Journal of Multivariate Analysis 146 (2016) 164–176

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Sharp minimax tests for large Toeplitz covariance matrices with repeated observations

Cristina Butucea, Rania Zgheib *

Université Paris-Est Marne-la-Vallée, LAMA(UMR 8050), UPEMLV F-77454, Marne-la-Vallée, France
ENSAE-CREST-GENES 3, ave. P. Larousse 92245 MALAKOFF Cedex, France

ARTICLE INFO

Article history:
Received 31 January 2015
Available online 26 September 2015

AMS 2000 subject classifications:
62G10
62H15
62G20
62H10

Keywords:
Toeplitz matrix
Covariance matrix
High-dimensional data
U-statistic
Minimax hypothesis testing
Optimal separation rates
Sharp asymptotic rates

ABSTRACT

We observe a sample of n independent p-dimensional Gaussian vectors with Toeplitz covariance matrix $\Sigma = [\sigma_{|i-j|}]_{1 \le i,j \le p}$ and $\sigma_0 = 1$. We consider the problem of testing the hypothesis that Σ is the identity matrix asymptotically when $n \to \infty$ and $p \to \infty$. We suppose that the covariances σ_k decrease either polynomially ($\sum_{k \ge 1} k^{2\alpha}\sigma_k^2 \le L$ for $\alpha > 1/4$ and $L > 0$) or exponentially ($\sum_{k \ge 1} e^{2Ak}\sigma_k^2 \le L$ for $A, L > 0$).

We consider a test procedure based on a weighted U-statistic of order 2, with optimal weights chosen as solution of an extremal problem. We give the asymptotic normality of the test statistic under the null hypothesis for fixed n and $p \to +\infty$ and the asymptotic behavior of the type I error probability of our test procedure. We also show that the maximal type II error probability, either tend to 0, or is bounded from above. In the latter case, the upper bound is given using the asymptotic normality of our test statistic under alternatives close to the separation boundary. Our assumptions imply mild conditions: $n = o(p^{2\alpha-1/2})$ (in the polynomial case), $n = o(e^p)$ (in the exponential case).

We prove both rate optimality and sharp optimality of our results, for $\alpha > 1$ in the polynomial case and for any $A > 0$ in the exponential case.

A simulation study illustrates the good behavior of our procedure, in particular for small n, large p.

1. Introduction

In the last decade, both functional data analysis (FDA) and high-dimensional (HD) problems have known an unprecedented expansion both from a theoretical point of view (as they offer many mathematical challenges) and for the applications (where data have complex structure and grow larger every day). Therefore, both areas share a large number of trends, see [3] and the review by [12], like regression models with functional or large-dimensional covariates, supervised or unsupervised classification, testing procedures, covariance operators.

Functional data analysis proceeds very often by discretizing curve datasets in time domain or by projecting on suitable orthonormal systems and produces large dimensional vectors with size possibly larger than the sample size. Hence methods and techniques from HD problems can be successfully implemented (see e.g. [1]). However, in some cases, HD vectors can be transformed into stochastic processes, see [9], and then techniques from FDA bring new insights into HD problems. Our work is of the former type.

* Corresponding author at: Université Paris-Est Marne-la-Vallée, LAMA(UMR 8050), UPEMLV F-77454, Marne-la-Vallée, France.
E-mail address: rania.zgheib@u-pem.fr (R. Zgheib).

http://dx.doi.org/10.1016/j.jmva.2015.09.003
0047-259X/

We observe independent, identically distributed Gaussian vectors $X_1, \ldots, X_n$, $n \geq 2$, which are p-dimensional, centered and with a positive definite Toeplitz covariance matrix Σ. We denote by $X_k = (X_{k,1}, \ldots, X_{k,p})^\top$ the coordinates of the vector X_k in $\mathbb{R}^p$ for all k.

Our model is that of a stationary Gaussian time series, repeatedly and independently observed n times, for $n \geq 2$. We assume that n and p are large. In functional data analysis, it is quite often that curves are observed in an independent way: electrocardiograms of different patients, power supply for different households and so on, see other datasets in [3]. After modelization of the discretized curves, the statistician will study the normality and the whiteness of the residuals in order to validate the model. Our problem is to test from independent samples of high-dimensional residual vectors that the standardized Gaussian coordinates are uncorrelated.

Let us denote by $\sigma_{|j|} = \mathrm{Cov}(X_{k,h}, X_{k,h+j})$, for all integer numbers h and j, for all $k \in \mathbb{N}^*$, where $\mathbb{N}^*$ is the set of positive integers. We assume that $\sigma_0 = 1$, therefore σ_j are correlation coefficients. We recall that $\{\sigma_j\}_{j\in\mathbb{N}}$ is a sequence of non-negative type, or, equivalently, the associated Toeplitz matrix Σ is non-negative definite. We assume that the sequence $\{\sigma_j\}_{j\in\mathbb{N}}$ belongs to $\ell_1(\mathbb{N}) \cap \ell_2(\mathbb{N})$, where $\ell_1(\mathbb{N})$ (resp. $\ell_2(\mathbb{N})$) is the set all absolutely (resp. square) summable sequences. It is therefore possible to construct a positive, periodic function

$$f(x) = \frac{1}{2\pi}\left(1 + 2\sum_{j=1}^{\infty} \sigma_j \cos(jx)\right), \quad \text{for } x \in (-\pi, \pi),$$

belonging to $\mathbb{L}_2(-\pi, \pi)$ the set of all square-integrable functions f over $(-\pi, \pi)$. This function is known as the spectral density of the stationary series $\{X_{k,i}, i \in \mathbb{Z}\}$.

We solve the following test problem,

$$H_0 : \Sigma = I \tag{1.1}$$

versus the alternative

$$H_1 : \Sigma \in \mathcal{T}(\alpha, L) \quad \text{such that } \sum_{j\geq 1} \sigma_j^2 \geq \psi^2, \tag{1.2}$$

for $\psi = (\psi_{n,p})_{n,p}$ a positive sequence converging to 0. From now on, $C_{>0}$ denotes the set of squared symmetric and positive definite matrices. The set $\mathcal{T}(\alpha, L)$ is an ellipsoid of Sobolev type

$$\mathcal{T}(\alpha, L) = \left\{ \Sigma \in C_{>0}, \Sigma \text{ is Toeplitz}; \sum_{j\geq 1} \sigma_j^2 j^{2\alpha} \leq L \text{ and } \sigma_0 = 1 \right\}, \quad \alpha > 1/4,\ L > 0.$$

We shall also test (1.1) against

$$H_1 : \Sigma \in \mathcal{E}(A, L) \quad \text{such that } \sum_{j\geq 1} \sigma_j^2 \geq \psi^2, \text{ for } \psi > 0, \tag{1.3}$$

where the ellipsoid of covariance matrices is given by

$$\mathcal{E}(A, L) = \left\{ \Sigma \in C_{>0}, \Sigma \text{ is Toeplitz}; \sum_{j\geq 1} \sigma_j^2 e^{2Aj} \leq L \text{ and } \sigma_0 = 1 \right\}, \quad A, L > 0.$$

This class contains the covariance matrices whose elements decrease exponentially, when moving away from the diagonal. We denote by $G(\psi)$ either $G(\mathcal{T}(\alpha, L), \psi)$ the set of matrices under the alternative (1.2) or $G(\mathcal{E}(A, L), \psi)$ under the alternative (1.3).

We stress the fact that a matrix Σ in $G(\psi)$ is such that $1/(2p)\|\Sigma - I\|_F^2 \geq \sum_{j\geq 1} \sigma_j^2 \geq \psi^2$, i.e. Σ is outside a neighborhood of I with radius ψ in Frobenius norm.

Our test can be applied in the context of model fitting for testing the whiteness of the standard Gaussian residuals. In this context, it is natural to assume that the covariance matrix under the alternative hypothesis has small entries like in our classes of covariance matrices. Such tests have been proposed by [15], where it is noted that weighted test statistics can be more powerful.

Note that, most of the literature on testing the null hypothesis (1.1), either focus on finding the asymptotic behavior of the test statistic under the null hypothesis, or control in addition the type II error probability for one fixed unknown matrix under the alternative, whereas our main interest is to quantify the worst type II error probabilities, i.e. uniformly over a large set of possible covariance matrices.

Various test statistics in high dimensional settings have been considered for testing (1.1), as it was known for some time that likelihood ratio tests do not converge when dimension grows. Therefore, a corrected Likelihood Ratio Test is proposed in [2] when $p/n \to c \in (0, 1)$, and its asymptotic behavior is given under the null hypothesis, based on the random matrix theory. In [24] the result is extended to $c = 1$. An exact test based on one column of the covariance matrix is constructed by [20]. A series of papers propose test statistics based on the Frobenius norm of $\Sigma - I$, see [25,30,31,10]. Different test statistics are introduced and their asymptotic distribution is studied. In particular in [10] the test statistic is a U-statistic with constant weights. An unbiased estimator of $\mathrm{tr}(\Sigma - B_k(\Sigma))^2$ is constructed in [28], where $B_k(\Sigma) = (\sigma_{ij} \cdot I\{|i-j| \leq k\})$, in order to develop a test statistic for the problem of testing the bandedness of a given matrix. Another extension of our test problem is to test the sphericity hypothesis $\Sigma = \sigma^2 I$, where $\sigma^2 > 0$ is unknown. [16] introduced a test statistic based on functionals of order 4 of the covariance matrix. Motivated by these results, the test $H_0 : \Sigma = I$ is revisited by [14]. The maximum value of non-diagonal elements of the empirical covariance matrix was also investigated as a test statistic. Its asymptotic extreme-value distribution was given under the identity covariance matrix by [6] and for other covariance matrices by [32]. We propose here a new test statistic to test (1.1) which is a weighted U-statistic of order 2 and study its probability errors uniformly over the set of matrices given by (1.2) and (1.3).

The test problem with alternative (1.2) and with one sample ($n = 1$) was solved in the sharp asymptotic framework, as $p \to \infty$, by [13]. Indeed, [13] studies sharp minimax testing of the spectral density f of the Gaussian process. Note that under the null hypothesis we have a constant spectral density $f_0(x) = 1/(2\pi)$ for all x and the alternative can be described in $\mathbb{L}_2$ norm as we have the following isometry $\|f - f_0\|_2^2 = (2\pi)^{-1}\|\Sigma - I\|_F^2$. Moreover, the ellipsoid of covariance matrices $\mathcal{T}(\alpha, L)$ are in bijection with Sobolev ellipsoids of spectral densities f. Let us also recall that the adaptive rates for minimax testing are obtained for the spectral density problem by [19] by a non constructive method using the asymptotic equivalence with a Gaussian white noise model. Finding explicit test procedures which adapt automatically to parameters α and/or L of our class of matrices will be the object of future work. Our efforts go here into finding sharp minimax rates for testing.

Our results generalize the results in [13] to the case of repeatedly observed stationary Gaussian process. We stress the fact that repeated sampling of the stationary process $(X_{1,1}, \ldots, X_{1,p})$ to $(X_{n,1}, \ldots, X_{n,p})$ can be viewed as one sample of size $n \times p$ under the null hypothesis. However, this sample will not fit the assumptions of our alternative. Indeed, under the alternative, its covariance matrix is not Toeplitz, but block diagonal. Moreover, we can summarize the n independent vectors into one p-dimensional vector $X = n^{-1/2}\sum_{k=1}^n X_k$ having Gaussian distribution $\mathcal{N}_p(0, \Sigma)$. The results by [13] will produce a test procedure with rate that we expect optimal as a function of p, but more biased and suboptimal as a function of n. The test statistic that we suggest removes cross-terms and has smaller bias. Therefore, results in [13] do not apply in a straightforward way to our setup.

A conjecture in the sense of asymptotic equivalence of the model of repeatedly observed Gaussian vectors and a Gaussian white noise model was given by [8]. Our rates go in the sense of the conjecture.

The test of $H_0 : \Sigma = I$ against (1.2), with Σ not necessary Toeplitz, is given in [5]. Their rates show a loss of a factor p when compared to the rates for Toeplitz matrices obtained here. This can be interpreted heuristically by the size of the set of unknown parameters which is $p(p-1)/2$ for [5] whereas here it is p. We can see that the family of Toeplitz matrices is a subfamily of general covariance matrices in [5]. Therefore, the lower bounds are different, they are attained through a particular family of Toeplitz large covariance matrices. The upper bounds take into account as well the fact that we have repeated information on the same diagonal elements. The test statistic is different from the one used in [5].

The test problem with alternative hypothesis (1.3) has not been studied in this model. The class $\mathcal{E}(A, L)$ contains matrices with exponentially decaying elements when further from the main diagonal. The spectral density function associated to this process belongs to the class of functions which are in $\mathbb{L}_2$ and admit an analytic continuation on the strip of complex numbers z with $|\mathrm{Im}(z)| \leq A$. Such classes of analytic functions are very popular in the literature of minimax estimation, see [18].

In times series analysis such covariance matrices describe among others the linear ARMA processes. The problem of adaptive estimation of the spectral density of an ARMA process has been studied by [17] (for known α) and adaptively to α via wavelet based methods by [27] and by model selection by [11]. In the case of an ARFIMA process, obtained by fractional differentiation of order $d \in (-1/2, 1/2)$ of a casual invertible ARMA process, [29] gave adaptive estimators of the spectral density based on the log-periodogram regression model when the covariance matrix belongs to $\mathcal{E}(A, L)$.

Before describing our results let us define more precisely the quantities we are interested in evaluating.

1.1. Formalism of the minimax theory of testing

Let χ be a test, that is a measurable function of the observations $X_1, \ldots, X_n$ taking values in $\{0, 1\}$ and recall that $G(\psi)$ corresponds to the set of covariance matrices under the alternative hypothesis. Let

$\eta(\chi) = \mathbb{E}_I(\chi)$ be its type I error probability, and

$\beta(\chi, G(\psi)) = \sup_{\Sigma \in G(\psi)} \mathbb{E}_\Sigma(1 - \chi)$ be its maximal type II error probability.

We consider two criteria to measure the performance of the test procedure. The first one corresponds to the classical Neyman–Pearson criterion. For $w \in (0, 1)$, we define,

$$\beta_w(G(\psi)) = \inf_{\chi;\, \eta(\chi) \leq w} \beta(\chi, G(\psi)).$$

The test χ_w is asymptotically minimax according to the Neyman–Pearson criterion if

$$\eta(\chi_w) \leq w + o(1) \quad \text{and} \quad \beta(\chi_w, G(\psi)) = \beta_w(G(\psi)) + o(1).$$

The second criterion is the total error probability, which is defined as follows:

$$\gamma(\chi, G(\psi)) = \eta(\chi) + \beta(\chi, G(\psi)).$$

Table 1
Separation rates $\widetilde{\psi}$ and $b(\psi)$ in the sharp asymptotic bounds where $C(\alpha, L) = (2\alpha+1)(4\alpha+1)^{-(1+\frac{1}{2\alpha})}L^{-\frac{1}{2\alpha}}$.

Σ	$\mathcal{T}(\alpha, L)$	$\mathcal{E}(A, L)$	Not Toeplitz and $\mathcal{T}(\alpha, L)$ [5]
$\widetilde{\psi}$	$\left(C(\alpha, L) \cdot n^2p^2\right)^{-\frac{\alpha}{4\alpha+1}}$	$\left(\frac{2\ln(n^2p^2)}{An^2p^2}\right)^{1/4}$	$\left(C(\alpha, L) \cdot n^2p\right)^{-\frac{\alpha}{4\alpha+1}}$
$b(\psi)^2$	$C(\alpha, L) \cdot \psi^{\frac{4\alpha+1}{\alpha}}$	$\frac{A\psi^4}{2\ln\left(\frac{1}{\psi}\right)}$	$C(\alpha, L) \cdot \psi^{\frac{4\alpha+1}{\alpha}}$

Define also the minimax total error probability γ as $\gamma(G(\psi)) = \inf_\chi \gamma(\chi, G(\psi))$, where the infimum is taken over all possible tests.

Note that the two criteria are related since $\gamma(G(\psi)) = \inf_{w\in(0,1)}(w + \beta_w(G(\psi)))$ (see Ingster and Suslina [23]).

A test χ is asymptotically minimax if: $\gamma(G(\psi)) = \gamma(\chi, G(\psi)) + o(1)$. We say that $\widetilde{\psi}$ is a (asymptotic) separation rate, if the following lower bounds hold

$$\gamma(G(\psi)) \longrightarrow 1 \quad \text{as } \frac{\psi}{\widetilde{\psi}} \longrightarrow 0$$

together with the following upper bounds: there exists a test χ such that,

$$\gamma(\chi, G(\psi)) \longrightarrow 0 \quad \text{as } \frac{\psi}{\widetilde{\psi}} \longrightarrow +\infty.$$

The sharp optimality corresponds to the study of the asymptotic behavior of the maximal type II error probability $\beta_w(G(\psi))$ and the total error probability $\gamma(G(\psi))$. In our study we obtain asymptotic behavior of Gaussian type, i.e. we show that, under some assumptions,

$$\beta_w(G(\psi)) = \Phi(z_{1-w} - npb(\psi)) + o(1) \quad \text{and} \quad \gamma(G(\psi)) = 2\Phi(-npb(\psi)) + o(1), \tag{1.4}$$

where Φ is the cumulative distribution function of a standard Gaussian random variable, z_{1-w} is the $1-w$ quantile of the standard Gaussian distribution for any $w \in (0, 1)$, and $b(\psi)$ has an explicit form for each ellipsoid of Toeplitz covariance matrices.

Separation rates and sharp asymptotic results for different testing problem were studied under this formalism by [22]. We refer for precise definitions of sharp asymptotic and non asymptotic rates to [26]. Note that throughout this paper, asymptotics and symbols o, O, $\sim$ and $\asymp$ are considered as p tends to infinity, unless we specify that n tends to infinity. Recall that, given sequences of real numbers u and real positive numbers v, we say that they are asymptotically equivalent, $u \sim v$, if $\lim u/v = 1$. Moreover, we say that the sequences are asymptotically of the same order, $u \asymp v$, if there exist two constants $0 < c \le C < \infty$ such that $c \le \liminf u/v$ and $\limsup u/v \le C$.

1.2. Overview of the results

In this paper, we describe the separation rates $\widetilde{\psi}$ and sharp asymptotics for the error probabilities for testing the identity matrix against $G(\mathcal{T}(\alpha, L), \psi)$ and $G(\mathcal{E}(A, L), \psi)$ respectively.

We propose here a test procedure whose type II error probability tends to 0 uniformly over the set of $G(\psi)$, that is even for a covariance matrix that gets closer to the identity matrix at distance $\widetilde{\psi} \to 0$ as n and p increase. The radius $\widetilde{\psi}$ in Table 1 is the smallest vicinity around the identity matrix which still allows testing error probabilities to tend to 0. Our test statistic is a weighted quadratic form and we show how to choose these weights in an optimal way over each class of alternative hypotheses.

Under mild assumptions we obtain the sharp optimality in (1.4), where $b(\psi)$ is described in Table 1 and compared to the case of non Toeplitz matrices in [5].

This paper is structured as follows. In Section 2, we study the test problem with alternative hypothesis defined by the class $G(\mathcal{T}(\alpha, L), \psi)$, $\alpha > 1/4$, L, $\psi > 0$. We define explicitly the test statistic and give its first and second moments under the null and the alternative hypotheses. We derive its Gaussian asymptotic behavior under the null hypothesis and under the alternative submitted to the constraints that ψ is close to the separation rate $\widetilde{\psi}$ and that Σ is closed to the solution of an extremal problem Σ^*. We deduce the asymptotic separation rates. Their optimality is shown only for $\alpha > 1$. Our lower bounds are original in the literature of minimax lower bounds, as in this case we cannot reduce the proof to the vector case, or diagonal matrices. We give the sharp rates for $\psi \asymp \widetilde{\psi}$. Our assumptions imply that necessarily $n = o(p^{2\alpha-1/2})$ as $p \to \infty$. That does not prevent n to be larger than p for sufficiently large α.

In Section 3, we derive analogous results over the class $G(\mathcal{E}(A, L), \psi)$, with A, L, $\psi > 0$. We show how to choose the parameters in this case and study the test procedure similarly. We give asymptotic separation rates. The sharp bounds are attained as $\psi \asymp \widetilde{\psi}$. Our assumptions involve that $n = o(\exp(p))$ which allows n to grow exponentially fast with p. That can be explained by the fact that the elements of Σ decay much faster over exponential ellipsoids than over the polynomial ones. In Section 4 we implement our procedure and show the power of testing over two families of covariance matrices.

The proofs of our results are postponed to the Section 5 and to the Supplementary material (see Appendix A).

2. Testing procedure and results for polynomially decreasing covariances

We introduce a weighted U-statistic of order 2, which is an estimator of the functional $\sum_{j\geq 1}\sigma_j^2$ that defines the separation between a Toeplitz covariance matrix under the alternative hypothesis from the identity matrix under the null. Indeed, in nonparametric estimation of quadratic functionals such as $\sum_{j\geq 1}\sigma_j^2$ weighted estimators are often considered (see e.g. [4]). These weights have finite support of length T, where T is optimal in some sense. Intuitively, as the coefficients $\{\sigma_j\}_j$ belong to an ellipsoid, they become smaller when j increases and thus the bias due to the truncation and the weights becomes as small as the variance for estimating the weighted finite sum.

2.1. Test statistic

Let us denote by $T_p(\{\sigma_j\}_{j\geq 1})$ the symmetric $p\times p$ Toeplitz matrix $\Sigma = [\sigma_{lk}]_{1\leq l,k\leq p}$ such that the diagonal elements of Σ are equal to 1, and $\sigma_{lk} = \sigma_{kl} = \sigma_{|l-k|}$, for all $l \neq k$. Now we define the weighted test statistic in this setup

$$\widehat{\mathcal{A}}_n := \widehat{\mathcal{A}}_n^{\mathcal{T}} = \frac{1}{n(n-1)(p-T)^2}\sum_{1\leq k\neq l\leq n}\sum_{j=1}^{T} w_j^* \sum_{T+1\leq i_1,i_2\leq p} X_{k,i_1}X_{k,i_1-j}X_{l,i_2}X_{l,i_2-j} \tag{2.5}$$

where the weights $\{w_j^*\}_j$ and the parameters T, λ, $b^2(\psi)$ are obtained by solving the following extremal problem:

$$b(\psi) := \sum_{j\geq 1} w_j^*\sigma_j^{*2} = \sup_{\left\{\substack{(w_j)_j : w_j\geq 0;\\ \sum_{j\geq 1} w_j^2 = \frac{1}{2}}\right\}} \inf_{\left\{\substack{\Sigma:\Sigma = T_p(\{\sigma_j\}_{j\geq 1});\\ \Sigma\in\mathcal{T}(\alpha,L),\ \sum_{j\geq 1}\sigma_j^2\geq\psi^2}\right\}} \sum_{j\geq 1} w_j\sigma_j^2. \tag{2.6}$$

This extremal problem appears heuristically as we want that the expected value of our test statistic for the worst parameter Σ under the alternative hypothesis (closest to the null) to be as large as possible for the weights we use. This problem will provide the optimal weights $\{w_j^*\}_{j\geq 1}$ in order to control the worst type II error probability, but also the critical matrix $\Sigma^* = T_p(\{\sigma_j^*\})$ that will be used in the lower bounds. Indeed, Σ^* is positive definite for small enough ψ (see [5]).

The solution of the extremal problem (2.6) can be found in [23]:

$$\begin{aligned} &w_j^* = \frac{\lambda}{2b(\psi)}\left(1-\left(\frac{j}{T}\right)^{2\alpha}\right), \qquad \sigma_j^{*2} = \lambda\left(1-\left(\frac{j}{T}\right)^{2\alpha}\right), \qquad T = \left\lfloor (L(4\alpha+1))^{\frac{1}{2\alpha}}\cdot\psi^{-\frac{1}{\alpha}}\right\rfloor \\ &\lambda = \frac{2\alpha+1}{2\alpha(L(4\alpha+1))^{\frac{1}{2\alpha}}}\cdot\psi^{\frac{2\alpha+1}{\alpha}}, \qquad b^2(\psi) = \frac{1}{2}\sum_j \sigma_j^{*4} = \frac{2\alpha+1}{L^{\frac{1}{2\alpha}}(4\alpha+1)^{1+\frac{1}{2\alpha}}}\cdot\psi^{\frac{4\alpha+1}{\alpha}}. \end{aligned} \tag{2.7}$$

Remark that T is a finite number but grows to infinity as $\psi\to 0$. Moreover, the test statistic will have optimality properties under the additional condition that $T/p\to 0$ which is equivalent to $p\psi^{1/\alpha}\to\infty$. It is obvious that in practice it might happen that $T\geq p$ and then we have no solution but to use $T = p-1$, with the inconvenient that the procedure does not behave as well as the theory predicts.

Proposition 1. *Under the null hypothesis, the test statistic $\widehat{\mathcal{A}}_n$ is centered, $\mathbb{E}_I(\widehat{\mathcal{A}}_n) = 0$, with variance:*

$$\mathrm{Var}_I(\widehat{\mathcal{A}}_n) = \frac{1}{n(n-1)(p-T)^2}.$$

Moreover, under the alternative hypothesis with $\alpha > 1/4$, if we assume that $\psi\to 0$ we have:

$$\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) = \sum_{j=1}^{T} w_j^*\sigma_j^2 \geq b(\psi) \quad \text{and} \quad \mathrm{Var}_\Sigma(\widehat{\mathcal{A}}_n) = \frac{R_1}{n(n-1)(p-T)^4} + \frac{R_2}{n(p-T)^2},$$

uniformly over Σ in $G(\mathcal{T}(\alpha,L),\psi)$, where

$$R_1 \leq (p-T)^2\cdot\{1+o(1)+\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)\cdot(O(\sqrt{T})+O(T^{3/2-2\alpha}))+\mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n)\cdot O(T^2)\} \tag{2.8}$$

$$R_2 \leq (p-T)\cdot\{\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)\cdot o(1)+\mathbb{E}_\Sigma^{3/2}(\widehat{\mathcal{A}}_n)\cdot(O(T^{1/4})+O(T^{3/4-\alpha}))+\mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n)\cdot O(T)\}. \tag{2.9}$$

In the next proposition we prove asymptotic normality of the test statistic under the null and under the alternative hypothesis with additional assumptions. More precisely, we need that ψ is of the same order as the separation rate and that the matrix Σ is close to the optimal Σ^*. This is not a drawback, since the asymptotic constant for probability errors are attained under the same assumptions or tend to 0 otherwise.

Proposition 2. *Suppose that $n, p \to +\infty$, $\alpha > 1/4$, $\psi \to 0$, $p\psi^{1/\alpha} \to +\infty$ and moreover assume that $n(p-T)b(\psi) \asymp 1$, the test statistic $\widehat{\mathcal{A}}_n$ defined by (2.5) with parameters given in (2.7), verifies:*

$$n(p-T)\left(\widehat{\mathcal{A}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)\right) \longrightarrow \mathcal{N}(0,1)$$

for all $\Sigma \in G(\mathcal{T}(\alpha, L), \psi)$, such that $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) = O(b(\psi))$.

Moreover, $n(p-T)\widehat{\mathcal{A}}_n$ has asymptotical $\mathcal{N}(0,1)$ distribution under H_0, as $p \to \infty$ for any fixed $n \geq 2$.

2.2. Separation rate and sharp asymptotic optimality

Based on the test statistic $\widehat{\mathcal{A}}_n$, we define the test procedure

$$\chi^* = \chi^*(t) = \mathbb{1}(\widehat{\mathcal{A}}_n > t), \tag{2.10}$$

for conveniently chosen $t > 0$, where $\widehat{\mathcal{A}}_n$ is the estimator defined in (2.5) with parameters in (2.7).

The next theorem gives the separation rate under the assumption that $T = o(p)$, or equivalently, that $p\psi^{1/\alpha} \to \infty$. The upper bounds are attained for arbitrary $\alpha > 1/4$, but the lower bounds require $\alpha > 1$.

Theorem 1. *Suppose that asymptotically*

$$\psi \to 0 \quad \text{and} \quad p\psi^{1/\alpha} \to +\infty. \tag{2.11}$$

***Lower bound**. If $\alpha > 1$ and $n^2p^2\, b^2(\psi) = C(\alpha, L)n^2p^2\, \psi^{\frac{4\alpha+1}{\alpha}} \to 0$ then*

$$\gamma = \inf_\chi \gamma(\chi, G(\mathcal{T}(\alpha, L), \psi)) \longrightarrow 1,$$

where the infimum is taken over all test statistics χ.

***Upper bound**. The test procedure χ^* defined in (2.10) with $t > 0$ has the following properties:*
Type I error probability: if $np \cdot t \to +\infty$ then $\eta(\chi^*) \to 0$.
Type II error probability: if

$$\alpha > 1/4 \quad \text{and} \quad n^2p^2\, b^2(\psi) = C(\alpha, L)n^2p^2\, \psi^{\frac{4\alpha+1}{\alpha}} \to +\infty \tag{2.12}$$

then, uniformly over t such that $t \leq c \cdot C^{1/2}(\alpha, L) \cdot \psi^{\frac{4\alpha+1}{2\alpha}}$, for some constant $0 < c < 1$, we have

$$\beta(\chi^*, G(\mathcal{T}(\alpha, L), \psi)) \longrightarrow 0.$$

Under the assumptions given in (2.11) and (2.12), with t verifying the assumptions of Theorem 1, we get:

$$\gamma(\chi^*, G(\mathcal{T}(\alpha, L), \psi)) \longrightarrow 0.$$

As a consequence of the previous theorem, we get that χ^* is an asymptotically minimax test procedure if $\psi/\widetilde{\psi} \longrightarrow +\infty$. From the lower bounds we deduce that, if $\psi/\widetilde{\psi} \longrightarrow 0$, there is no test procedure to distinguish between the null and the alternative hypotheses, with errors tending to 0. The minimax separation rate $\widetilde{\psi}$ is therefore:

$$\widetilde{\psi} = \left(\frac{2\alpha+1}{L^{\frac{1}{2\alpha}}(4\alpha+1)^{1+\frac{1}{2\alpha}}} \cdot n^2p^2\right)^{-\frac{\alpha}{4\alpha+1}}. \tag{2.13}$$

It is obtained from the relation $n^2p^2b^2(\psi) = 1$. Naturally the constant does not play any role here. Remark that the condition $T/p \to 0 \asymp p\widetilde{\psi}^{1/\alpha} \to +\infty$ implies that $n = o(p^{2\alpha - \frac{1}{2}})$.

The maximal type II error probability either tends to 0, see Theorem 1, or is less than $\Phi(np(t - b(\psi))) + o(1)$ when $npt < npb(\psi) \asymp 1$. The latter case is the object of the next theorem giving sharps bounds for the asymptotic errors. The upper bounds are attained for arbitrary $n \geq 2$ and for $\alpha > 1/4$, while our proof of the sharp lower bounds requires additionally that $n \to \infty$ and $\alpha > 1$.

Theorem 2. *Suppose that $\psi \to 0$ such that $p/T \asymp p\psi^{1/\alpha} \to +\infty$ and, moreover, that*

$$n^2p^2\, b^2(\psi) \asymp 1. \tag{2.14}$$

***Lower bound**. If $\alpha > 1$, then*

$$\inf_{\chi:\eta(\chi)\leq w} \beta(\chi, G(\mathcal{T}(\alpha, L), \psi)) \geq \Phi(z_{1-w} - npb(\psi)) + o(1),$$

where the infimum is taken over all test statistics χ with type I *error probability less than or equal to w. Moreover,*

$$\gamma = \inf_{\chi} \gamma(\chi, G(\mathcal{T}(\alpha, L), \psi)) \geq 2\Phi\left(-np\,\frac{b(\psi)}{2}\right) + o(1).$$

Upper bound. *The test procedure χ^* defined in (2.10) with $t > 0$ has the following properties.*
Type I *error probability: $\eta(\chi^*) = 1 - \Phi(np \cdot t) + o(1)$.*
Type II *error probability: under the assumption (2.14), and for all $\alpha > 1/4$, we have that, uniformly over t:*

$$\beta(\chi^*, G(\mathcal{T}(\alpha, L), \psi)) \leq \Phi(np \cdot (t - b(\psi))) + o(1).$$

In particular, for $t = t^w$, such that $np \cdot t^w = z_{1-w}$, we have $\eta(\chi^*(t^w)) \leq w + o(1)$ and also,

$$\beta(\chi^*(t^w), G(\mathcal{T}(\alpha, L), \psi)) = \Phi(z_{1-w} - np \cdot b(\psi)) + o(1).$$

Another important consequence of the previous theorem, is that the test procedure χ^*, with $t^* = b(\psi)/2$ is such that

$$\gamma(\chi^*(t^*), G(\mathcal{T}(\alpha, L), \psi)) = 2\,\Phi\left(-np\,\frac{b(\psi)}{2}\right) + o(1).$$

Then we can deduce that the minimax separation rate $\widetilde{\psi}$ defined in (2.13) is sharp.

3. Exponentially decreasing covariances

In this section we want to test (1.1) against (1.3), where the alternative set is $G(\mathcal{E}(A, L), \psi)$, for some $A, L, \psi > 0$. It is well known in the nonparametric minimax theory that $\mathcal{E}(A, L)$ is in bijection with ellipsoids of analytic spectral densities admitting analytic continuation on the strip $\{z \in \mathbb{C} : |\mathrm{Im}(z)| \leq A\}$ of the complex plane. On this class nearly parametric rates are attained for testing in the Gaussian noise model, see Ingster [21].

Let us define $\widehat{\mathcal{A}}_n^{\mathcal{E}}$ in (2.5)

$$\widehat{\mathcal{A}}_n^{\mathcal{E}} = \frac{1}{n(n-1)(p-T)^2} \sum_{1 \leq k \neq l \leq n} \sum_{j=1}^{T} w_j^* \sum_{T+1 \leq i_1, i_2 \leq p} X_{k,i_1} X_{k,i_1-j} X_{l,i_2} X_{l,i_2-j}, \tag{3.15}$$

where the weights $\{w_j^*\}_{j \geq 1}$, are obtained by solving the optimization problem (2.6), with the class $\mathcal{T}(\alpha, L)$ replaced by $\mathcal{E}(A, L)$. The solution given in [21] is as follows:

$$\begin{aligned} &w_j^* = \frac{\lambda}{2b(\psi)}\left(1 - \left(\frac{e^j}{e^T}\right)^{2A}\right)_+, \qquad \sigma_j^* = \sqrt{\lambda}\left(1 - \left(\frac{e^j}{e^T}\right)^{2A}\right)_+^{1/2}, \qquad T = \left\lfloor \frac{1}{A} \ln\left(\frac{1}{\psi}\right)\right\rfloor, \\ &\lambda = \frac{A\psi^2}{\ln\left(\frac{1}{\psi}\right)}, \qquad b^2(\psi) = \frac{A\psi^4}{2\ln\left(\frac{1}{\psi}\right)}. \end{aligned} \tag{3.16}$$

Note that all parameters above are free of the radius $L > 0$. Moreover, we have:

$$\sup_j w_j^* \leq \frac{\lambda}{2b(\psi)} \asymp \frac{1}{2(\ln(1/\psi))^{1/2}} \longrightarrow 0.$$

Under the null hypothesis, we still have $\mathbb{E}_I(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = 0$, $\mathrm{Var}_I(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = 1/(n(n-1)(p-T)^2)$ and

$$n(p-T)\widehat{\mathcal{A}}_n^{\mathcal{E}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \quad \text{for fixed } n \geq 2 \text{ and } p \to +\infty.$$

In the following proposition, we see how the upper bounds of the variance have changed under Σ in $G(\mathcal{E}(A, L), \psi)$.

Proposition 3. *Under the alternative, for all $\Sigma \in G(\mathcal{E}(A, L), \psi)$, we have:*

$$\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = \sum_{j=1}^{T} w_j^* \sigma_j^2 \geq b(\psi) \quad \text{and} \quad \mathrm{Var}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = \frac{R_1}{n(n-1)(p-T)^4} + \frac{R_2}{n(p-T)^2}$$

where, for all $A > 0$, and as $\psi \longrightarrow 0$:

$$R_1 \leq (p-T)^2 \cdot \{1 + o(1) + \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot O(\sqrt{T}) + \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot O(T^2)\} \tag{3.17}$$

$$R_2 \leq (p-T) \cdot \{\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot o(1) + \mathbb{E}_\Sigma^{3/2}(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot O(T^{1/4}) + \mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_n^{\mathcal{E}}) \cdot O(T)\}. \tag{3.18}$$

Moreover, if $n(p-T)b(\psi) \asymp 1$, we show that $n(p-T)(\widehat{\mathcal{A}}_n^{\mathcal{E}} - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}})) \to \mathcal{N}(0, 1)$, for all $\Sigma \in \mathcal{E}(A, L)$, such that $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n^{\mathcal{E}}) = O(b(\psi))$.

Now we define the test procedure as follows,

$$\Delta^* = \Delta^*(t) = \mathbb{1}(\widehat{\mathcal{A}}_n^{\mathcal{E}} > t).$$

We describe next the separation rate. We stress the fact that Lemma 2 in the Supplementary material (see Appendix A) shows that the optimal sequence $\{\sigma_j^*\}_j$ in (3.16) provides a Toeplitz positive definite covariance matrix. The sharp results are obtained under the additional assumption that $\psi \asymp \widetilde{\psi}$ and the lower bounds require that n tends also to infinity.

Theorem 3. *Suppose that asymptotically $\psi \to 0$ and $p/T \asymp p/\ln(1/\psi) \to \infty$.*
1. ***Separation rate. Lower bound**: if $n^2p^2b^2(\psi) = n^2p^2 \cdot A\psi^4/(2\ln(1/\psi)) \longrightarrow 0$ then*

$$\gamma = \inf_{\Delta} \gamma(\Delta, G(\psi)) \longrightarrow 1,$$

where the infimum is taken over all test statistics Δ.

***Upper bound**: the test procedure Δ^* defined previously with $t > 0$ has the following properties:*
Type I *error probability: if $np \cdot t \to +\infty$ then $\eta(\Delta^*) \to 0$.*
Type II *error probability: if $n^2p^2\, b^2(\psi) = n^2p^2 \cdot A\psi^4/(2\ln(1/\psi)) \longrightarrow +\infty$ then, uniformly over t such that $t \le c \cdot A^{\frac{1}{2}}\psi^2/(2\ln(1/\psi))^{\frac{1}{2}}$, for some constant c; $0 < c < 1$,*

$$\beta(\Delta^*, G(\psi)) \longrightarrow 0.$$

2. ***Sharp asymptotic bounds. Lower bound**: suppose that $n \to +\infty$ and that*

$$n^2p^2\, b^2(\psi) \asymp 1, \tag{3.19}$$

then we get $\inf_{\Delta:\eta(\Delta)\le w} \beta(\Delta, G(\psi)) \ge \Phi(z_{1-w} - npb(\psi)) + o(1)$, where the infimum is taken over all test statistics Δ with type I error probability less than or equal to w for $w \in (0, 1)$. Moreover,

$$\gamma = \inf_{\Delta} \gamma(\Delta, G(\psi)) \ge 2\Phi\left(-np\,\frac{b(\psi)}{2}\right) + o(1).$$

***Upper bound**: we have*
Type I *error probability: $\eta(\Delta^*) = 1 - \Phi(npt) + o(1)$.*
Type II *error probability: under the condition (3.19), we get that, uniformly over t,*

$$\beta(\,\Delta^*, G(\psi)) \le \Phi(np \cdot (t - b(\psi))) + o(1).$$

In particular, the test procedure $\Delta^*(b(\psi)/2)$, is such that $\gamma(\Delta^*(b(\psi)/2), G(\psi)) = 2\Phi(-np\,\frac{b(\psi)}{2}) + o(1)$. We get the sharp minimax separation rate: $\widetilde{\psi} = \left(\frac{2\ln(n^2p^2)}{An^2p^2}\right)^{1/4}$. Remark that, in this case the condition $T/p \to 0$ implies that $n = o(e^p)$, which is considerably less restrictive than the condition $n = o(p^{2\alpha-\frac{1}{2}})$ of the previous case and allows for exponentially large n, e.g. $n = e^{p/2}$.

4. Numerical implementation and extensions

In this section we implement the test procedure χ in (2.10) with empirically chosen threshold $t > 0$ and study its numerical performance over two families of covariance matrices. We estimate the type I and type II errors by Monte Carlo sampling with 1000 repetitions. First, we choose $\Sigma = \Sigma(M) = [\sigma_j]_j$; $\sigma_j = j^{-2}/M$ under the alternative hypothesis, for various values of $M \in \{2, 2.5, 3, 4, 6, 8, 16, 30, 60, 80\}$. We implement the test statistic $\widehat{\mathcal{A}}_n^{\mathcal{T}}$ defined in (2.5) and (2.7), for parameters $\alpha = 1, L = 1$ and $\psi = \psi(M) = \left(\sum_{j=1}^{p-1} j^{-4}\right)^{\frac{1}{2}}/M$. Our choice of the values for M provides positive definite matrices. We denote by $A(M)$ the random variable $n(p-T)\widehat{\mathcal{A}}_n^{\mathcal{T}}$ when $\Sigma = \Sigma(M)$, and by $A(0)$ when $\Sigma = I$. Note that large values of M give $\Sigma(M)$ with small off-diagonal entries, which is very close to the identity matrix.
Fig. 1, shows that $n(p-T)\widehat{\mathcal{A}}_n^{\mathcal{T}}$ is distributed as a standard normal random variable, when $\Sigma = I$ and $\Sigma(M)$ close enough to the identity. And as a non-centered normal distribution when $\Sigma(M)$ is far from the identity matrix.

To evaluate the performance of our test procedure we compute its power. For each value of n and p, we estimate the 95th percentile t of the distribution of $n(p-T)\widehat{\mathcal{A}}_n^{\mathcal{T}}$ under the null hypothesis $\Sigma = I$. We use t previously defined to estimate the type II error probability, and then plot the associated power. In Fig. 2, we plot the power function of our test procedure χ-test as function of $\psi(M)$, for a fixed value of n and different values of p.

The vertical lines in Fig. 2 represent the different $\tilde{\psi}(n, p)$ associated to different values of p and $n = 10$. We remark that, on the one hand the power grows with $\psi(M)$ for all $p \in \{10, 30, 50, 70\}$. On the other hand the power is an increasing function of p for a fixed covariance matrix $\Sigma(M)$.

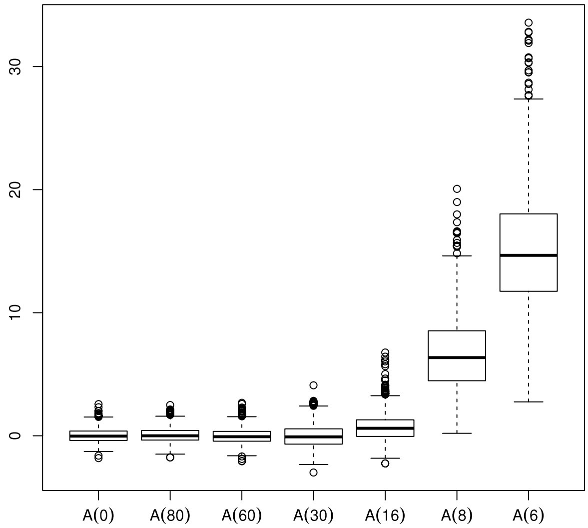

Fig. 1. Distributions of $A(M) = n(p-T)\widehat{\mathcal{A}}_n^{\mathcal{T}}$ for $I = \Sigma(0)$ and $\Sigma = \Sigma(M)$, when $p = 60$ and $n = 40$.

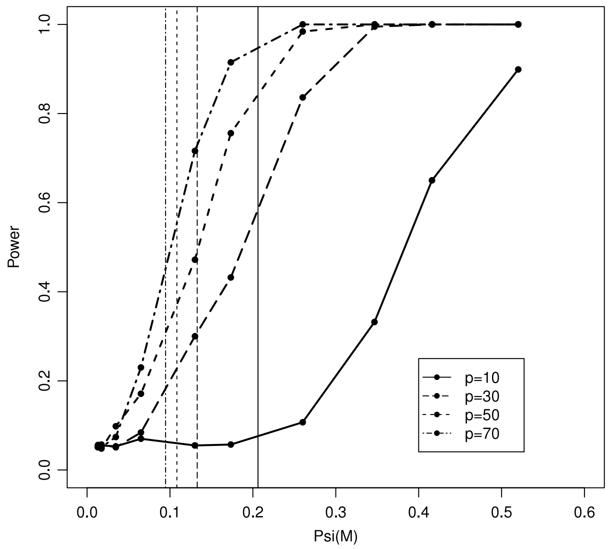

Fig. 2. Power curves of the χ-test as function of $\psi(M)$ for $n = 10$ and $p \in \{10, 30, 50, 70\}$.

We also compare our test procedure with the one defined in [7]. Recall that the test statistic defined by [7] is given by:

$$\widehat{T}_n^{CM} = \frac{2}{n(n-1)} \sum\sum_{1\le k<l\le n} \Big((X_k^\top X_l)^2 - X_k^\top X_k - X_l^\top X_l + p \Big).$$

Note that for matrices $\Sigma \in \mathcal{T}(1, 1)$, we have $(1/p)\|\Sigma - I\|_F^2 \sim \sum_{j=1}^{p-1} \sigma_j^2$, thus we implement $\widehat{T}_n^{CM}/p$ as CM-test statistic. To have fair comparison, we estimate the 95th percentile under the null hypothesis for both tests.
Fig. 3, shows that when n is bigger than or equal to p the powers of the χ-test and the CM-test take close values. While when n is smaller then p, the gap between the power values of the two tests is large, and the χ-test is more powerful than the CM-test.

Second, we consider tridiagonal matrices under the alternative. We define $\Sigma = \Sigma(\rho) = [\sigma_j]_j$; $\sigma_j = \rho \cdot \mathbb{1}\{j = 1\}$, for $\rho \in (0, 1)$. In this case the parameter ψ is $\psi(\rho) = \rho$, for a grid of 10 points ρ belonging to the interval $(0, 0.35]$ and as previously we take $\alpha = 1$ and $L = 1$.
Fig. 4 shows that, the χ-test performs better than the U-test, in the three cases: p smaller than n, p equal n and p larger than n. Moreover, we see that the power curves of the χ-test and the CM-test are closer, when the ratio p/n is smaller. We expect even better results in this particular example if we use a larger value of α, or the procedure defined by (3.15) and (3.16). The question arises of a test statistic free of parameters α, respectively A, which is beyond the scope of this paper.

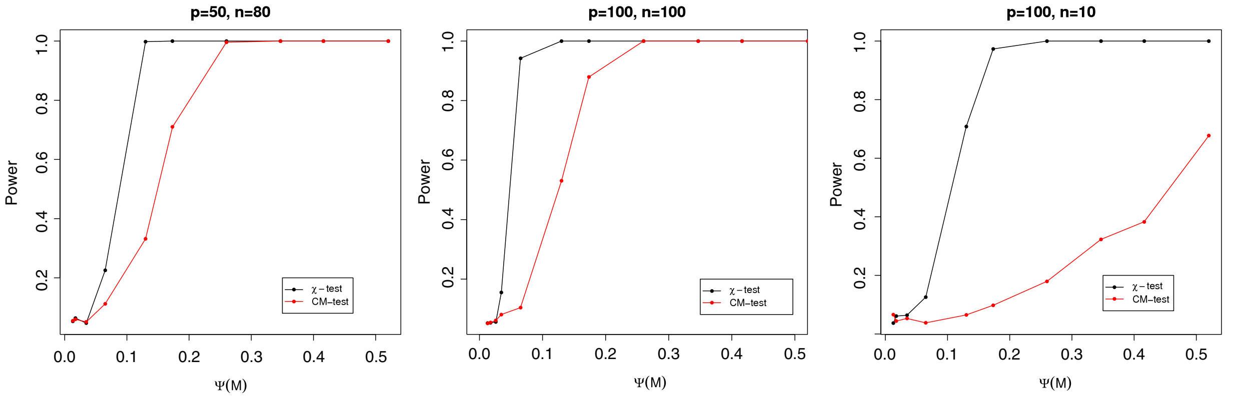

Fig. 3. Power curves of the χ-test and the CM-test as functions of $\psi(M)$, when the alternative consists of matrices whose elements decrease polynomially when moving away from the main diagonal.

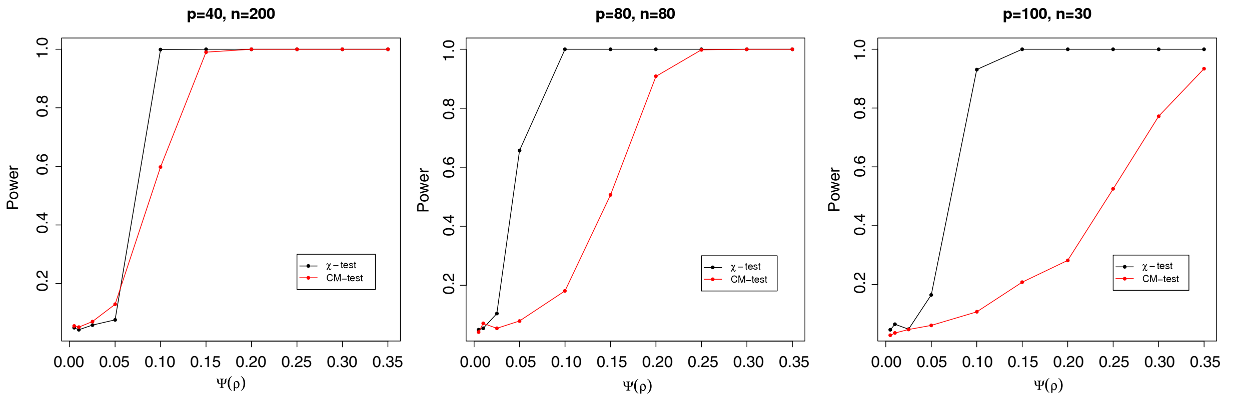

Fig. 4. Power curves of the χ-test and the CM-test as functions of $\psi(\rho)$, when the alternative consists of tridiagonal matrices.

5. Proofs

Proof of Theorems 1 and 2. Recall the assumptions $n,\ p \to +\infty$, $\psi \to 0$ and $T/p \asymp 1/(p\psi^{1/\alpha}) \to 0$.

Lower bounds: In order to show the lower bound, we first reduce the set of parameters to a convenient parametric family. Let $\Sigma^* = T_p(\{\sigma_k^*\}_{k\geq 1})$ be the Toeplitz matrix such that,

$$\sigma_k^* = \sqrt{\lambda}\left(1 - \left(\frac{k}{T}\right)^{2\alpha}\right)_+^{\frac{1}{2}} \quad \text{for } 1 \leq k \leq p-1, \tag{5.20}$$

with λ and T are given by (2.7).

Let us define G^* a subset of $G(\mathcal{T}(\alpha, L), \psi)$ as follows

$$G^* = \{\Sigma_U^* : \Sigma_U^* = T_p(\{u_k\sigma_k\}_{k\geq 1}),\ U \in \mathcal{U}\},$$

where

$$\mathcal{U} = \{U = T_p(\{u_k\}_{k\geq 1}) - I_p \text{ and } u_k = \pm 1 \cdot I(k \leq T-1),\ \text{for } 1 \leq k \leq T-1\}.$$

The cardinality of $\mathcal{U}$ is 2^{T-1}.

From Proposition 3 in [5], we can see that if $\alpha > 1/2$, for all $U \in \mathcal{U}$, the matrix Σ_U^* is positive definite, for $\psi > 0$ small enough. In contrast with [5], we change the signs randomly on each diagonal of the upper triangle of Σ^* and not of all its elements. That allows us to stay into the model of Toeplitz covariance matrices and will actually change the rates of these lower bounds.

Assume that $X_1, \ldots, X_n \sim N(0, I)$ under the null hypothesis and denote by P_I the likelihood of these random variables. Moreover assume that $X_1, \ldots, X_n \sim N(0, \Sigma_U^*)$ under the alternative, and we denote P_U the associated likelihood. In addition let

$$P_\pi = \frac{1}{2^{T-1}} \sum_{U \in \mathcal{U}} P_U$$

be the average likelihood over G^*.

The problem can be reduced to the test $H_0 : X_1, \ldots, X_n \sim P_I$ against the averaged distribution $H_1 : X_1, \ldots, X_n \sim P_\pi$, in the sense that

$$\begin{aligned} \inf_{\chi:\eta(\chi)\le w} \beta(\chi, G(\mathcal{T}(\alpha, L), \psi)) &= \inf_{\chi:\eta(\chi)\le w} \sup_{\Sigma \in G(\mathcal{T}(\alpha,L),\psi)} \mathbb{E}_\Sigma(1-\chi) \ge \inf_{\chi:\eta(\chi)\le w} \sup_{\Sigma\in G^*} \mathbb{E}_\Sigma(1-\chi) \\ &\ge \inf_{\chi:\eta(\chi)\le w} \frac{1}{2^{T-1}} \mathbb{E}_\Sigma(1-\chi) = \inf_{\chi:\eta(\chi)\le w} \mathbb{E}_\pi(1-\chi) := \inf_{\chi:\eta(\chi)\le w} \beta(\chi, \{P_\pi\}) \end{aligned}$$

and that

$$\inf_\chi \gamma(\chi, G(\mathcal{T}(\alpha, L), \psi)) \ge \inf_\chi \gamma(\chi, \{P_\pi\}) + o(1)$$

where, with an abuse of notation, $\beta(\chi, \{P_\pi\}) = \mathbb{E}_\pi(1-\chi)$ and $\gamma(\chi, \{P_\pi\}) = \mathbb{E}_I(\chi) + \mathbb{E}_\pi(1-\chi)$.

It is therefore sufficient to show that, when $u_n \asymp 1$,

$$\inf_{\chi:\eta(\chi)\le w} \beta(\chi, \{P_\pi\}) \ge \Phi(z_{1-w} - npb(\psi)) + o(1) \tag{5.21}$$

and that

$$\inf_\chi \gamma(\chi, \{P_\pi\}) \ge 2\Phi\left(-np\,\frac{b(\psi)}{2}\right) + o(1), \tag{5.22}$$

while, for $u_n = o(1)$, we need that

$$\gamma(\chi, \{P_\pi\}) \to 1. \tag{5.23}$$

Lemma 1. *Assume that $\psi \to 0$ such that $p\psi^{1/\alpha} \to \infty$ and let f_π be the probability density associated to the likelihood P_π previously defined. Then*

$$L_{n,p} := \log \frac{f_\pi}{f_I}(X_1, \ldots, X_n) = u_n Z_n - \frac{u_n^2}{2} + o_P(1), \quad \textit{in } P_I \textit{ probability}, \tag{5.24}$$

where Z_n is asymptotically distributed as a standard Gaussian distribution and $u_n = npb(\psi)$ is such that either $u_n \to 0$ or $u_n \asymp 1$. Moreover, $L_{n,p}$ is uniformly integrable.

In order to obtain (5.21) and (5.22), we apply results in Section 4.3.1 of [23] giving the sufficient condition is (5.24).

It is known that $\gamma(\chi, \{P_\pi\}) = 1 - \frac{1}{2}\|P_I - P_\pi\|_1$ and we bound the L_1 norm by the Kullback–Leibler divergence

$$\frac{1}{2}\|P_I - P_\pi\|_1^2 \le K(P_I, P_\pi).$$

Therefore to show (5.23), we apply Lemma 1 to see that the log likelihood $\log f_\pi/f_I(X_1, \ldots, X_n)$ is an uniformly integrable sequence. This implies that $K(P_I, P_\pi) = \mathbb{E}_I(\log f_\pi/f_I(X_1, \ldots, X_n)) \to 0$. ■

Upper bounds: By Proposition 1, we have that under the null hypothesis $n(p-T)\widehat{\mathcal{A}}_n \to \mathcal{N}(0, 1)$. Then we can deduce that the Type I error probability of χ^* has the following form:

$$\eta(\chi^*) = \mathbb{P}(\widehat{\mathcal{A}}_n > t) = 1 - \Phi(npt) + o(1).$$

For the Type II error probability of χ^*, we shall distinguish two cases, when $n^2p^2b^2(\psi)$ tends to infinity or is bounded by some finite constant. First, assume that $\psi/\widetilde{\psi} \to +\infty$ or, equivalently, that $n^2p^2b^2(\psi) \to +\infty$. Then by the Markov inequality,

$$\mathbb{P}_\Sigma(\widehat{\mathcal{A}}_n \le t) \le \mathbb{P}_\Sigma(|\widehat{\mathcal{A}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)| \ge \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) - t) \le \frac{\mathrm{Var}_\Sigma(\widehat{\mathcal{A}}_n)}{(\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) - t)^2}$$

for all $\Sigma \in G(\mathcal{T}(\alpha, L), \psi)$ and $t \le c \cdot b(\psi)$ such that $0 < c < 1$. Recall that under the alternative, we have $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \ge b(\psi)$ which gives:

$$\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) - t \ge (1-c)\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) \ge (1-c)b(\psi). \tag{5.25}$$

Therefore from the first part of the inequality (5.25) and the variance expression of $\widehat{\mathcal{A}}_n$ under H_1, given in Proposition 1, we have:

$$\mathbb{P}_\Sigma(\widehat{\mathcal{A}}_n \le t) \le \frac{R_1}{n(n-1)(p-T)^4(1-c)^2\mathbb{E}^2_\Sigma(\widehat{\mathcal{A}}_n)} + \frac{R_2}{n(p-T)^2(1-c)^2\mathbb{E}^2_\Sigma(\widehat{\mathcal{A}}_n)} := U_1 + U_2.$$

Let us bound from above U_1, using (2.8) and the second part of the inequality (5.25):

$$U_1 \le \frac{1+o(1)}{n(n-1)(p-T)^2(1-c)^2b^2(\psi)} + \frac{O(\sqrt{T}) + O(T^{3/2-2\alpha})}{n(n-1)(p-T)^2 b(\psi)} + \frac{O(T^2)}{n(n-1)(p-T)^2}.$$

We have $T^{(3/2-2\alpha)}b(\psi) \asymp T^2b^2(\psi) \asymp \psi^{4-\frac{1}{\alpha}} = o(1)$, for all $\alpha > 1/4$, which proves that:

$$U_1 \le \frac{1+o(1)}{n(n-1)(p-T)(1-c)^2b^2(\psi)} = o(1).$$

Indeed, $n^2(p-T)^2b^2(\psi) \to +\infty$, since $n^2p^2b^2(\psi) \to +\infty$ and $T/p \to 0$.
We can check using (2.9) that the term U_2 tends to zero as well:

$$\begin{aligned} U_2 &\le \frac{o(1)}{n(p-T)b(\psi)} + \frac{O(T^{1/4}) + O(T^{3/4-\alpha})}{n(p-T)b^{1/2}(\psi)} + \frac{O(T)}{n(p-T)} \\ &= o(1) \quad \text{for all } \alpha > 1/4, \text{ as soon as } n^2p^2b^2(\psi) \longrightarrow +\infty. \end{aligned}$$

Finally, when ψ is of the same order of the separation rate, i.e. $n^2p^2b^2(\psi) \asymp 1$, we may have either $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n)/b(\psi)$ tends to infinity, or $\mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n) = O(b(\psi))$. In the first case it is easy to see that $U_1 + U_2 \longrightarrow 0$. In the latter the Proposition 2 gives the asymptotic normality of $n(p-T)(\widehat{\mathcal{A}}_n - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n))$. Thereby,

$$\begin{aligned} \sup_{\Sigma \in G(\mathcal{T}(\alpha,L),\psi)} \mathbb{P}_\Sigma(\widehat{\mathcal{A}}_n \le t) &\le \sup_{\Sigma \in G(\mathcal{T}(\alpha,L),\psi)} \Phi(np \cdot (t - \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n))) + o(1) \\ &\le \Phi(np \cdot (t - \inf_{\Sigma \in G(\mathcal{T}(\alpha,L),\psi)} \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_n))) + o(1) \\ &= \Phi(np \cdot (t - b(\psi))) + o(1). \end{aligned}$$

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jmva.2015.09.003.

References

[1] G. Aneiros, Ph. Vieu, Variable selection in infinite-dimensional problems, Statist. Probab. Lett. 94 (2014) 12–20.
[2] Z. Bai, D. Jiang, J.-F. Yao, S. Zheng, Corrections to lrt on large-dimensional covariance matrix by RMT, Ann. Statist. 37 (6B) (2009) 3822–3840. 12.
[3] E.G. Bongiorno, E. Salinelli, A. Goia, Ph. Vieu, Contributions in Infinite-Dimensional Statistics and Related Topics, Esculapio, 2014.
[4] C. Butucea, K. Meziani, Quadratic functional estimation in inverse problems, Stat. Methodol. 8 (1) (2011) 31–41.
[5] C. Butucea, R. Zgheib, Sharp minimax tests for large covariance matrices, ArXiv e-prints, 2014.
[6] T.T. Cai, T. Jiang, Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices, Ann. Statist. 39 (3) (2011) 1496–1525.
[7] T.T. Cai, Z. Ma, Optimal hypothesis testing for high dimensional covariance matrices, Bernoulli 19 (5B) (2013) 2359–2388. 11.
[8] T. Cai, Z. Ren, H. Zhou, Optimal rates of convergence for estimating toeplitz covariance matrices, Probab. Theory Related Fields 156 (2013) 101–143.
[9] K. Chen, K. Chen, H.-G. Müller, J.-L. Wang, Stringing high-dimensional data for functional analysis, J. Amer. Statist. Assoc. 106 (493) (2011) 275–284.
[10] S.X. Chen, L.-X. Zhang, P.-S. Zhong, Tests for high-dimensional covariance matrices, J. Amer. Statist. Assoc. 105 (490) (2010) 810–819.
[11] F. Comte, Adaptive estimation of the spectrum of a stationary Gaussian sequence, Bernoulli 7 (2) (2001) 267–298.
[12] A. Cuevas, A partial overview of the theory of statistics with functional data, J. Statist. Plann. Inference 147 (2014) 1–23.
[13] M.S. Ermakov, A minimax test for hypotheses on a spectral density, J. Math. Sci. 68 (4) (1994) 475–483.
[14] T.J. Fisher, On testing for an identity covariance matrix when the dimensionality equals or exceeds the sample size, J. Statist. Plann. Inference 142 (1) (2012) 312–326.
[15] T.J. Fisher, C.M. Gallagher, New weighted portmanteau statistics for time series goodness of fit testing, J. Amer. Statist. Assoc. 107 (498) (2012) 777–787.
[16] T.J. Fisher, X. Sun, C.M. Gallagher, A new test for sphericity of the covariance matrix for high dimensional data, J. Multivariate Anal. 101 (10) (2010) 2554–2570.
[17] G. Golubev, Nonparametric estimation of smooth spectral densities of Gaussian stationary sequences, Theory Probab. Appl. 38 (4) (1994) 630–639.
[18] Y.K. Golubev, B.Y. Levit, A.B. Tsybakov, Asymptotically efficient estimation of analytic functions in Gaussian noise, Bernoulli 2 (2) (1996) 167–181. 06.
[19] G.K. Golubev, M. Nussbaum, H.H. Zhou, Asymptotic equivalence of spectral density estimation and Gaussian white noise, Ann. Statist. 38 (2010) 181–214.
[20] A.K. Gupta, T. Bodnar, An exact test about the covariance matrix, J. Multivariate Anal. 125 (0) (2014) 176–189.
[21] Yu.I. Ingster, Asymptotically minimax hypothesis testing for nonparametric alternatives. I, Math. Methods Statist. 2 (1993) 85–114. 171–189, 249–268.
[22] Yu.I. Ingster, T. Sapatinas, Minimax goodness-of-fit testing in multivariate nonparametric regression, Math. Methods Statist. 18 (3) (2009) 241–269.
[23] Yu.I. Ingster, I.A. Suslina, Nonparametric Goodness-of-Fit Testing Under Gaussian Models, in: Lecture Notes in Statistics, vol. 169, Springer-Verlag, New York, 2003.
[24] D. Jiang, T. Jiang, F. Yang, Likelihood ratio tests for covariance matrices of high-dimensional normal distributions, J. Statist. Plann. Inference 142 (8) (2012) 2241–2256.
[25] O. Ledoit, M. Wolf, Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size, Ann. Statist. 30 (4) (2002) 1081–1102.

[26] C. Marteau, T. Sapatinas, A unified treatment for non-asymptotic and asymptotic approaches to minimax signal detection, ArXiv e-prints, June 2014.
[27] M.H. Neumann, Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series, J. Time Ser. Anal. 17 (6) (1996) 601–633.
[28] Y. Qiu, S.X. Chen, Test for bandedness of high-dimensional covariance matrices and bandwidth estimation, Ann. Statist. 40 (3) (2012) 1285–1314. 06.
[29] Ph. Soulier, Adaptive estimation of the spectral density of a weakly or strongly dependent Gaussian process, Math. Methods Statist. 10 (3) (2001) 331–354. Meeting on Mathematical Statistics (Marseille, 2000).
[30] M.S. Srivastava, Some tests concerning the covariance matrix in high dimensional data, J. Japan Statist. Soc. 35 (2) (2005) 251–272.
[31] M.S. Srivastava, H. Yanagihara, T. Kubokawa, Tests for covariance matrices in high dimension with less sample size, J. Multivariate Anal. 130 (0) (2014) 289–309.
[32] H. Xiao, W.B. Wu, Asymptotic theory for maximum deviations of sample covariance matrix estimation, Stochastic Process. Appl. 123 (2013) 2899–2920.