2^*(\mathbf{Y}) = \begin{cases} 1 & \text{if } \nu_{\mathbf{C}}(\text{vec}(\mathbf{Z} - \boldsymbol{\theta}_0^{\text{NC}})) > \lambda_{nm}(\tau_2, \gamma) \\ 0 & \text{otherwise.} \end{cases} \quad (27)$$

Given that the tolerance τ_1 is fixed and known, how do we choose a reasonable value for τ_2 and what is the performance loss induced by compression?

To simplify the analysis, we assume that the measurement matrices are random orthoprojectors, *i.e.*, $\mathbf{A}\mathbf{A}^T = \mathbf{I}_n$ and $\mathbf{B}\mathbf{B}^T = \mathbf{I}_m$, and consider a diagonal noise covariance matrix $\boldsymbol{\Sigma} = \mathbf{I}_{NM}$. The norm in both tests (26) and (27) then becomes the ℓ_2 -norm. The choice of τ_2 can be made by assuming that $\|\text{vec}(\mathbf{Z} - \boldsymbol{\theta}_0)\|_2$ concentrates around $\sqrt{\frac{nm}{NM}}\|\text{vec}(\mathbf{Y} - \boldsymbol{\theta}_0^{\text{NC}})\|_2$. In fact, for $\epsilon \in (0, 1)$ and $c > 0$, the following relation is satisfied [24, 25]

$$\begin{aligned} \sqrt{1 - \epsilon}\|\text{vec}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0^{\text{NC}})\|_2 &\leq \sqrt{\frac{NM}{nm}}\|\text{vec}(\mathbf{A}\boldsymbol{\Theta}\mathbf{B}^T - \boldsymbol{\theta}_0)\|_2 \\ &\leq \sqrt{1 + \epsilon}\|\text{vec}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0^{\text{NC}})\|_2, \end{aligned} \quad (28)$$

with probability $1 - 2e^{-cnm\epsilon^2}$. Based on (28), a reasonable tolerance for problem (25) is therefore $\tau_2 = \sqrt{\frac{nm}{NM}}\tau_1$. The loss of performance due to compression cannot be assessed exactly as it depends on the distribution of the distortion around $\boldsymbol{\theta}_0^{\text{NC}}$, which is unknown. However, based on (28), a rough analysis can be conducted. From Sec. 3, the conditional power of \mathfrak{I}_1^* and \mathfrak{I}_2^* can be derived, that is

$$\mathbb{P}[\mathfrak{I}_1^*(\mathbf{Y}) = 1 \mid \|\text{vec}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0^{\text{NC}})\|_2 = \tau'_1 > \tau_1] = Q_{NM/2}(\tau'_1, \lambda_{NM}(\tau_1, \gamma)) \quad (29)$$

and

$$\mathbb{P}[\mathfrak{I}_2^*(\mathbf{Z}) = 1 \mid \|\text{vec}(\mathbf{A}\boldsymbol{\Theta}\mathbf{B}^T - \boldsymbol{\theta}_0)\|_2 = \tau'_2 > \tau_2] = Q_{nm/2}(\tau'_2, \lambda_{nm}(\tau_2, \gamma)). \quad (30)$$

The actual conditional power of interest useful to quantify the price of compression is $\mathbb{P}[\mathfrak{I}_2^*(\mathbf{Z}) = 1 \mid \|\text{vec}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0^{\text{NC}})\|_2 = \tau'_1 > \tau_1]$, *i.e.*, the probability of making the right decision in the compressed domain given that the distortion in the original domain is greater than the tolerance τ_1 . Based on (28) and (30), and following the same approach as in [25, Eq. (20)], the following approximation is obtained

$$\mathbb{P}[\mathfrak{I}_2^*(\mathbf{Z}) = 1 \mid \|\text{vec}(\boldsymbol{\Theta} - \boldsymbol{\theta}_0^{\text{NC}})\|_2 = \tau'_1 > \tau_1] \approx Q_{nm/2}\left(\sqrt{\frac{nm}{NM}}\tau'_1, \lambda_{nm}(\tau_2, \gamma)\right). \quad (31)$$

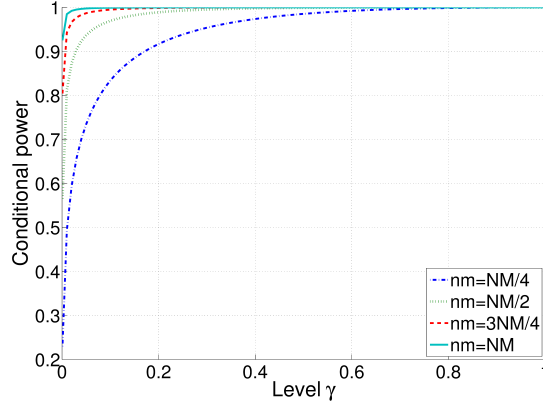


Figure 1: ROC curves, DNR=-5dB, $NM = 600$, $\tau_1 = \tau_1'/4$, $\tau_2 = \sqrt{\frac{nm}{NM}}\tau_1$.

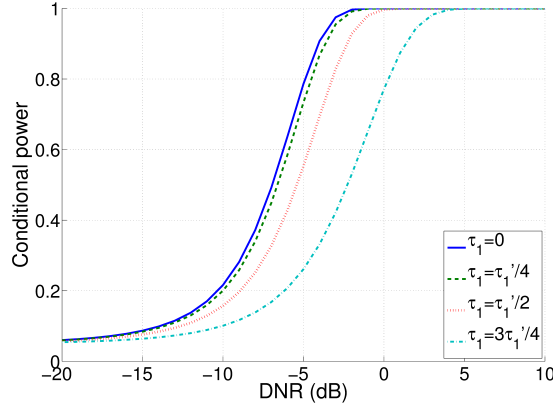


Figure 2: Conditional power as a function of DNR, $\gamma = 0.05$, $NM = 600$, $nm = 150$, $\tau_2 = \sqrt{\frac{nm}{NM}}\tau_1$.

Setting the same size γ for both tests, the performance loss is then the difference between (29) and (31). Note that in practice, depending on the given instance of matrices \mathbf{A} and \mathbf{B} , the actual performance may slightly deviate from (31). However, numerical simulations (not shown here) have confirmed that the expected performance curves with respect to random draws of \mathbf{A} and \mathbf{B} match with (31).

Figures 1 and 2 illustrate the impact on the performance of the various

parameters involved in problem (24) and (25). For both figures, the conditional power displayed is the one predicted by (31). The distortion-to-noise ratio (DNR) is defined as

$$\text{DNR} = \frac{\tau_1'^2}{\mathbb{E}[\|\text{vec}(\mathbf{X})\|_2^2]} = \frac{\tau_1'^2}{NM}. \quad (32)$$

Note that when $\boldsymbol{\theta}_0^{\text{NC}} = 0$, the DNR is homogeneous to a signal-to-noise ratio.

Fig. 1 shows the receiver operating characteristic (ROC) curve of \mathfrak{F}_2^* for several compression ratio. As expected, increasing nm improves the performance. However, for compression ratio smaller than 2 (*i.e.*, $NM/nm < 2$), the improvement becomes rather marginal.

The impact of the DNR on (31) is shown in Fig. 2 for several values of tolerance τ_1 . The choice of τ_1 is based on prior knowledge of the problem and a large value for τ_1 is sometimes desirable. This is the case for instance when we want to detect a random signal in noise with the possible presence of interference whose norm is bounded with a high probability. τ_1 is therefore useful to avoid recurrent triggering of false alarms when this interference is present. Consequently, increasing τ_1 , when τ_1' remains constant, makes the detector become more conservative, which yields a decrease in probability of correct detection as illustrated in Fig. 2.

5. Conclusions

On the basis of noisy observations, we have proposed a random distortion testing approach to decide whether a multi-dimensional signal of interest deviates significantly from some known model after some linear transformations and possible dimensionality reductions. As in [1], this problem does not concern a parameterized family of distributions for the observation, but the huge class of all the random matrices modeled by (1). In addition, the test bears on a random matrix with unknown distribution, whereas usual approaches in statistical signal processing generally assume deterministic unknown parameters or suppose that the random parameters to test have known prior. The proposed test, named $\text{RDT}\ell m$, generalizes the results presented in [1] and relies on the computation of Mahalanobis norm of the difference between the linearly transformed and compressed observation and the known model. The $\text{RDT}\ell m$ test is shown to satisfy optimal properties such as maximal constant conditional power and it is also uniformly most powerful among invariant tests. The framework described in this paper is very general and can be applied to a vast range of applications where decisions must be made

after linear transformations and/or compression of the observation. A few examples have been presented to illustrate the possible scope of the method.

Some findings presented in this paper open new prospects in connection with other decision theories. To begin with, the present paper embraces two different GLRT problems and tests. Indeed, in addition to Section 4.1, Section 4.3 establishes that the GLRT test for matched subspace detection [9] is a RDT ℓm test and, as such, satisfies the optimality properties of the latter. It can then be wondered whether such links between RDT ℓm and GLRT tests are merely coincidental or more fundamental. In this respect, algebraic properties encountered in this paper and [1] are certainly crucial. Such algebraic mechanisms may also deserve attention for the crucial role they have played above to transfer optimality from one representation domain to another. It should be profitable to better understand such mechanisms so as to generalize them for multiple transfer of optimality, from one single representation domain to possibly a sheaf of different representation domains. Finally, the RDT ℓm approach should also be relevant in machine learning for the following reasons. As many approaches in statistical inference, RDT ℓm tests can be regarded as means to decide whether observations after transformations are similar enough to models describing our prior knowledge. One main feature of the RDT ℓm approach is to allow for unknown observation distributions, which should therefore be beneficial to address novelty, outlier, anomaly and change-point detection issues in various machine learning approaches [26].

Appendix A.

With the notation introduced in Section 2, let us prove that there exists no UMP test \mathfrak{T}^* , that is, a test satisfying (5) in (iii). To this end, consider the specific case where Θ takes two values θ_0 and θ_1 in $\mathbb{R}^{N \times M}$ only, such that $\nu(\theta_0 - \theta_0) \leq \tau$ and $\nu(\theta_1 - \theta_0) > \tau$. In this case, the hypothesis $\nu(\mathbf{A} \Theta \mathbf{B}^T - \theta_0) \leq \tau$ is equivalent to the hypothesis $h_0 : \text{vec}(\mathbf{Y}) \sim \mathcal{N}(\text{vec}(\theta_0), \Sigma)$ and the alternative $\nu(\mathbf{A} \Theta \mathbf{B}^T - \theta_0) > \tau$ is equivalent to the hypothesis $h_1 : \text{vec}(\mathbf{Y}) \sim \mathcal{N}(\text{vec}(\theta_1), \Sigma)$. If a UMP test $\mathfrak{T}^{\text{UMP}}$ existed, then its size — in the usual sense [27] — for testing h_0 against h_1 would be less than or equal to γ . This follows from the expression of $\alpha_{\mathfrak{T}^{\text{UMP}}}$ deriving from (3). In addition, under h_1 , $\beta_{\mathfrak{T}^{\text{UMP}}}^{\text{[RDT}\ell m]}(\Theta) = \mathbb{P}[\mathfrak{T} \circ \text{vec}^{-1}(\text{vec}(\theta_1) + \text{vec}(\mathbf{X})) = 1]$ for any $N \times M$ -dimensional test. Therefore, the $\mathfrak{T}^{\text{UMP}}$ would verify $\mathbb{P}[\mathfrak{T}^{\text{UMP}} \circ \text{vec}^{-1}(\text{vec}(\theta_1) + \text{vec}(\mathbf{X})) = 1] \geq \mathbb{P}[\mathfrak{T} \circ \text{vec}^{-1}(\text{vec}(\theta_1) + \text{vec}(\mathbf{X})) = 1]$. In other words, $\mathfrak{T}^{\text{UMP}} \circ \text{vec}^{-1}$ would have power better — in the usual sense [27], again — than that of any other test. Therefore, this test should necessarily

equate the Neyman-Pearson test, which is known to exist for testing h_0 against h_1 . But this should hold for any pair $(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)$ of elements of $\mathbb{R}^{N \times M}$, which is impossible since the Neyman-Pearson test is different in each case.

Appendix B. Proof of Theorem 1

Our proof requires a series of elementary results. In this respect, we begin with the following two nested lemmas.

Lemma 1. *The linear map $\boldsymbol{x} \in \mathbb{R}^{N \times M} \rightarrow \boldsymbol{A} \boldsymbol{x} \boldsymbol{B}^T \in \mathbb{R}^{n \times m}$ is surjective.*

Proof. Let $\boldsymbol{y} \in \mathbb{R}^{n \times m}$. \boldsymbol{A} has full rank. Therefore $\boldsymbol{A} \boldsymbol{A}^T$ is invertible. Set $\boldsymbol{z} = \boldsymbol{A}^T (\boldsymbol{A} \boldsymbol{A}^T)^{-1} \boldsymbol{y}$. Since $\boldsymbol{z} \in \mathbb{R}^{N \times M}$, $\boldsymbol{v} = \boldsymbol{z}^T \in \mathbb{R}^{m \times N}$. Denote by \boldsymbol{v}_i with $i \in \llbracket 1, N \rrbracket$, the N columns of \boldsymbol{v} . Each \boldsymbol{v}_i is an element of \mathbb{R}^m and $\boldsymbol{v} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_N)$. Since \boldsymbol{B} has full rank then, for each $i \in \llbracket 1, N \rrbracket$, there is at least one $\boldsymbol{u}_i \in \mathbb{R}^M$ such that $\boldsymbol{B} \boldsymbol{u}_i = \boldsymbol{v}_i$. Set now $\boldsymbol{x} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_N)^T$. We have:

$$\begin{aligned} \boldsymbol{B} \boldsymbol{x}^T &= \boldsymbol{B}(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_N) \\ &= (\boldsymbol{B} \boldsymbol{u}_1, \boldsymbol{B} \boldsymbol{u}_2, \dots, \boldsymbol{B} \boldsymbol{u}_N) \\ &= (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_N) \\ &= \boldsymbol{v} \\ &= \boldsymbol{z}^T. \end{aligned}$$

Therefore, $\boldsymbol{x} \boldsymbol{B}^T = \boldsymbol{z}$ and $\boldsymbol{A} \boldsymbol{x} \boldsymbol{B}^T = \boldsymbol{A} \boldsymbol{z} = \boldsymbol{A} \boldsymbol{A}^T (\boldsymbol{A} \boldsymbol{A}^T)^{-1} \boldsymbol{y} = \boldsymbol{y}$. \square

Lemma 2. *An $N \times M$ -dimensional test $\mathfrak{T} : \mathbb{R}^{N \times M} \rightarrow \{0, 1\}$ satisfies [RDT ℓm -invariance] if and only if there exists a unique $n \times m$ -dimensional test $\bar{\mathfrak{T}} : \mathbb{R}^{n \times m} \rightarrow \{0, 1\}$ such that, for any $\boldsymbol{x} \in \mathbb{R}^{N \times M}$, $\mathfrak{T}(\boldsymbol{x}) = \bar{\mathfrak{T}}(\boldsymbol{A} \boldsymbol{x} \boldsymbol{B}^T)$. We call $\bar{\mathfrak{T}}$ the reduced form of the test \mathfrak{T} .*

Proof. Let $\mathfrak{T} : \mathbb{R}^{N \times M} \rightarrow \{0, 1\}$ be some given $N \times M$ -dimensional test. The existence of an $n \times m$ -dimensional test $\bar{\mathfrak{T}} : \mathbb{R}^{n \times m} \rightarrow \{0, 1\}$ such that $\mathfrak{T}(\boldsymbol{x}) = \bar{\mathfrak{T}}(\boldsymbol{A} \boldsymbol{x} \boldsymbol{B}^T)$ for all $\boldsymbol{x} \in \mathbb{R}^{N \times M}$ straightforwardly implies that \mathfrak{T} satisfies [RDT ℓm -invariance]. We thus limit our attention to the direct implication of the statement.

Given $\boldsymbol{y} \in \mathbb{R}^{n \times m}$, Lemma 1 implies that $\mathcal{K}(\boldsymbol{y}) = \{\boldsymbol{x} \in \mathbb{R}^{N \times M} : \boldsymbol{A} \boldsymbol{x} \boldsymbol{B}^T = \boldsymbol{y}\}$ is not empty. Since \mathfrak{T} is assumed to verify [RDT ℓm -invariance], \mathfrak{T} is constant on $\mathcal{K}(\boldsymbol{y})$. We can thus set $\bar{\mathfrak{T}}(\boldsymbol{y}) = \mathfrak{T}(\boldsymbol{x})$ where \boldsymbol{x} is any element of $\mathcal{K}(\boldsymbol{y})$. Now, for any $\boldsymbol{y} \in \mathbb{R}^{n \times m}$, $\boldsymbol{y} \in \mathcal{K}(\boldsymbol{A} \boldsymbol{y} \boldsymbol{B}^T)$ and $\bar{\mathfrak{T}}(\boldsymbol{A} \boldsymbol{y} \boldsymbol{B}^T) = \bar{\mathfrak{T}}(\boldsymbol{y})$.

We now establish that the reduced form $\bar{\mathfrak{T}}$ is unique. To this end, suppose that $\mathfrak{T}' : \mathbb{R}^{n \times m} \rightarrow \{0, 1\}$ is such that $\mathfrak{T}'(\boldsymbol{x}) = \mathfrak{T}'(\boldsymbol{A} \boldsymbol{x} \boldsymbol{B}^T)$. For any $\boldsymbol{y} \in \mathbb{R}^{n \times m}$, the surjectivity of $\boldsymbol{y} \mapsto \boldsymbol{A} \boldsymbol{y} \boldsymbol{B}^T$ guarantees the existence of $\boldsymbol{x} \in \mathbb{R}^{N \times M}$ such

that $\mathbf{y} = \mathbf{A} \mathbf{x} \mathbf{B}^T$. We thus have: $\mathfrak{T}'(\mathbf{y}) = \mathfrak{T}'(\mathbf{A} \mathbf{x} \mathbf{B}^T) = \overline{\mathfrak{T}}(\mathbf{A} \mathbf{x} \mathbf{B}^T) = \overline{\mathfrak{T}}(\mathbf{y})$. It follows that $\mathfrak{T}' = \overline{\mathfrak{T}}$.

Note that the binary valuation of \mathfrak{T} plays no actual role in the proof. \square

The results above lead to the following ones that will prove useful in the sequel. The notation introduced below will be kept throughout the rest of the proof with always the same meaning. To begin with, we define $\tilde{\mathbf{y}} \in \mathbb{R}^{nm}$ for each $\mathbf{y} \in \mathbb{R}^{N \times M}$ by setting:

$$\tilde{\mathbf{y}} = \text{vec}(\mathbf{A} \mathbf{y} \mathbf{B}^T). \quad (\text{B.1})$$

It then follows from [28, Lemma 2.2.2., Sec. 2.2., p. 74] that, for any $\mathbf{y} \in \mathbb{R}^{N \times M}$:

$$\tilde{\mathbf{y}} = \mathbf{H} \text{vec}(\mathbf{y}) \text{ with } \mathbf{H} = \mathbf{B} \otimes \mathbf{A}. \quad (\text{B.2})$$

In particular, it follows from the foregoing definition that:

[P1] $\widetilde{\mathbf{X}} \sim \mathcal{N}(0, \mathbf{C})$ with $\mathbf{C} = (\mathbf{B} \otimes \mathbf{A}) \boldsymbol{\Sigma} (\mathbf{B} \otimes \mathbf{A})^T$

[P2] The independence of $\boldsymbol{\Theta}$ and \mathbf{X} implies that of $\widetilde{\boldsymbol{\Theta}}$ and $\widetilde{\mathbf{X}}$.

It is also worth noticing that the bijectivity of vec and Lemma 1 directly entails that:

[P3] The map that assigns $\tilde{\mathbf{y}} \in \mathbb{R}^{nm}$ to any given $\mathbf{y} \in \mathbb{R}^{N \times M}$ is surjective.

For each $N \times M$ -dimensional test \mathfrak{T} satisfying [RDT ℓm -invariance], we now define the test $\widehat{\mathfrak{T}} : \mathbb{R}^{nm} \rightarrow \{0, 1\}$ by setting:

$$\widehat{\mathfrak{T}} = \overline{\mathfrak{T}} \circ \text{vec}^{-1}. \quad (\text{B.3})$$

It follows from this definition that:

$$\forall \mathbf{y} \in \mathbb{R}^{N \times M}, \mathfrak{T}(\mathbf{y}) = \widehat{\mathfrak{T}}(\tilde{\mathbf{y}}). \quad (\text{B.4})$$

The bijectivity of vec and the uniqueness of $\overline{\mathfrak{T}}$ guaranteed by Lemma 2 for any \mathfrak{T} satisfying [RDT ℓm -invariance] imply that:

[P4] The map that assigns $\widehat{\mathfrak{T}}$ to any given \mathfrak{T} satisfying [RDT ℓm -invariance] is injective.

A straightforward computation based on the results above yields the equality:

$$\forall \mathbf{y} \in \mathbb{R}^{N \times M}, \nu(\mathbf{A} \mathbf{y} \mathbf{B}^T - \boldsymbol{\theta}_0) = \nu_{\mathbf{C}}(\tilde{\mathbf{y}} - \text{vec}(\boldsymbol{\theta}_0)). \quad (\text{B.5})$$

Given any $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$, it straightforwardly follows from (B.5) that:

$$\mathbb{P}_{\nu(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T - \boldsymbol{\theta}_0)} = \mathbb{P}_{\nu_{\mathbf{C}}(\widetilde{\boldsymbol{\Theta}} - \text{vec}(\boldsymbol{\theta}_0))} \quad (\text{B.6})$$

and, for any $\rho \in [0, \infty)$, with the help of (B.4), we have:

$$\begin{aligned} & \mathbb{P}[\mathfrak{T}(\mathbf{\Theta} + \mathbf{X}) = 1 \mid \nu(\mathbf{A}\mathbf{\Theta}\mathbf{B}^T - \boldsymbol{\theta}_0) = \rho] \\ &= \mathbb{P}[\widehat{\mathfrak{T}}(\widetilde{\mathbf{\Theta}} + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathcal{C}}(\widetilde{\mathbf{\Theta}} - \text{vec}(\boldsymbol{\theta}_0)) = \rho] \end{aligned} \quad (\text{B.7})$$

On the basis of the foregoing results, we can now tackle the proof of Theorem 1. As a first step, let us consider the RDT problem — as defined in [1, see Sec. III, Eq. (5)] — of testing the hypothesis $\text{H}_0(\boldsymbol{\Xi})$ that $\nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0)) \leq \tau$ against its alternative $\text{H}_1(\boldsymbol{\Xi})$ that $\nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0)) > \tau$, when we observe $\mathbf{Z} = \boldsymbol{\Xi} + \widetilde{\mathbf{X}}$ where $\boldsymbol{\Xi} \in \mathbb{R}^{nm}$. This problem can be summarized as:

$$[\text{RDT}]: \begin{cases} \text{Observation: } \mathbf{Z} = \boldsymbol{\Xi} + \widetilde{\mathbf{X}} \begin{cases} \boldsymbol{\Xi} \in \mathcal{M}(\Omega, \mathbb{R}^{nm}), \\ \boldsymbol{\Xi} \text{ and } \widetilde{\mathbf{X}} \text{ independent,} \end{cases} \\ \text{H}_0(\boldsymbol{\Xi}) : \nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0)) \leq \tau, \\ \text{H}_1(\boldsymbol{\Xi}) : \nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0)) > \tau. \end{cases} \quad (\text{B.8})$$

Similarly to Section 2, we set:

$$\begin{cases} \mathbb{P}[\text{H}_0(\boldsymbol{\Xi})] \stackrel{\text{def}}{=} \mathbb{P}[\nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0)) \leq \tau] \\ \mathbb{P}[\text{H}_1(\boldsymbol{\Xi})] \stackrel{\text{def}}{=} \mathbb{P}[\nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0)) > \tau] \end{cases}$$

and given $F \in \mathcal{B}$:

$$\begin{cases} \mathbb{P}[F \mid \text{H}_0(\boldsymbol{\Xi})] \stackrel{\text{def}}{=} \mathbb{P}[F \mid \nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0)) = \tau] \\ \mathbb{P}[F \mid \text{H}_1(\boldsymbol{\Xi})] \stackrel{\text{def}}{=} \mathbb{P}[F \mid \nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0)) = \tau] \end{cases}$$

Given $\gamma \in (0, 1)$, define the class $\mathcal{C}_\gamma^{[\text{RDT}]}$ of those nm -dimensional tests $\mathfrak{T} : \mathbb{R}^{nm} \rightarrow \{0, 1\}$ with [RDT]-size:

$$\alpha_{\mathfrak{T}}^{[\text{RDT}]} \stackrel{\text{def}}{=} \sup_{\boldsymbol{\Xi} \in \mathcal{M}(\Omega, \mathbb{R}^{nm}) : \mathbb{P}[\mathcal{H}_0(\boldsymbol{\Xi})] \neq 0} \mathbb{P}[\mathfrak{T}(\boldsymbol{\Xi} + \widetilde{\mathbf{X}}) = 1 \mid \text{H}_0(\boldsymbol{\Xi})] \leq \gamma \quad (\text{B.9})$$

and [RDT-constant conditional power]: for any $\boldsymbol{\Xi} \in \mathcal{M}(\Omega, \mathbb{R}^{nm})$ independent of $\widetilde{\mathbf{X}}$, there exists a domain \mathcal{D} of $\nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0))$ such that, for any $\rho \in \mathcal{D} \cap (\tau, \infty)$,

$$\mathbb{P}[\mathfrak{T}(\boldsymbol{\Xi} + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0)) = \rho] \text{ is independent of } \mathbb{P}_{\nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0))}.$$

As straightforwardly verified, [RDT] is the particular [RDT ℓm] problem where, in (2), $\mathbb{R}^{N \times M}$, $\boldsymbol{\Theta}$ and \mathbf{X} are replaced by $\mathbb{R}^{n \times m}$, $\boldsymbol{\Xi}$ and $\widetilde{\mathbf{X}}$, respectively, and the $nm \times nm$ identity matrix is substituted for both \mathbf{A} and \mathbf{B} . In

this specific case, it follows from (7) that the test \mathfrak{T}^* of Theorem 1 reduces to $\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}$ defined in (10). [RDT ℓm]-size and [RDT ℓm -invariance] then reduce to [RDT]-size and [RDT-constant conditional power], respectively. The class $\mathcal{C}_\gamma^{\text{[RDT]}}$ is thus the particular instance of $\mathcal{C}_\gamma^{\text{[RDT}\ell m]}$ for the [RDT] problem (B.8).

We can then endow $\mathcal{C}_\gamma^{\text{[RDT]}}$ with a partial order $\preceq^{\text{[RDT]}}$ as follows. Given \mathfrak{T} and \mathfrak{T}' of $\mathcal{C}_\gamma^{\text{[RDT]}}$, say that $\mathfrak{T} \preceq^{\text{[RDT]}} \mathfrak{T}'$ if, given any $\Xi \in \mathcal{M}(\Omega, \mathbb{R}^{nm})$, there exists a domain \mathcal{D} of $\nu_{\mathbf{C}}(\Xi - \text{vec}(\boldsymbol{\theta}_0))$ such that, for every $\rho \in \mathcal{D} \cap (\tau, \infty)$, both $\mathbb{P}[\mathfrak{T}(\Xi + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathbf{C}}(\Xi - \text{vec}(\boldsymbol{\theta}_0)) = \rho]$ and $\mathbb{P}[\mathfrak{T}'(\Xi + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathbf{C}}(\Xi - \text{vec}(\boldsymbol{\theta}_0)) = \rho]$ are independent of $\nu_{\mathbf{C}}(\Xi - \text{vec}(\boldsymbol{\theta}_0)) = \rho$ and such that:

$$\begin{aligned} & \mathbb{P}[\mathfrak{T}(\Xi + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathbf{C}}(\Xi - \text{vec}(\boldsymbol{\theta}_0)) = \rho] \\ & \leq \mathbb{P}[\mathfrak{T}'(\Xi + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathbf{C}}(\Xi - \text{vec}(\boldsymbol{\theta}_0)) = \rho]. \end{aligned} \quad (\text{B.10})$$

According to [1, Theorem 2], it then turns out that $\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}$ is maximal in $\mathcal{C}_\gamma^{\text{[RDT]}}$ with [RDT]-size equal to γ , so that:

$$\alpha_{\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}}^{\text{[RDT]}} = \gamma \quad (\text{B.11})$$

and

$$\forall \mathfrak{T} \in \mathcal{C}_\gamma^{\text{[RDT]}}, \mathfrak{T} \preceq^{\text{[RDT]}} \mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}, \quad (\text{B.12})$$

This establishes Theorem 1 in this particular [RDT ℓm] problem that [RDT] is.

We now consider the general case. We begin by proving that \mathbf{C} is non-singular, which is necessary to guarantee the existence of the Mahalanobis norm involved in the definition of ν . Since $\mathbf{A} \in \mathbb{R}^{n \times N}$ and $\mathbf{B} \in \mathbb{R}^{m \times M}$, $\mathbf{B} \otimes \mathbf{A} \in \mathbb{R}^{nm \times NM}$. In addition, \mathbf{A} and \mathbf{B} have full rank. Therefore, $\text{rank}(\mathbf{B} \otimes \mathbf{A}) = \text{rank}(\mathbf{A})\text{rank}(\mathbf{B}) = nm$. The full rank of $\mathbf{B} \otimes \mathbf{A}$ then implies the nonsingularity of \mathbf{C} and the \mathbf{C} -Mahalanobis norm is actually well defined.

We are now going to show that:

[P5] If $\mathfrak{T} \in \mathcal{C}_\gamma^{\text{[RDT}\ell m]}$, then $\widehat{\mathfrak{T}} \in \mathcal{C}_\gamma^{\text{[RDT]}}$.

To this end, suppose that $\mathfrak{T} \in \mathcal{C}_\gamma^{\text{[RDT}\ell m]}$. We begin by comparing $\alpha_{\widehat{\mathfrak{T}}}^{\text{[RDT}\ell m]}$ to $\alpha_{\widehat{\mathfrak{T}}}^{\text{[RDT]}}$. Given $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ such that $\mathbb{P}[\mathcal{H}_0(\boldsymbol{\Theta})] \neq 0$, it follows from (B.4) and (B.5) that:

$$\mathbb{P}[\mathfrak{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1 \mid \mathcal{H}_0(\boldsymbol{\Theta})] = \mathbb{P}[\widehat{\mathfrak{T}}(\widetilde{\boldsymbol{\Theta}} + \widetilde{\mathbf{X}}) = 1 \mid \text{H}_0(\widetilde{\boldsymbol{\Theta}})] \quad (\text{B.13})$$

Therefore, according to Properties **(P1)** and **(P2)**, we obtain:

$$\begin{aligned} & \mathbb{P}[\mathfrak{I}(\boldsymbol{\Theta} + \mathbf{X}) = 1 \mid \mathcal{H}_0(\boldsymbol{\Theta})] \\ & \leq \sup_{\boldsymbol{\Xi} \in \mathcal{M}(\Omega, \mathbb{R}^{nm}): \mathbb{P}[\mathcal{H}_0(\boldsymbol{\Xi})] \neq 0} \mathbb{P}[\widehat{\mathfrak{I}}(\boldsymbol{\Xi} + \widetilde{\mathbf{X}}) = 1 \mid \mathcal{H}_0(\boldsymbol{\Xi})] \end{aligned} \quad (\text{B.14})$$

From (3) & (B.9), the inequality above implies that $\alpha_{\widehat{\mathfrak{I}}}^{[\text{RDT}\ell m]} \leq \alpha_{\mathfrak{I}}^{[\text{RDT}]}$.

Conversely, given any $\boldsymbol{\Xi} \in \mathcal{M}(\Omega, \mathbb{R}^{nm})$ such that $\mathbb{P}[\mathcal{H}_0(\boldsymbol{\Xi})] \neq 0$, **[P3]** and (B.5) induce the existence of $\boldsymbol{\Theta}_0 \in \mathbb{R}^{N \times M}$ such that $\boldsymbol{\Xi} = \widetilde{\boldsymbol{\Theta}}_0$ and $\mathbb{P}[\mathcal{H}_0(\boldsymbol{\Theta}_0)] \neq 0$. It results from (B.13) applied to $\boldsymbol{\Theta}_0$, that:

$$\mathbb{P}[\widehat{\mathfrak{I}}(\widetilde{\boldsymbol{\Theta}}_0 + \widetilde{\mathbf{X}}) = 1 \mid \mathcal{H}_0(\widetilde{\boldsymbol{\Theta}}_0)] = \mathbb{P}[\mathfrak{I}(\boldsymbol{\Theta}_0 + \mathbf{X}) = 1 \mid \mathcal{H}_0(\boldsymbol{\Theta}_0)]$$

As a consequence of this equality and the definition (3) of [RDT ℓm]-size, we obtain $\alpha_{\widehat{\mathfrak{I}}}^{[\text{RDT}]} \leq \alpha_{\mathfrak{I}}^{[\text{RDT}\ell m]}$, and we conclude from the foregoing that

$$\alpha_{\widehat{\mathfrak{I}}}^{[\text{RDT}\ell m]} = \alpha_{\mathfrak{I}}^{[\text{RDT}]}. \quad (\text{B.15})$$

Now, still given $\mathfrak{I} \in \mathcal{C}_\gamma^{[\text{RDT}\ell m]}$, we aim to show that $\widehat{\mathfrak{I}}$ has [RDT-constant conditional power]. To this end, let us consider an element $\boldsymbol{\Xi}$ of $\mathcal{M}(\Omega, \mathbb{R}^{nm})$. According to **[P3]**, there exists $\boldsymbol{\Theta}_0 \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ such that $\boldsymbol{\Xi} = \widetilde{\boldsymbol{\Theta}}_0$. Consider a domain \mathcal{D} of $\nu(\mathbf{A}\boldsymbol{\Theta}_0\mathbf{B}^T - \boldsymbol{\theta}_0)$. As a consequence of (B.5), \mathcal{D} is a domain of $\nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0))$ as well. It follows from (B.7) and [RDT ℓm -constant conditional power] satisfied by \mathfrak{I} that, for all $\rho \in \mathcal{D} \cap (\tau, \infty)$, $\mathbb{P}[\widehat{\mathfrak{I}}(\boldsymbol{\Xi} + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0)) = \rho]$ is independent of $\mathbb{P}_{\nu_{\mathcal{C}}(\boldsymbol{\Xi} - \text{vec}(\boldsymbol{\theta}_0))}$. Therefore, $\widehat{\mathfrak{I}}$ has [RDT-constant conditional power].

Conversely, suppose that $\widehat{\mathfrak{I}}$ has [RDT-constant conditional power]. Given any $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$, there exists a domain \mathcal{D} of $\nu_{\mathcal{C}}(\widetilde{\boldsymbol{\Theta}} - \text{vec}(\boldsymbol{\theta}_0))$ such that, for any $\rho \in \mathcal{D} \cap (\tau, \infty)$, $\mathbb{P}[\widehat{\mathfrak{I}}(\widetilde{\boldsymbol{\Theta}} + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathcal{C}}(\widetilde{\boldsymbol{\Theta}} - \text{vec}(\boldsymbol{\theta}_0)) = \rho]$ is independent of $\mathbb{P}_{\nu_{\mathcal{C}}(\widetilde{\boldsymbol{\Theta}} - \text{vec}(\boldsymbol{\theta}_0))}$. It then follows from (B.6) and (B.7) that $\mathbb{P}[\mathfrak{I}(\boldsymbol{\Theta} + \mathbf{X}) = 1 \mid \nu(\boldsymbol{\Theta} - \boldsymbol{\theta}_0) = \rho]$ is independent of $\mathbb{P}_{\nu(\mathbf{A}\boldsymbol{\Theta}\mathbf{B}^T - \boldsymbol{\theta}_0)}$. This exactly means that \mathfrak{I} satisfies [RDT ℓm -constant conditional power]. We thus have shown that \mathfrak{I} has [RDT ℓm -constant conditional power] if and only if $\widehat{\mathfrak{I}}$ has [RDT-constant conditional power]. This property and (B.15) establish **[P5]**.

We can now conclude the proof of Theorem 1. First, Lemma 2 and (10) imply that $\mathfrak{I}^* \in \mathcal{C}_\gamma^{[\text{RDT}\ell m]}$ with reduced form $\overline{\mathfrak{I}}^* = \mathfrak{I}_{\lambda_{nm}(\tau, \gamma)} \circ \text{vec}$. It then follows from this equality and (B.3) that

$$\widehat{\mathfrak{I}}^* = \overline{\mathfrak{I}}^* \circ \text{vec}^{-1} = \mathfrak{I}_{\lambda_{nm}(\tau, \gamma)}. \quad (\text{B.16})$$

and [P4] guarantees that \mathfrak{T}^* is the sole $N \times M$ -dimensional test satisfying (B.16). Therefore, thanks to (B.11) and (B.15), we have $\alpha_{\mathfrak{T}^*}^{[\text{RDT}\ell m]} = \alpha_{\widehat{\mathfrak{T}^*}}^{[\text{RDT}]} = \alpha_{\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}^{[\text{RDT}]}} = \gamma$, which establishes the $[\text{RDT}\ell m]$ -size of \mathfrak{T}^* . Second, \mathfrak{T}^* satisfies $[\text{RDT}\ell m\text{-constant conditional power}]$ and (8) as a direct consequence of (B.7), (B.16) and [1, Theorem 2, (i)]. It thus remain to prove that \mathfrak{T}^* is actually maximal in $\mathcal{C}_\gamma^{[\text{RDT}\ell m]}$. In this respect, let $\mathfrak{T} \in \mathcal{C}_\gamma^{[\text{RDT}\ell m]}$ and $\Theta \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$. By [P5], $\widehat{\mathfrak{T}} \in \mathcal{C}_\gamma^{[\text{RDT}]}$. We derive from (B.12) and (B.10) the existence of a domain \mathcal{D} of $\nu_{\mathcal{C}}(\widetilde{\Theta} - \text{vec}(\theta_0))$ such that, for every $\rho \in \mathcal{D} \cap (\tau, \infty)$, both $\mathbb{P}[\widehat{\mathfrak{T}}(\widetilde{\Theta} + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathcal{C}}(\widetilde{\Theta} - \text{vec}(\theta_0)) = \rho]$ and $\mathbb{P}[\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}(\widetilde{\Theta} + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathcal{C}}(\widetilde{\Theta} - \text{vec}(\theta_0)) = \rho]$ are independent of $\nu_{\mathcal{C}}(\widetilde{\Theta} - \text{vec}(\theta_0)) = \rho$ and such that:

$$\begin{aligned} & \mathbb{P}[\widehat{\mathfrak{T}}(\widetilde{\Theta} + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathcal{C}}(\widetilde{\Theta} - \text{vec}(\theta_0)) = \rho] \\ & \leq \mathbb{P}[\mathfrak{T}_{\lambda_{nm}(\tau, \gamma)}(\widetilde{\Theta} + \widetilde{\mathbf{X}}) = 1 \mid \nu_{\mathcal{C}}(\widetilde{\Theta} - \text{vec}(\theta_0)) = \rho]. \end{aligned}$$

It then suffices to inject (B.4), (B.7) and (B.16) into the inequality above to conclude that $\mathfrak{T} \preceq^{[\text{RDT}\ell m]} \mathfrak{T}^*$.

Appendix C. \mathfrak{T}^* without vectorization

Suppose that $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M \stackrel{iid}{\sim} \mathcal{N}(0, \mathbf{K})$ where $\mathbf{K} \in \mathbb{R}^{N \times N}$ is positive definite. We then have $\text{vec}(\mathbf{X}) \sim \mathcal{N}(0, \mathbf{I}_M \otimes \mathbf{K})$. In this case, $\Sigma = \mathbf{I}_M \otimes \mathbf{K}$. \mathbf{B} full rank implies that the symmetric matrix $\mathbf{B}\mathbf{B}^T \in \mathbb{R}^{m \times m}$ is positive definite and thus invertible. \mathbf{A} full rank implies that the symmetric matrix $\mathbf{A}\mathbf{K}\mathbf{A}^T \in \mathbb{R}^{n \times n}$ is positive definite and thus invertible. It follows from [28, statements (d), (f) & (g), p. 74, Sec. 2.2, Chap. 2] that $\mathbf{C}^{-1} = (\mathbf{B}\mathbf{B}^T)^{-1} \otimes (\mathbf{A}\mathbf{K}\mathbf{A}^T)^{-1}$. The compact expression we can give to \mathfrak{T}^* in this case is stated by Lemma 4, where Ψ and Φ are defined as follows.

With the same notation as above, let $\mathbf{B}\mathbf{B}^T = \mathbf{V}\boldsymbol{\delta}\mathbf{V}^T$ be an eigenvalue decomposition of the positive definite symmetric matrix $\mathbf{B}\mathbf{B}^T \in \mathbb{R}^{m \times m}$, where $\boldsymbol{\delta} = \text{diag}(\zeta_1, \zeta_2, \dots, \zeta_m)$ is a diagonal matrix such that $\zeta_1, \zeta_2, \dots, \zeta_m$ are the eigenvalues of $\mathbf{B}\mathbf{B}^T$ and $\mathbf{V} \in \mathbb{R}^{m \times m}$ is orthogonal. In the same way, let $\mathbf{A}\mathbf{K}\mathbf{A}^T = \mathbf{U}\boldsymbol{\Delta}\mathbf{U}^T$ be an eigenvalue decomposition of the positive definite symmetric matrix $\mathbf{A}\mathbf{K}\mathbf{A}^T$, where $\boldsymbol{\Delta} = \text{diag}(\xi_1, \xi_2, \dots, \xi_n)$ is a diagonal matrix such that $\xi_1, \xi_2, \dots, \xi_n$ are the eigenvalues of $\mathbf{A}\mathbf{K}\mathbf{A}^T$ and $\mathbf{U} \in \mathbb{R}^{n \times n}$ is orthogonal. Set $\Psi = \boldsymbol{\delta}^{-1/2} \mathbf{V}^T \in \mathbb{R}^{m \times m}$ with $\boldsymbol{\delta}^{-1/2} = \text{diag}(\zeta_1^{-1/2}, \zeta_2^{-1/2}, \dots, \zeta_m^{-1/2})$ and $\Phi = \boldsymbol{\Delta}^{-1/2} \mathbf{U}^T \in \mathbb{R}^{n \times n}$ with $\boldsymbol{\Delta}^{-1/2} = \text{diag}(\xi_1^{-1/2}, \xi_2^{-1/2}, \dots, \xi_n^{-1/2})$. We then have the following lemma.

Lemma 3. $\forall \mathbf{x} \in \mathbb{R}^{n \times m}, \nu_C(\text{vec}(\mathbf{x})) = \|\Phi \mathbf{x} \Psi^T\|_F$.

Proof. Since $(\mathbf{B}\mathbf{B}^T)^{-1} = \Psi^T \Psi = \mathbf{V} \delta^{-1} \mathbf{V}^T$ and $(\mathbf{A}\mathbf{K}\mathbf{A}^T)^{-1} = \Phi^T \Phi = \mathbf{U} \Delta^{-1} \mathbf{U}^T$, we derive from [28, Property (f), Section 2.2, p. 74] that:

$$\mathbf{C}^{-1} = (\mathbf{B}\mathbf{B}^T)^{-1} \otimes (\mathbf{A}\mathbf{K}\mathbf{A}^T)^{-1} = \Psi^T \Psi \otimes \Phi^T \Phi.$$

Therefore, for any $\mathbf{x} \in \mathbb{R}^{nm}$,

$$\nu_C(\mathbf{x}) = \sqrt{\mathbf{x}^T (\Psi^T \Psi \otimes \Phi^T \Phi) \mathbf{x}}.$$

For all $\mathbf{x} \in \mathbb{R}^{n \times m}$, it follows from [28, Lemma 2.2.3. (iii)]:

$$\begin{aligned} \text{vec}(\mathbf{x})^T (\Psi^T \Psi \otimes \Phi^T \Phi) \text{vec}(\mathbf{x}) &= \\ \text{tr}(\Psi \mathbf{x}^T \Phi^T \Phi \mathbf{x} \Psi^T) &= \text{tr}\left(\left(\Phi \mathbf{x} \Psi^T\right)^T \left(\Phi \mathbf{x} \Psi^T\right)\right). \end{aligned}$$

On the other hand, $\sqrt{\text{tr}(\mathbf{x}^T \mathbf{x})} = \|\mathbf{x}\|_F$. It follows from the foregoing that:

$$\nu_C(\mathbf{x}) = \sqrt{\text{tr}\left(\left(\Phi \mathbf{x} \Psi^T\right)^T \left(\Phi \mathbf{x} \Psi^T\right)\right)} = \|\Phi \mathbf{x} \Psi^T\|_F,$$

which concludes the proof. \square

The compact form for \mathfrak{T}^* is a straightforward consequence of the previous lemma and can be stated as follows.

Lemma 4. For any $\mathbf{x} \in \mathbb{R}^{N \times M}$,

$$\mathfrak{T}^*(\mathbf{x}) = \begin{cases} 1 & \text{if } \|\Phi (\mathbf{A} \mathbf{x} \mathbf{B}^T - \boldsymbol{\theta}_0) \Psi^T\|_F > \lambda_{nm}(\tau, \gamma) \\ 0 & \text{otherwise} \end{cases}$$

where $\|\bullet\|_F$ is the Froebenius norm in $\mathbb{R}^{N \times M}$.

Appendix D. Proof of Theorem 2

We begin with a partial extension of [1, Lemma 7].

Proposition 1. Given $\mathfrak{T} : \mathbb{R}^{N \times M} \rightarrow \{0, 1\}$ and $\mathbf{X} \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$, define the power function of \mathfrak{T} with respect to \mathbf{X} as the function $\beta_{\mathfrak{T}}(\bullet)$ defined for every $\boldsymbol{\theta} \in \mathbb{R}^{N \times M}$ by:

$$\beta_{\mathfrak{T}}(\boldsymbol{\theta}) = \mathbb{P}[\mathfrak{T}(\boldsymbol{\theta} + \mathbf{X}) = 1]$$

Let \equiv be a relation of equivalence in $\mathbb{R}^{N \times M}$. Suppose the existence of a function $\mu : \mathbb{R}^{N \times M} \rightarrow \mathbb{R}^q$ with $q \in \mathbb{N}$ such that:

$$\mu(\boldsymbol{\theta}) = \mu(\boldsymbol{\theta}') \quad \text{iff} \quad \boldsymbol{\theta} \equiv \boldsymbol{\theta}' \quad (\text{D.1})$$

for any given pair $(\boldsymbol{\theta}, \boldsymbol{\theta}')$ of elements of $\mathbb{R}^{N \times M}$. Let $\mu(\mathbb{R}^{N \times M})$ stand for the image of $\mathbb{R}^{N \times M}$ by μ .

If the power function of \mathfrak{T} is constant on each equivalence class of \equiv then, for any $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ independent of \mathbf{X} , \mathfrak{T} has constant conditional power given $\mu(\boldsymbol{\Theta}) = \mathbf{x}$ for $\mathbb{P}_{\mu(\boldsymbol{\Theta})}$ – almost every $\mathbf{x} \in \mu(\mathbb{R}^{N \times M})$, in the sense that, for $\mathbb{P}_{\mu(\boldsymbol{\Theta})}$ – almost every $\mathbf{x} \in \mu(\mathbb{R}^{N \times M})$:

$$\mathbb{P}[\mathfrak{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1 \mid \mu(\boldsymbol{\Theta}) = \mathbf{x}] = \beta_{\mathfrak{T}}(\boldsymbol{\theta})$$

for any $\boldsymbol{\theta}$ in the pre-image $\mu^{-1}(\{\mathbf{x}\})$ of \mathbf{x} .

Proof. Note that (D.1) amounts to saying that μ is constant on every equivalence class of \equiv . It follows from (D.1) that, given $\boldsymbol{\theta} \in \mathbb{R}^{N \times M}$, $\mu^{-1}(\{\mu(\boldsymbol{\theta})\})$ is the equivalence class of $\boldsymbol{\theta}$. The proof now mimics [1, Lemma 7] as follows.

If \mathfrak{T} has constant power function with respect to \mathbf{X} on each equivalence class of \equiv , we can define the map $\mathcal{R} : \mathbb{R}^q \rightarrow [0, 1]$ such that, for every $\mathbf{x} \in \mu(\mathbb{R}^{N \times M})$:

$$\mathcal{R}(\mathbf{x}) = \beta_{\mathfrak{T}}(\boldsymbol{\theta}) \quad (\text{D.2})$$

where $\boldsymbol{\theta}$ is any element of $\mu^{-1}(\{\mathbf{x}\})$.

Let $\boldsymbol{\Theta} \in \mathcal{M}(\Omega, \mathbb{R}^{N \times M})$ independent of \mathbf{X} and B be any borel set of \mathbb{R}^q . We derive from the definition of \mathcal{R} and the independence of $\boldsymbol{\Theta}$ and \mathbf{X} that:

$$\begin{aligned} \mathbb{P}[\mathfrak{T}(\boldsymbol{\Theta} + \mathbf{X}) = 1, \mu(\boldsymbol{\Theta}) \in B \mid \boldsymbol{\Theta} = \boldsymbol{\theta}] &= \mathbb{1}_B(\mu(\boldsymbol{\theta}))\beta_{\mathfrak{T}}(\boldsymbol{\theta}) \\ &= \mathbb{1}_B(\mu(\boldsymbol{\theta}))\mathcal{R}(\mu(\boldsymbol{\theta})) \end{aligned}$$

where $\mathbb{1}_B$ stands for the indicator function of B : given $\mathbf{x} \in \mathbb{R}^q$, $\mathbb{1}_A(\mathbf{x}) = 1$ if $\mathbf{x} \in A$ and $\mathbb{1}_A(\mathbf{x}) = 0$, otherwise. It follows from the standard change of variable [29, Theorem 17.2, p. 225]:

$$\begin{aligned} &\int \mathbb{1}_B(\mu(\boldsymbol{\theta}))\mathcal{R}(\mu(\boldsymbol{\theta}))\mathbb{P}_{\mu(\boldsymbol{\Theta})}(\text{d}\boldsymbol{\theta}) \\ &= \mathbb{E}[\mathbb{1}_B(\mu(\boldsymbol{\Theta}))\mathcal{R}(\mu(\boldsymbol{\Theta}))] \\ &= \int_{B \cap \mu(\mathbb{R}^{N \times M})} \mathcal{R}(\mathbf{x})\mathbb{P}_{\mu(\boldsymbol{\Theta})}(\text{d}\mathbf{x}) \end{aligned}$$

On the one hand, we derive from the foregoing that:

$$\begin{aligned}
& \mathbb{P}[\mathfrak{T}(\Theta + \mathbf{X}) = 1, \mu(\Theta) \in B] \\
&= \int \mathbb{P}[\mathfrak{T}(\Theta + \mathbf{X}) = 1, \mu(\Theta) \in B \mid \Theta = \boldsymbol{\theta}] \mathbb{P}_{\Theta}(\mathrm{d}\boldsymbol{\theta}) \\
&= \int_{B \cap \mu(\mathbb{R}^{N \times M})} \mathcal{R}(\mathbf{x}) \mathbb{P}_{\mu(\Theta)}(\mathrm{d}\mathbf{x})
\end{aligned}$$

On the other hand:

$$\begin{aligned}
& \mathbb{P}[\mathfrak{T}(\Theta + \mathbf{X}) = 1, \mu(\Theta) \in B] \\
&= \int_{B \cap \mu(\mathbb{R}^{N \times M})} \mathbb{P}[\mathfrak{T}(\Theta + \mathbf{X}) = 1 \mid \mu(\Theta) = \mathbf{x}] \mathbb{P}_{\mu(\Theta)}(\mathrm{d}\mathbf{x})
\end{aligned}$$

Since B is arbitrary, it follows from the equalities above and the definition of conditional probability that $\mathbb{P}[\mathfrak{T}(\Theta + \mathbf{X}) = 1 \mid \mu(\Theta) = \mathbf{x}] = \mathcal{R}(\mathbf{x})$ for $\mathbb{P}_{\mu(\Theta)}$ -almost every $\mathbf{x} \in \mu(\mathbb{R}^{N \times M})$. \square

We now complete the proof of Theorem 2. First, by definition of \mathcal{G} (see (14)), $\mathfrak{T}^* \in \mathcal{C}_\gamma$ is basically \mathcal{G} -invariant. Let $\mathfrak{T} \in \mathcal{C}_\gamma$ be another \mathcal{G} -invariant test. Since the orbits of \mathcal{G} form a partition of $\mathbb{R}^{N \times M}$, we can define the relation of equivalence \equiv for every given pair $(\boldsymbol{\theta}, \boldsymbol{\theta}')$ of $\mathbb{R}^{N \times M}$ by setting: $\boldsymbol{\theta} \equiv \boldsymbol{\theta}'$ if $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ belong to the same orbit. In addition, the map $\mu : \mathbb{R}^{N \times M} \rightarrow [0, \infty)$ defined for each $\mathbf{x} \in \mathbb{R}^{N \times M}$ by $\mu(\mathbf{x}) = \nu(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T - \boldsymbol{\theta}_0)$, is a maximal invariant of \mathcal{G} and, as such, satisfies (D.1). It thus follows from Proposition 1 that \mathfrak{T} has constant conditional power given $\nu(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T - \boldsymbol{\theta}_0) = \rho$ for $\mathbb{P}_{\nu(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T - \boldsymbol{\theta}_0)}$ -almost every $\rho \geq 0$. By Bayes's axiom, we have:

$$\begin{aligned}
\beta_{\mathfrak{T}}(\Theta) &= \frac{1}{\mathbb{P}[\mathcal{H}_1(\Theta)]} \times \\
& \int_{(\tau, \infty)} \mathbb{P}[\mathfrak{T}(\Theta + \mathbf{X}) = 1 \mid \nu(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T - \boldsymbol{\theta}_0) = \rho] \mathbb{P}_{\nu(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T - \boldsymbol{\theta}_0)} \mathrm{d}\rho
\end{aligned} \tag{D.3}$$

and

$$\begin{aligned}
\beta_{\mathfrak{T}^*}(\Theta) &= \frac{1}{\mathbb{P}[\mathcal{H}_1(\Theta)]} \\
& \int_{(\tau, \infty)} \mathbb{P}[\mathfrak{T}^*(\Theta + \mathbf{X}) = 1 \mid \nu(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T - \boldsymbol{\theta}_0) = \rho] \mathbb{P}_{\nu(\mathbf{A} \boldsymbol{\Theta} \mathbf{B}^T - \boldsymbol{\theta}_0)} \mathrm{d}\rho
\end{aligned} \tag{D.4}$$

By the maximality property established for \mathfrak{T}^* by Theorem 1, the integrand in (D.3) does not exceed the integrand in (D.4), which completes the proof.

References

- [1] D. Pastor, Q.-T. Nguyen, Random Distortion Testing and Optimality of Thresholding Tests, *IEEE Transactions on Signal Processing* 61 (16) (2013) 4161 – 4171.
- [2] J. Neyman, E. Pearson, On the use and interpretation of certain test criteria for purpose of statistical inference, *Biometrika* 20 (1928) 175 – 240.
- [3] C. Rao, Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Proceedings of the Cambridge Philosophical Society II* (1948) 50 – 57.
- [4] A. Wald, Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Transactions of the American Mathematical Society* 49 (3) (1943) 426 – 482.
- [5] P. Mahalanobis, On the generalised distance in statistics, *Proceedings of the National Institute of Sciences of India* 2 (1) (1936) 49 – 55.
- [6] D. Pastor, Q.-T. Nguyen, Random distortion testing and applications, in: *ICASSP 2013 : IEEE international conference on acoustics, speech and signal processing*, Vancouver, Canada, Mai 26-31, 2013.
- [7] F.-X. Socheleau, D. Pastor, Testing the energy of random signals in a known subspace: An optimal invariant approach, *IEEE Signal Processing Letters* 21 (10) (2014) 1182–1186.
- [8] D. Pastor, Q.-T. Nguyen, Robust statistical process control in block-rdt framework, in: *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, 2015, pp. 3896–3900.
- [9] L. L. Scharf, B. Friedlander, Matched subspace detectors, *IEEE Transactions on Signal Processing* 42 (8) (1994) 2146 – 2157.
- [10] S. Kay, J. Gabriel, An invariance property of the generalized likelihood ratio test, *IEEE Transactions on Signal Processing* 10 (12) (2003) 352 – 355.
- [11] D. Ciunzo, A. De Maio, D. Orlando, A unifying framework for adaptive radar detection in homogeneous plus structured interference-part i: On the maximal invariant statistic, *IEEE Transactions on Signal Processing* 64 (11) (2016) 2894–2906.

- [12] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume II, Detection Theory*, 14th printing, Prentice Hall, 2009.
- [13] S. M. Kay, J. R. Gabriel, Optimal invariant detection of a sinusoid with unknown parameters, *IEEE Transactions on Signal Processing* 50 (1) (2002) 27 – 40.
- [14] J. Gabriel, S. Kay, On the relationship between the GLRT and UMPI tests for the detection of signals with unknown parameters, *IEEE Transactions on Signal Processing* 53 (11) (2005) 4193 – 4203.
- [15] W. A. Shewhart, *Economic Control of Quality Manufactured Product*, D. Van Nostrand Reinhold, Princeton, NJ, 1931.
- [16] M. Basseville, I. Nikiforov, *Detection of Abrupt Changes - Theory and Application*, Prentice-Hall, Inc., 1993.
- [17] A. Tartakovsky, M. Basseville, I. Nikiforov, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, Chapman & Hall/CRC, 2014.
- [18] M. Rangaswamy, F. C. Lin, K. R. Gerlach, Robust adaptive signal processing methods for heterogeneous radar clutter scenarios, *Signal Processing* 84 (9) (2004) 1653–1665.
- [19] B. A. Ardekani, J. Kershaw, K. Kashikura, I. Kanno, Activation detection in functional mri using subspace modeling and maximum likelihood estimation, *IEEE Transactions on Medical Imaging* 18 (2) (1999) 101–114.
- [20] F.-X. Socheleau, E. Leroy, A. Carvallo Pecci, F. Samaran, J. Bonnel, J.-Y. Royer, Automated detection of antarctic blue whale calls, *The Journal of the Acoustical Society of America* 138 (5) (2015) 3105–3117.
- [21] H. Kwon, N. M. Nasrabadi, Kernel matched subspace detectors for hyperspectral target detection, *IEEE transactions on pattern analysis and machine intelligence* 28 (2) (2006) 178–194.
- [22] C.-L. Chang, G.-S. Huang, Low-complexity spatial-temporal filtering method via compressive sensing for interference mitigation in a gnss receiver, *International Journal of Antennas and Propagation* 2014 (501025).

- [23] E. J. Candes, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies?, *IEEE Trans. Inf. Theor.* 52 (12) (2006) 5406–5425.
- [24] S. Dasgupta, A. Gupta, An elementary proof of a theorem of johnson and lindenstrauss, *Random Structures & Algorithms* 22 (1) (2003) 60–65.
- [25] M. A. Davenport, P. T. Boufounos, M. B. Wakin, R. G. Baraniuk, Signal processing with compressive measurements, *IEEE Journal of Selected Topics in Signal Processing* 4 (2) (2010) 445–460.
- [26] M. A. F. Pimentel, D. A. Clifton, C. L., L. Tarassenko, A review of novelty detection, *Signal Processing* 99 (2014) 215 – 249.
- [27] E. L. Lehmann, J. P. Romano, *Testing Statistical Hypotheses*, 3rd edition, Springer, 2005.
- [28] R. J. Muirhead, *Aspects of multivariate statistical theory*, Wiley, 1982.
- [29] P. Billingsley, *Probability and Measure*, 3rd edition, Wiley, 1995.