



Adaptive test for large covariance matrices in presence of missing observations

Cristina Butucea and Rania Zgheib

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

E-mail address: { cristina.butucea , rania.zgheib } @u-pem.fr

Abstract. We observe n independent p -dimensional Gaussian vectors with missing coordinates, that is each value (which is assumed standardized) is observed with probability $a > 0$. We investigate the problem of minimax nonparametric testing that the high-dimensional covariance matrix Σ of the underlying Gaussian distribution is the identity matrix, using these partially observed vectors. Here, n and p tend to infinity and $a > 0$ tends to 0, asymptotically.

We assume that Σ belongs to a Sobolev-type ellipsoid with parameter $\alpha > 0$. When α is known, we give asymptotically minimax consistent test procedure and find the minimax separation rates $\tilde{\varphi}_{n,p} = (a^2 n \sqrt{p})^{-\frac{2\alpha}{4\alpha+1}}$, under some additional constraints on n , p and a . We show that, in the particular case of Toeplitz covariance matrices, the minimax separation rates are faster, $\tilde{\phi}_{n,p} = (a^2 np)^{-\frac{2\alpha}{4\alpha+1}}$. We note how the "missingness" parameter a deteriorates the rates with respect to the case of fully observed vectors ($a = 1$).

We also propose adaptive test procedures, that is free of the parameter α in some interval, and show that the loss of rate is $(\ln \ln(a^2 n \sqrt{p}))^{\alpha/(4\alpha+1)}$ in general, and $(\ln \ln(a^2 np))^{\alpha/(4\alpha+1)}$ for Toeplitz covariance matrices, respectively.

1. Introduction

Recently, problems related to high-dimensional data became more popular. In particular, in many areas such as genetics, meteorology and others, the generated

Received by the editors February 24th, 2016; accepted April 1st, 2017.

2010 *Mathematics Subject Classification.* 62G10, 62H15.

Key words and phrases. adaptive test, covariance matrices, goodness-of-fit tests, minimax separation rate, missing observation, Toeplitz matrices.

data sets are high-dimensional and incomplete, in the sense that they contain missing values. In this paper we investigate the problem of testing large covariance matrices from a sample of partially observed vectors.

Let X_1, \dots, X_n , be n independent and identically distributed p -vectors following a multivariate normal distribution $\mathcal{N}_p(0, \Sigma)$, where $\Sigma = [\sigma_{ij}]_{1 \leq i, j \leq p}$ is the normalized covariance matrix, with $\sigma_{ii} = 1$, for all $i = 1$ to p . Let us denote $X_k = (X_{k,1}, \dots, X_{k,p})^\top$ for all $k = 1, \dots, n$. Let $\{\varepsilon_{k,j}\}_{1 \leq k \leq n, 1 \leq j \leq p}$ be a sequence of i.i.d. Bernoulli random variables with parameter $a \in (0, 1]$, $\mathcal{B}(a)$, and independent from X_1, \dots, X_n . We observe n i.i.d. random vectors Y_1, \dots, Y_n such that

$$Y_k = (\varepsilon_{k,1} \cdot X_{k,1}, \dots, \varepsilon_{k,p} \cdot X_{k,p})^\top \quad \text{for all } k = 1, \dots, n.$$

Each component of the vector Y_k is observed with probability equal to a and this is the context of missing observations. We denote by $P_{a,\Sigma}$ the probability distribution of the random vector Y_k when $X_k \sim \mathcal{N}_p(0, \Sigma)$ and $\varepsilon_{k,j} \sim \mathcal{B}(a)$. We also denote by $\mathbb{E}_{a,\Sigma}$ and $\text{Var}_{a,\Sigma}$ the expected value and the variance associated to $P_{a,\Sigma}$. Given the partially observed vectors Y_1, \dots, Y_n , we want to test the null hypothesis

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

against a composite alternative hypothesis

$$H_1 : \Sigma \in Q(\mathcal{G}(\alpha), \varphi) \tag{1.2}$$

where $\alpha > 1/2$ and $\mathcal{G}(\alpha)$ is either

$$\mathcal{F}(\alpha) = \left\{ \Sigma > 0, \text{ symmetric}; \right. \\ \left. \lim_{p \rightarrow \infty} \frac{1}{p} \sum_{1 \leq i < j \leq p} \sigma_{ij}^2 |i - j|^{2\alpha} \leq 1 \text{ and } \sigma_{ii} = 1 \text{ for all } 1 \leq i \leq p \right\}$$

in the general case or

$$\mathcal{T}(\alpha) = \left\{ \Sigma > 0, \text{ symmetric, } \Sigma \text{ is Toeplitz}; \sum_{j \geq 1} \sigma_j^2 j^{2\alpha} \leq 1 \text{ and } \sigma_0 = 1 \right\}$$

for the case of Toeplitz matrices. Note that the assumption that second order moments of a stationary series (which is the case for the Toeplitz covariance matrices) exist implies that the associated spectral density is a continuous function. The coefficients σ_k are Fourier coefficients of the spectral density in this case and it is therefore natural to consider Sobolev type ellipsoids with smoothness parameter $\alpha > 1/2$. Thus, we define the following ℓ_2 ellipsoids with ℓ_2 balls removed:

$$Q(\mathcal{F}(\alpha), \varphi) = \left\{ \Sigma \in \mathcal{F}(\alpha) \quad \text{such that } \frac{1}{p} \sum_{1 \leq i < j \leq p} \sigma_{ij}^2 \geq \varphi^2 \right\} \tag{1.3}$$

and

$$Q(\mathcal{T}(\alpha), \varphi) = \left\{ \Sigma \in \mathcal{T}(\alpha) \quad \text{such that } \sum_{j \geq 1} \sigma_j^2 \geq \varphi^2 \right\}. \tag{1.4}$$

Typically, the test procedures depend on the parameter α and it is therefore useful to construct a test procedure that is adaptive to α in some interval. Here we propose minimax and adaptive procedures for testing in the context of missing observations.

From now on, we assume that

$$\varphi \rightarrow 0 \quad \text{and} \quad p \cdot \varphi^{1/\alpha} \rightarrow \infty. \tag{1.5}$$

Let us note that in the case of Toeplitz covariance matrices where $\sigma_{i,j} = \tilde{\sigma}_{|i-j|}$ we have

$$\frac{1}{p} \sum_{1 \leq i < j \leq p} \sigma_{i,j}^2 = \sum_{k=1}^{p-1} \frac{p-k}{p} \tilde{\sigma}_k^2 \geq \varphi^2 - \frac{1}{p} \sum_{k=1}^{p-1} k \left(\frac{k}{p}\right)^{2\alpha-1} \tilde{\sigma}_k^2 \geq \varphi^2 \left(1 - \frac{L}{p^{2\alpha} \varphi^2}\right),$$

leading to the fact that $\frac{1}{p} \|\Sigma - I\|_F^2$ is equivalent to $\sum_{k=1}^p \tilde{\sigma}_k^2$ as $p \rightarrow \infty$ under assumption (1.5). This choice makes the rates easier to compare and discuss heuristically.

The problem of estimating a covariance matrix of partially observed vectors was investigated several times in the literature. The simplest method to deal with missing data is to ignore the missing values and restrict the study to a subset of fully observed variables. This method is not always reliable mainly when the number of missing values is relatively high. Hence, in order to treat this problem, methods based on filling in the missing values were developed, in particular the Expectation-Maximization (EM) algorithm see [Schneider \(2001\)](#). Recently, [Lounici \(2014\)](#) proposed an estimating procedure that does not need imputation of the missing values. Instead, the setup with missing values is treated as an inverse problem. We will also follow this approach for the test problem.

The problem of testing large covariance matrices was considered only in the case of complete data. Out of the large amount of results in the literature on this latter problem, we mention only the most related papers where procedures to test the null hypothesis H_0 in (1.1) are derived. We refer to [Bai et al. \(2009\)](#), [Jiang et al. \(2012\)](#) and [Wang et al. \(2013\)](#), where test procedures based on the likelihood ratio are proposed, and to [Ledoit and Wolf \(2002\)](#), [Srivastava \(2005\)](#), [Chen et al. \(2010\)](#) and [Cai and Ma \(2013\)](#), where test statistics based on the quadratic loss function $\text{tr}(\Sigma - I)^2$ are used. Note that in [Butucea and Zgheib \(2016a\)](#) and [Butucea and Zgheib \(2016b\)](#) asymptotically consistent test procedures were given in order to test (1.1) against (1.2), when the covariance matrices belongs to (1.3) and to (1.4), respectively. They describe the minimax and sharp minimax separation rates. Here, we give the minimax separation rates when assuming that we have partially observed vectors. We describe how the "missingness" parameter a deteriorates the minimax rates in this context. Moreover we develop consistent test procedures free of the class parameter α , via an aggregation procedure of tests.

Missing observations appeared recently in random matrix theory, see [Jurczak and Rohde \(2015\)](#). They show that the sequence of the spectral measures of sample covariance matrices with missing observations converge weakly to a sequence of non random measures. Also they studied the limits of the extremes eigenvalues in the same context.

In this paper, we describe the minimax separation rate for testing H_0 given in (1.1) against the composite alternative H_1 in (1.2), when the data contains missing values. A test procedure Δ is defined as any measurable function of the observations Y_1, \dots, Y_n taking values in $\{0, 1\}$ (we do not consider randomized procedures taking values within the set $[0, 1]$). For a test procedure Δ , we define the type I error probability by $\eta(\Delta) = P_I(\Delta = 1)$, the maximal type II error probability by $\beta(\Delta, Q(\mathcal{G}(\alpha), \varphi)) = \sup_{\Sigma \in Q(\mathcal{G}(\alpha), \varphi)} P_\Sigma(\Delta = 0)$ and the total error probability by

$$\gamma(\Delta, Q(\mathcal{G}(\alpha), \varphi)) = \eta(\Delta) + \beta(\Delta, Q(\mathcal{G}(\alpha), \varphi)).$$

Moreover, we define the minimax total error probability over the class $Q(\mathcal{G}(\alpha), \varphi)$ by

$$\gamma := \inf_{\Delta} \gamma(\Delta, Q(\mathcal{G}(\alpha), \varphi))$$

where the infimum is taken over all possible test procedures. We define the minimax separation rate $\tilde{\varphi}_\alpha$. On the one hand, we construct a test procedure Λ and derive the conditions on φ for which $\gamma(\Lambda, Q(\alpha, \varphi)) \rightarrow 0$. The test Λ will be called asymptotically minimax consistent. On the other hand we give the conditions on φ for which $\gamma \rightarrow 1$. The previous conditions together allow us to determine the minimax separations rate $\tilde{\varphi}_\alpha$, such that there exists the test Λ with

$$\gamma(\Lambda, Q(\mathcal{G}(\alpha), \varphi)) \rightarrow 0 \quad \text{if } \frac{\varphi}{\tilde{\varphi}_\alpha} \rightarrow +\infty,$$

and

$$\gamma = \inf_{\Delta} \gamma(\Delta, Q(\mathcal{G}(\alpha), \varphi)) \rightarrow 1 \quad \text{if } \frac{\varphi}{\tilde{\varphi}_\alpha} \rightarrow 0.$$

In other words, when $\varphi \gg \tilde{\varphi}_\alpha$ there exists an asymptotically minimax consistent test procedure and when $\varphi \ll \tilde{\varphi}_\alpha$, there is no asymptotically consistent test procedure which can distinguish between the null and the alternative hypothesis.

We also consider the problem of adaptation with respect to the parameter α . To treat this problem we first assume that $\alpha \in A$, for A an interval, and define a larger class of matrices under the alternative than (1.2). The testing problem we are interested in now, is to test H_0 in (1.1) against

$$H_1 : \Sigma \in \bigcup_{\alpha \in A} Q(\mathcal{F}(\alpha), \mathcal{C}\psi_\alpha),$$

where $\psi_\alpha = \rho_{n,p}/\tilde{\varphi}_\alpha$, and $\tilde{\varphi}_\alpha$ is the minimax separation rate of testing H_0 given in (1.1) against H_1 in (1.2) for a known α . Our aim is to construct a test procedure Δ_{ad} and to find the loss $\rho_{n,p}$ such that for a large enough constant $\mathcal{C} > 0$:

$$\gamma(\Delta_{ad}, \bigcup_{\alpha \in A} Q(\mathcal{F}(\alpha), \mathcal{C}\psi_\alpha)) \rightarrow 0.$$

In this case we say that Δ_{ad} is an asymptotically adaptive consistent test procedure.

The paper is structured as follows: in section 2 we solve the case of general covariance matrices in $\mathcal{F}(\alpha)$ and in section 3 the particular case of Toeplitz covariance matrices in $\mathcal{T}(\alpha)$. In section 2.1, we study the test problem with alternative hypothesis $Q(\mathcal{F}(\alpha), \varphi)$. We construct an asymptotically minimax consistent test procedure based on the data with missing observations and show that the minimax separation rate is

$$\tilde{\varphi}_\alpha(\mathcal{F}) = (a^2 n \sqrt{p})^{-\frac{2\alpha}{4\alpha+1}}.$$

In section 2.2, we propose a test procedure adaptive to the unknown parameter α . In section 3, we study the problem with alternative hypothesis $Q(\mathcal{T}(\alpha), \varphi)$ and derive analogous results. The minimax separation rate is

$$\tilde{\varphi}_\alpha(\mathcal{T}) = (a^2 np)^{-\frac{2\alpha}{4\alpha+1}}.$$

We can view the vectors X_k in this case as a sample of size p from a stationary Gaussian process. However, due to the missing data, this is not true anymore for vectors Y_k . Recall the assumption (1.5) that $p \cdot \varphi^{1/\alpha} \rightarrow +\infty$. When applied to $\tilde{\varphi}_\alpha(\mathcal{F})$ and to $\tilde{\varphi}_\alpha(\mathcal{T})$, this assumption will restrict the range of parameters n , p and a where our results hold.

Minimax and adaptive rates of testing are faster by a factor \sqrt{p} over classes $\mathcal{T}(\alpha)$ than over the classes $\mathcal{F}(\alpha)$, fact that was already known in the case $a = 1$.

Moreover, note that the rates decrease faster to 0 as p is large. Indeed, the larger are the vectors, more information they bring on the covariance matrix, thus p is not a nuisance parameter.

The adaptive procedure attains the rates $(\sqrt{\ln \ln(a^2 n \sqrt{p})}/(a^2 n \sqrt{p}))^{2\alpha/(4\alpha+1)}$ over \mathcal{F} and $(\sqrt{\ln \ln(a^2 np)}/(a^2 np))^{2\alpha/(4\alpha+1)}$ over \mathcal{T} , respectively. However, the parameter a describing the probability of observing a coordinate appears similarly in both cases. It actually deteriorates the rates with respect to the case $a = 1$ of fully observed data. Proofs are given in section 4.

For the rest of the paper asymptotic values will be taken when $n \rightarrow +\infty$, $p \rightarrow +\infty$ and a is either fix or tends to 0 under further constraints.

2. Test for covariance matrices

We want to test from the data with missing coordinates Y_1, \dots, Y_n the null hypothesis (1.1) against the alternative (1.2) that we recall here:

$$H_1 : \Sigma \in Q(\mathcal{F}(\alpha), \varphi)$$

where $Q(\mathcal{F}(\alpha), \varphi)$ is given in (1.3). This testing problem is treated in Butucea and Zgheib (2016a), for the case of fully observed data, which correspond to $a = 1$ in our case. For the sake of clarity, let us recall that in Butucea and Zgheib (2016a), the following weighted U-statistic was studied

$$\widehat{\mathcal{D}}_{n,p} = \frac{1}{n(n-1)p} \sum_{1 \leq k \neq l \leq n} \sum_{\substack{1 \leq i < j \leq p \\ |i-j| < m}} w_{ij}^* X_{k,i} X_{k,j} X_{l,i} X_{l,j}.$$

The test based on $\widehat{\mathcal{D}}_{n,p}$ was shown to achieve minimax and sharp minimax separation rates, i.e. asymptotic equivalents of the type II error and the total error probabilities are also given when $\varphi \asymp \tilde{\varphi}_\alpha$. The weights $\{w_{ij}^*\}_{1 \leq i < j \leq p}$ depend on the parameter α and are given by

$$w_{ij}^* = \frac{\lambda}{2b(\varphi)} \left(1 - \left(\frac{|i-j|}{T} \right)^{2\alpha} \right)_+$$

where $T = \lfloor C_T(\alpha) \cdot \varphi^{-\frac{1}{\alpha}} \rfloor$, $\lambda = C_\lambda(\alpha) \cdot \varphi^{\frac{2\alpha+1}{\alpha}}$ and $b(\varphi) = C^{\frac{1}{2}}(\alpha) \varphi^{2+\frac{1}{2\alpha}}$ with $C_T(\alpha)$, $C_\lambda(\alpha)$ and $C(\alpha)$ are some explicit constants that depend on α . These parameters are obtained in Butucea and Zgheib (2016a) under the assumptions: $\varphi \rightarrow 0$ and $p\varphi^{1/\alpha} \rightarrow +\infty$, as solutions of an optimization problem.

In the next section we introduce a simpler U-statistic for the case of partially observed vectors and give the asymptotic minimax separation rates, then we aggregate these tests in order to construct a procedure free of the parameter α .

2.1. Test procedure and minimax separation rate. Let us introduce the asymptotically minimax consistent test procedure. In the theory of nonparametric tests developed since Ermakov (1988) and Ingster (1993) it is known that optimal test statistics are obtained from estimators of the functional measuring the separation between the null and the alternative set of hypotheses. In our case, the functional is given by $\frac{1}{p} \|\Sigma - I\|_F^2 = \frac{1}{p} \sum_{1 \leq i < j \leq p} \sigma_{ij}^2$. We introduce now a U-statistic with simpler form than $\widehat{\mathcal{D}}_{n,p}$ defined above. For an integer $m \in \mathbb{N}$ large enough, such that it verifies

$$D \leq m^\alpha \cdot \varphi \leq K^{-2\alpha} \text{ for some constants } D > 1 \text{ and } K > 0, \quad (2.1)$$

we define the following test statistic

$$\widehat{\mathcal{D}}_{n,p,m} = \frac{1}{n(n-1)p} \cdot \frac{1}{\sqrt{2m}} \sum_{1 \leq k \neq l \leq n} \sum_{\substack{1 \leq i < j \leq p \\ |i-j| < m}} Y_{k,i} Y_{k,j} Y_{l,i} Y_{l,j}. \quad (2.2)$$

Note that, as in [Butucea and Zgheib \(2016a\)](#) we only use m diagonals of the sample covariance matrix $\bar{Y}\bar{Y}^\top$, but the weights are constant and equal to $1/\sqrt{2m}$. Indeed, in [Butucea and Zgheib \(2016a\)](#) authors focused on sharp separation rates that were attained by $\widehat{\mathcal{D}}_{n,p}$. The test statistic $\widehat{\mathcal{D}}_{n,p}$ with X 's replaced by Y 's would provide minimax and adaptive separation rates in the setup of missing observations, but the sharp constants would be hard to establish in this context (in particular the lower bounds). Instead, we introduce a less cumbersome test statistic $\widehat{\mathcal{D}}_{n,p,m}$ and prove that it attains the minimax separation rates. The same test statistic will be used for building an adaptive test procedure (free of the parameter α) in a convenient way.

Proposition 2.1. *Under the null hypothesis, the test statistic $\widehat{\mathcal{D}}_{n,p,m}$ in (2.2) is a centered random variable with variance $\text{Var}_{a,I}(\widehat{\mathcal{D}}_{n,p,m}) = a^4/(n(n-1)p)$. Moreover,*

$$\frac{n\sqrt{p}}{a^2} \cdot \widehat{\mathcal{D}}_{n,p,m} \rightarrow \mathcal{N}(0, 1) \quad \text{under } P_I\text{-probability}$$

Under the alternative hypothesis, for all $\Sigma \in \mathcal{F}(\alpha)$ with $\alpha > 1/2$,

$$\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) = \frac{a^4}{p \cdot \sqrt{2m}} \sum_{\substack{1 \leq i < j \leq p \\ |i-j| \leq m}} \sigma_{ij}^2 \quad \text{and} \quad \text{Var}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) = \frac{T_1}{n(n-1)p^2} + \frac{T_2}{np^2}$$

where, for $m \rightarrow +\infty$ such that $m/p \rightarrow 0$ and that (2.1) holds,

$$\begin{aligned} T_1 &\leq p \cdot a^4(1 + o(1)) + p \cdot \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) \cdot O(a^2 m \sqrt{m}), \\ T_2 &\leq p \cdot \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) \cdot O(a^2 \sqrt{m}) \\ &\quad + p^{3/2} \left(\mathbb{E}_{a,\Sigma}^{3/2}(\widehat{\mathcal{D}}_{n,p,m}) O(a^2 m^{3/4}) + \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) \cdot o(a^4) \right). \end{aligned}$$

Proof of Proposition 2.1: The proof follows the same steps as the proof of Proposition 1 of [Butucea and Zgheib \(2016a\)](#). We use repeatedly the independence of $(\varepsilon_{k,i})_{k,i}$ and $(X_{k,\cdot})_k$ and obvious properties of the Bernoulli random variables. \square

Now, we propose the following test procedure

$$\Delta_m = \Delta_m(t) = \mathbb{1}(\widehat{\mathcal{D}}_{n,p,m} > t), \quad t > 0 \quad (2.3)$$

where $\widehat{\mathcal{D}}_{n,p,m}$ is the test statistic defined in (2.2).

Theorem 2.2. *Upper bound: let $m \rightarrow +\infty$ such that $m/p \rightarrow 0$ and that (2.1) holds. If $\alpha > 1/2$ and if*

$$\varphi \rightarrow 0 \quad \text{and} \quad a^2 n \sqrt{p} \varphi^{2 + \frac{1}{2\alpha}} \rightarrow +\infty$$

the test procedure defined in (2.3) with $t \leq c \cdot a^4 \varphi^{2 + \frac{1}{2\alpha}}$ for some constant $c < K(1 - D^{-2})/\sqrt{2}$ and with $n\sqrt{p}t/a^2 \rightarrow +\infty$, is asymptotically minimax consistent, that is

$$\gamma(\Delta_m(t), \mathcal{Q}(\mathcal{F}(\alpha), \varphi)) \rightarrow 0.$$

Lower bound: if $\alpha > 1/2$ and if

$$a^2 n \rightarrow +\infty, \quad p = o(1) \cdot (a^2 n)^{4\alpha-1} \quad \text{and} \quad a^2 n \sqrt{p} \varphi^{2+\frac{1}{2\alpha}} \rightarrow 0$$

then

$$\gamma = \inf_{\Delta} \gamma(\Delta, Q(\mathcal{F}(\alpha), \varphi)) \rightarrow 1$$

Proof of Theorem 2.2: The proof is given in section 4. \square

From the previous theorem we deduce that the minimax separation rate is given by:

$$\tilde{\varphi}_\alpha = \left(a^2 n \sqrt{p} \right)^{-\frac{2\alpha}{4\alpha+1}}$$

Thus the separation $\tilde{\varphi}_\alpha$ obtained for the observations with missing values is slower by the a^2 factor than the separation rate obtained in the case of fully observed vectors.

Note that the conditions on t , the threshold of $\Delta_m(t)$ in (2.3), are compatible. Indeed, $a^2/(n\sqrt{p}) \ll c \cdot a^4 \varphi^{2+1/(2\alpha)}$ is equivalent to our assumption that $a^2 n \sqrt{p} \varphi^{2+1/(2\alpha)} \rightarrow \infty$.

2.2. Adaptation. In this section we construct an asymptotically adaptive consistent test procedure Δ_{ad} free of the parameter $\alpha \in A := [\alpha_*, \alpha_{n,p}^*] \subset]1/2, +\infty[$, with $\alpha_{n,p}^* \rightarrow +\infty$ and $\alpha_{n,p}^* = o(1) \ln(a^2 n \sqrt{p})$, to test H_0 given in (1.1) against the large alternative

$$H_1 : \Sigma \in \bigcup_{\alpha \in A} \left\{ \mathcal{F}(\alpha) ; \frac{1}{2} \sum_{i < j} \sigma_{ij}^2 \geq (\mathcal{C} \Phi_\alpha)^2 \right\},$$

where $\mathcal{C} > 0$ is some constant and

$$\Phi_\alpha = \left(\frac{\sqrt{\ln \ln(a^2 n \sqrt{p})}}{a^2 n \sqrt{p}} \right)^{\frac{2\alpha}{4\alpha+1}}.$$

For each $\alpha \in [\alpha_*, \alpha_{n,p}^*]$, there exists $l \in \mathbb{N}^*$ such that

$$2^{l-1} \leq (\Phi_\alpha)^{-\frac{1}{\alpha}} < 2^l, \quad \text{it suffices to take } l \sim \frac{\frac{2}{4\alpha+1} \ln(a^2 n \sqrt{p})}{\ln(2)}.$$

Let $L_*, L^* \in \mathbb{N}^*$ be defined by

$$L_* = \left(\frac{2}{(4\alpha_{n,p}^* + 1) \ln 2} \right) \ln(a^2 n \sqrt{p}) \quad \text{and} \quad L^* = \left(\frac{2}{(4\alpha_* + 1) \ln 2} \right) \ln(a^2 n \sqrt{p}).$$

We see that L_* and L^* tend to infinity. We define the adaptive test procedure as follows

$$\Delta_{ad} = \max_{L_* \leq l \leq L^*} \mathbf{1}(\mathcal{D}_{n,p,2^l} > t_l), \quad (2.4)$$

where $\mathcal{D}_{n,p,2^l}$ is the test statistic defined in (2.2), with m replaced by 2^l .

Theorem 2.3. The test procedure Δ_{ad} defined in (2.4), with $t_l = a^2 \frac{\sqrt{\mathcal{C}^* \ln l}}{n \sqrt{p}}$,

verifies :

Type I error probability : $\eta(\Delta_{ad}) \rightarrow 0$, for $\mathcal{C}^* > 4$.

Type II error probability : if

$$a^2 n \sqrt{p} \rightarrow +\infty, \quad 2^{L^*}/p \rightarrow 0, \quad \ln(a^2 n \sqrt{p})/n \rightarrow 0 \quad \text{and} \quad \mathcal{C}^2 > 1 + 4\sqrt{\mathcal{C}^*}$$

we get

$$\beta(\Delta_{ad}, \cup_{\alpha \in A} Q(\mathcal{F}(\alpha), \mathcal{C}\Phi_\alpha) \rightarrow 0.$$

Note that the condition $2^{L^*}/p \rightarrow 0$ is equivalent to $a^2n \ll p^{2\alpha^*}$.

Proof of Theorem 2.3: The proof of this theorem is similar the the proof of the Theorem 3.4 which is given in section 4. □

3. Toeplitz covariance matrices

In this section we assume that the covariance matrix Σ is Toeplitz. In this case, we are interested to test (1.1) against the following alternative

$$H_1 : \Sigma \in Q(\mathcal{T}(\alpha), \phi) \tag{3.1}$$

where $Q(\mathcal{T}(\alpha), \phi)$ is defined in (1.4) for ϕ instead of φ . This testing problem is treated in Butucea and Zgheib (2016b), for the particular case $a = 1$, where a weighted U-statistic $\hat{\mathcal{A}}_{n,p}$ of order 2 is used to construct an asymptotically consistent test procedure that achieve the sharp separation rates. Similarly to the previous setup, we construct here a simpler test statistic with constant weights and prove that the associated test procedure has rate optimality.

3.1. *Test procedure and separation rates.* Take $m \in \mathbb{N}$ such that $m \rightarrow +\infty$ and m verifies (2.1) for ϕ instead of φ , we define the following test statistic:

$$\hat{\mathcal{A}}_{n,p,m} = \frac{1}{n(n-1)(p-m)^2} \cdot \frac{1}{\sqrt{2m}} \sum_{1 \leq k \neq l \leq n} \sum_{j=1}^m \sum_{m+1 \leq i_1, i_2 \leq p} Y_{k,i_1} Y_{k,i_1-j} Y_{l,i_2} Y_{l,i_2-j} \tag{3.2}$$

The main difference between the two test statistic $\hat{\mathcal{D}}_{n,p,m}$ and $\hat{\mathcal{A}}_{n,p,m}$ is that in this latter we take into consideration the fact that, we have repeated information on the same diagonal elements.

Now, we give bounds on the moments of this test statistic :

Proposition 3.1. *Under the null hypothesis $\hat{\mathcal{A}}_{n,p,m}$ is a centered random variable whose variance is $Var_{a,I}(\hat{\mathcal{A}}_{n,p,m}) = a^4/(n(n-1)(p-m)^2)$. Moreover, we have that $(n(p-m)/a^2) \cdot \hat{\mathcal{A}}_{n,p,m}/ \rightarrow \mathcal{N}(0, 1)$. Under the alternative hypothesis, for all $\Sigma \in \mathcal{T}(\alpha)$,*

$$\mathbb{E}_{a,\Sigma}(\hat{\mathcal{A}}_{n,p,m}) = (a^4/\sqrt{2m}) \sum_{j=1}^m \sigma_j^2$$

$$\text{and } Var_{a,\Sigma}(\hat{\mathcal{A}}_{n,p,m}) = \frac{R_1}{n(n-1)(p-m)^4} + \frac{R_2}{n(p-m)^2},$$

where

$$R_1 \leq (p-m)^2 \cdot \{a^4(1+o(1)) + \mathbb{E}_\Sigma(\hat{\mathcal{A}}_{n,p,m}) \cdot (O(a^2m) + O(a^3m^{3/2-2\alpha})) + \mathbb{E}_\Sigma^2(\hat{\mathcal{A}}_{n,p,m}) \cdot O(m^2)\}$$

$$R_2 \leq (p-m) \cdot \{a^2 \cdot \mathbb{E}_\Sigma(\hat{\mathcal{A}}_{n,p,m}) \cdot o(1) + \mathbb{E}_\Sigma^{3/2}(\hat{\mathcal{A}}_{n,p,m}) \cdot (O(a \cdot m^{1/4}) + O(a^2m^{3/4-\alpha})) + \mathbb{E}_\Sigma^2(\hat{\mathcal{A}}_{n,p,m}) \cdot O(m)\}.$$

It is easy to show that, since m verifies (2.1), we have for all $\Sigma \in \mathcal{T}(\alpha)$

$$\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{A}}_{n,p,m}) \geq a^4 B \cdot \phi^{2+\frac{1}{2\alpha}}$$

where B is given in (4.1).

To test (1.1) against (3.1), we define the following test procedure based on the statistic defined in (3.2) :

$$\Delta_m^T = \Delta_m^T(t) = \mathbb{1}(\widehat{\mathcal{A}}_{n,p,m} > t), \quad t > 0. \quad (3.3)$$

Theorem 3.2. *Upper bound: let $m \rightarrow +\infty$, such that $m/p \rightarrow 0$ and that (2.1) holds. If $\alpha > 1/4$ and if*

$$\phi \rightarrow 0 \text{ and } a^2 np \phi^{2+\frac{1}{2\alpha}} \rightarrow +\infty$$

the test procedure defined in (2.3) with $t \leq \kappa \cdot a^4 \phi^{2+\frac{1}{2\alpha}}$ for some constant $\kappa \leq B$ such that $np t/a^2 \rightarrow +\infty$ is consistent, that is $\gamma(\Delta_m^T(t), Q(\mathcal{T}(\alpha), \phi)) \rightarrow 0$.

Lower bound: if $\alpha > 1/2$ and if

$$a^2 np \rightarrow +\infty \quad \text{and} \quad a^2 np \phi^{2+\frac{1}{2\alpha}} \rightarrow 0$$

then

$$\gamma = \inf_{\Delta} \gamma(\Delta, Q(\mathcal{T}(\alpha), \phi)) \rightarrow 1.$$

The main consequence of Theorem 3.2 is that the separation rate is given as follows :

$$\tilde{\phi} = \left(a^2 np\right)^{-\frac{2\alpha}{4\alpha+1}}$$

Proof of Theorem 3.2: The proof follows the same steps as the proof of Theorem 2.2, we therefore omit it. The most significant difference is that in order to show the lower bound in this case, we consider a sub-class of Toeplitz matrices:

$$Q_T = \{\Sigma_U : [\Sigma_U]_{ij} = \mathbb{1}_{(i=j)} + u_{|i-j|} \sigma \mathbb{1}_{(|i-j| < T)} \text{ for all } 1 \leq i, j \leq p, U \in \mathcal{U}\},$$

where σ and T are defined in (4.2) and where

$$\mathcal{U} = \{U = [u_{|i-j|}]_{1 \leq |i-j| \leq p-1} : u_{ii} = 0, \forall i \text{ and } u_{|i-j|} \pm 1 \cdot \mathbb{1}_{(|i-j| < T)}, \text{ for } i \neq j\}.$$

Indeed, the signs are randomized but constant on each diagonal. We re-write the terms of $L_{n,p}$ taking into consideration the fact that the matrices are Toeplitz see, for example, the proof of lower bound in Butucea and Zgheib (2016b). \square

Remark 3.3. Remark that the conditions on m imply that m is of order of $\phi^{-\frac{1}{\alpha}}$ in the case of Toeplitz covariance matrices and of order of $\varphi^{-\frac{1}{\alpha}}$ in the case of general covariance matrices.

3.2. Adaptation. In this section, it is always assumed that the covariance matrices are Toeplitz. Our goal is to construct a consistent test procedure independent of the parameter $\alpha \in A := [\alpha_*, \alpha_{n,p}^*] \subset]1/4, +\infty[$, such that $\alpha_{n,p}^* \rightarrow +\infty$ and $\alpha_{n,p}^* = o(1) \ln(a^2 np)$, to test H_0 given in (1.1) against the large alternative

$$H_1 : \Sigma \in \bigcup_{\alpha \in A} \left\{ \mathcal{T}(\alpha) ; \sum_{j=1}^{p-1} \sigma_j^2 \geq (\mathcal{C}\psi_\alpha)^2 \right\}, \quad (3.4)$$

where $\mathcal{C} > 0$ is some constant and

$$\psi_\alpha = \left(\frac{\sqrt{\ln \ln(a^2 np)}}{a^2 np} \right)^{\frac{2\alpha}{4\alpha+1}}.$$

First, see that $\forall \alpha \in [\alpha_*, \alpha_{n,p}^*], \exists l \in \mathbb{N}^*$ such that

$$2^{l-1} \leq (\psi_\alpha)^{-\frac{1}{\alpha}} < 2^l, \quad \text{it suffices to take } l \sim \frac{\frac{2}{4\alpha+1} \ln(a^2 np)}{\ln(2)}$$

Let $L_*, L^* \in \mathbb{N}^*$ be defined by

$$L_* = \left(\frac{2}{(4\alpha_{n,p}^* + 1) \ln 2} \right) \ln(a^2 np) \quad \text{and} \quad L^* = \left(\frac{2}{(4\alpha_* + 1) \ln 2} \right) \ln(a^2 np)$$

We aggregate tests for all given values of l from L_* to L^* giving the following test procedure free of the parameter α :

$$\Delta_{ad} = \max_{L_* \leq l \leq L^*} \mathbf{1}(\mathcal{A}_{n,p,2^l} > t_l), \tag{3.5}$$

where $\mathcal{A}_{n,p,2^l}$ is the test statistic defined in (3.2), with m replaced by 2^l .

Theorem 3.4. *The test procedure Δ_{ad} defined in (3.5), with $t_l = a^2 \frac{\sqrt{\mathcal{C}^* \ln l}}{n(p - 2^l)}$, verifies :*

Type I error probability : $\eta(\Delta_{ad}) \rightarrow 0$, for $\mathcal{C}^ > 4$.*

Type II error probability : if

$$a^2 np \rightarrow +\infty, 2^{L^*}/p \rightarrow 0, \ln(a^2 np)/n \rightarrow 0 \quad \text{and} \quad \mathcal{C}^2 \geq 1 + 4\sqrt{\mathcal{C}^*}$$

we get

$$\beta(\Delta_{ad}, \cup_{\alpha \in A} Q(\mathcal{T}(\alpha), \mathcal{C}\psi_\alpha) \rightarrow 0.$$

Proof of Theorem 3.4: The proof is given in section 4. □

Remark that the condition $2^{L^*}/p$ gives that $a^2 n \ll p^{2\alpha_* - \frac{1}{2}}$ and $\ln(a^2 np)/n \rightarrow 0$ implies that $a^2 np \ll e^n$. Together, these conditions are mild as they give $a^2 np \ll \min\{p^{2\alpha_* + \frac{1}{2}}, e^n\}$.

4. Proofs

Proof of Theorem 2.2: Upper bound We use the asymptotic normality of $\widehat{\mathcal{D}}_{n,p,m}$ to show that, the type I error probability

$$\eta(\Delta_m(t)) = P_{a,I}(\Delta_m(t) = 1) = P_{a,I}(\widehat{\mathcal{D}}_{n,p,m} > t) = \Phi\left(-\frac{n\sqrt{pt}}{a^2}\right) + o(1) = o(1)$$

as soon as $n\sqrt{pt}/a^2 \rightarrow +\infty$. In order to control the maximal type II error probability, we use the Markov inequality to get that for all Σ in $Q(\mathcal{F}(\alpha), \varphi)$:

$$\begin{aligned} P_{a,\Sigma}(\Delta_m(t) = 0) &= P_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m} < t) \\ &\leq P_{a,\Sigma}\left(|\widehat{\mathcal{D}}_{n,p,m} - \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m})| < \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) - t\right) \\ &\leq \frac{\text{Var}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m})}{(\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) - t)^2} = \frac{T_1 + (n-1)T_2}{n(n-1)p^2(\mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,m}) - t)^2}, \end{aligned}$$

for t properly chosen. In order to bound the previous quantity uniformly in Σ over $Q(\mathcal{F}(\alpha), \varphi)$ we need to control

$$\inf_{\Sigma \in Q(\mathcal{F}(\alpha), \varphi)} \mathbb{E}_{a, \Sigma}(\widehat{\mathcal{D}}_{n,p,m}) = \inf_{\Sigma \in Q(\mathcal{F}(\alpha), \varphi)} \frac{a^4}{p \cdot \sqrt{2m}} \sum_{\substack{1 \leq i < j \leq p \\ |i-j| < m}} \sigma_{ij}^2.$$

For all $\Sigma \in Q(\mathcal{F}(\alpha), \varphi)$, we have

$$\begin{aligned} \mathbb{E}_{a, \Sigma}(\widehat{\mathcal{D}}_{n,p,m}) &= \frac{a^4}{p \cdot \sqrt{2m}} \sum_{\substack{1 \leq i < j \leq p \\ |i-j| < m}} \sigma_{ij}^2 = \frac{a^4}{p \cdot \sqrt{2m}} \left(\sum_{1 \leq i < j \leq p} \sigma_{ij}^2 - \sum_{\substack{1 \leq i < j \leq p \\ |i-j| \geq m}} \sigma_{ij}^2 \right) \\ &\geq \frac{a^4}{\sqrt{2m}} \left(\varphi^2 - \frac{1}{p} \sum_{i < j} \frac{|i-j|^{2\alpha}}{m^{2\alpha}} \sigma_{ij}^2 \right) \geq \frac{a^4 \varphi^2}{\sqrt{2m}} \left(1 - \frac{1}{m^{2\alpha} \varphi^2} \right). \end{aligned}$$

We use (2.1) to get that, for all $\Sigma \in Q(\mathcal{F}(\alpha), \varphi)$

$$\mathbb{E}_{a, \Sigma}(\widehat{\mathcal{D}}_{n,p,m}) \geq \frac{a^4 K}{\sqrt{2}} \cdot \varphi^{2+\frac{1}{2\alpha}} \left(1 - \frac{1}{D^2} \right) =: a^4 B \cdot \varphi^{2+\frac{1}{2\alpha}}, \tag{4.1}$$

where $B = K(1 - D^{-2})/\sqrt{2}$. Therefore, take $t \leq c \cdot a^4 \varphi^{2+\frac{1}{2\alpha}}$ for $c < B$ and use (4.1) to obtain that

$$\begin{aligned} \frac{T_1}{n(n-1)p^2(\mathbb{E}_{a, \Sigma}(\widehat{\mathcal{D}}_{n,p,m}) - t)^2} &\leq \frac{1 + o(1)}{a^4 n(n-1)p \varphi^{4+\frac{1}{\alpha}} (B-c)^2} \\ &\quad + \frac{a^2 \cdot O(m\sqrt{m})}{a^4 n(n-1)p \varphi^{2+\frac{1}{2\alpha}} (1-c/B)^2} = o(1), \end{aligned}$$

if $a^4 n(n-1)p \varphi^{4+1/\alpha} \rightarrow +\infty$ and for all $\alpha > 1/2$. Indeed, $a^2 \cdot m\sqrt{m} \varphi^{2+1/(2\alpha)} \asymp a^2 \varphi^{2-\frac{1}{\alpha}} = o(1)$. Similarly we show that under the previous conditions the term $T_2/np^2(\mathbb{E}_{a, \Sigma}(\widehat{\mathcal{D}}_{n,p,m}) - t)^2$ tends to 0.

Lower bound To show the lower bound we first restrict the class $Q(\mathcal{F}(\alpha), L)$ to the class

$$Q := \{ \Sigma_U : [\Sigma_U]_{ij} = \mathbb{1}_{(i=j)} + u_{ij} \sigma \mathbb{1}_{(|1 < |i-j| < T)} \text{ for all } 1 \leq i, j \leq p, U \in \mathcal{U} \},$$

where

$$\sigma = \varphi^{1+\frac{1}{2\alpha}}, \quad T \asymp \varphi^{-\frac{1}{\alpha}}, \tag{4.2}$$

and

$$\mathcal{U} = \{ U = [u_{ij}]_{1 \leq i, j \leq p} : u_{ii} = 0, \forall i \text{ and } u_{ij} = u_{ji} = \pm 1 \cdot \mathbb{1}_{(|i-j| < T)}, \text{ for } i \neq j \},$$

Denote by $\varepsilon_k = (\varepsilon_{k,1}, \dots, \varepsilon_{k,p})^\top$ the random vector with i.i.d. entries $\varepsilon_{k,i} \sim \mathcal{B}(a)$, for all $1 \leq k \leq n$. Moreover denote by P_ε and by P_{ε_k} the distributions of $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ and of ε_k , respectively. Recall that the observations Y_1, \dots, Y_n verify $Y_k = \varepsilon_k * X_k$ for all $1 \leq k \leq n$, where $*$ designate the Schur product.

Under the null hypothesis $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(0, I)$, thus the conditional random vectors $Y_k | \varepsilon_k$, are independent Gaussian vectors such that, for all $1 \leq k \leq n$, $Y_k | \varepsilon_k \sim \mathcal{N}(0, I * (\varepsilon_k \varepsilon_k^\top))$. We denote respectively by P_I and by $P_I^{(\varepsilon)}$ the distributions of (Y_1, \dots, Y_n) and of $(Y_1, \dots, Y_n) | (\varepsilon_1, \dots, \varepsilon_n)$ under the null hypothesis. Under the alternative hypothesis, for $X_1, \dots, X_n \sim \mathcal{N}(0, \Sigma_U)$, we get that

the conditional random vectors $Y_k|\varepsilon_k$ are independent Gaussian vectors such that $Y_k|\varepsilon_k \sim \mathcal{N}(0, \Sigma_U * (\varepsilon_k \varepsilon_k^\top))$ for all $1 \leq k \leq n$, where

$$(\Sigma_U * (\varepsilon_k \varepsilon_k^\top))_{ij} = \begin{cases} \varepsilon_{ik} & \text{for } i = j \\ \varepsilon_{ik} \varepsilon_{jk} \cdot \sigma & \text{if } 1 < |i - j| < T \\ 0 & \text{otherwise.} \end{cases}$$

We denote by $P_U = P_{\Sigma_U}$ and $P_U^{(\varepsilon)} = P_{\Sigma_U^{(\varepsilon)}}$ the distributions of (Y_1, \dots, Y_n) and of the conditional distribution $(Y_1, \dots, Y_n)|(\varepsilon_1, \dots, \varepsilon_n)$ respectively, when $X_1, \dots, X_n \sim \mathcal{N}(0, \Sigma_U)$.

We define the average distribution over Q by

$$P_\pi = \frac{1}{2^{p(T-1)/2}} \sum_{U \in \mathcal{U}} P_U.$$

It is known (see [Ingster and Suslina, 2003](#)) that the minimax total error probability satisfies

$$\gamma \geq 1 - \frac{1}{2} \|P_I - P_\pi\|_1$$

In order to prove that $\gamma \rightarrow 1$, we bound from above the L_1 distance by the Kullback-Leibler divergence (see [Tsybakov, 2009](#))

$$\|P_I - P_\pi\|_1^2 \leq \frac{1}{2} \cdot K(P_I, P_\pi), \quad \text{where } K(P_I, P_\pi) := \mathbb{E}_I \log \left(\frac{dP_I}{dP_\pi} \right).$$

Therefore, to complete the proof, it is sufficient to show that $K(P_I, P_\pi) \rightarrow 0$. In order to prove this we use the conditional likelihoods as follows:

$$K(P_I, P_\pi) = \mathbb{E}_I \log \left(\frac{d(P_\varepsilon \otimes P_I^{(\varepsilon)})}{d(P_\varepsilon \otimes P_\pi^{(\varepsilon)})} \right) = \mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \log \left(\frac{dP_I^{(\varepsilon)}}{dP_\pi^{(\varepsilon)}} \right).$$

Let $\varepsilon(w)$ be a realization of ε , we denote by $S_k \in \{1, \dots, p\}$ the support of $\varepsilon_k(w)$ that is $\varepsilon_{k,i}(w) = 1$ if and only if $i \in S_k$. Also we denote by $d_k = \text{Card}(S_k)$, $\Sigma_U^{\varepsilon_k(w)}$ the positive matrix $\in \mathbb{R}^{d_k \times d_k}$, defined as the sub-matrix of Σ_U obtained by removing all the i -th rows and columns corresponding to $i \notin S_k$ and $X_{\varepsilon_k(w)}$ the sub vector of X_k of dimension d_k in which we retain the coordinate with indices in S_k . Thus,

$$\begin{aligned} L((Y_1, \dots, Y_n)|\varepsilon(w)) &:= \mathbb{E}_I^{(\varepsilon(w))} \log \left(\frac{dP_I^{(\varepsilon(w))}}{dP_\pi^{(\varepsilon(w))}} (Y_1, \dots, Y_n) \right) \\ &= \mathbb{E}_I^{(\varepsilon(w))} \left(-\log \mathbb{E}_U \exp \left(-\frac{1}{2} \sum_{k=1}^n \left(X_{\varepsilon_k(w)}^\top ((\Sigma_U^{\varepsilon_k(w)})^{-1} - I_{\varepsilon_k}) X_{\varepsilon_k(w)} \right. \right. \right. \\ &\quad \left. \left. \left. + \log \det(\Sigma_U^{\varepsilon_k(w)}) \right) \right) \right) \end{aligned}$$

Therefore we have

$$\begin{aligned} &\mathbb{E}_\varepsilon \left(L((Y_1, \dots, Y_n)|\varepsilon) \right) \\ &= \mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \left(-\log \mathbb{E}_U \exp \left(-\frac{1}{2} \sum_{k=1}^n \left(X_{\varepsilon_k}^\top ((\Sigma_U^{\varepsilon_k})^{-1} - I_{\varepsilon_k}) X_{\varepsilon_k} + \log \det(\Sigma_U^{\varepsilon_k}) \right) \right) \right) \end{aligned}$$

Denote by

$$L_{n,p} := \log \mathbb{E}_U \exp \left(-\frac{1}{2} \sum_{k=1}^n \left(X_{\varepsilon_k}^\top ((\Sigma_U^{\varepsilon_k})^{-1} - I_{\varepsilon_k}) X_{\varepsilon_k} + \log \det(\Sigma_U^{\varepsilon_k}) \right) \right)$$

We define $\Delta_U^{\varepsilon_k} = \Sigma_U^{\varepsilon_k} - I_{\varepsilon_k}$, for all $U \in \mathcal{U}$ and any realization of ε_k , we have $\text{tr}(\Delta_U^{\varepsilon_k}) = 0$ and $\|\Delta_U^{\varepsilon_k}\| = O(\varphi^{1-\frac{1}{2\alpha}})$ which is $o(1)$, as soon as $\varphi \rightarrow 0$ and $\alpha > 1/2$. In fact, by the Gershgorin's theorem we have

$$\|\Delta_U^{\varepsilon_k}\| \leq \max_{\substack{i \in S_k \\ j \in S_k \\ j \neq i}} |(\Delta_U^{\varepsilon_k})_{ij}| \leq \max_{i \in S_k} \sum_{\substack{j \in S_k \\ 1 < |i-j| < T}} |u_{ij} \sigma| \leq 2T \cdot \sigma = O(\varphi^{1-\frac{1}{2\alpha}}).$$

For all $x \in [-\frac{1}{2}, +\frac{1}{2}]$ we have the following inequalities

$$\begin{aligned} x - x^2 + x^3 - 2x^4 &\leq -\left(\frac{1}{1+x} - 1\right) \leq x - x^2 + x^3 \\ -x + \frac{x^2}{2} - \frac{x^3}{3} &\leq -\log(1+x) \leq -x + \frac{x^2}{2} - \frac{x^3}{3} + \frac{x^4}{2}. \end{aligned}$$

By applying these inequalities to the eigenvalues of $\Delta_U^{\varepsilon_k}$ we get

$$\begin{aligned} \Delta_U^{\varepsilon_k} - (\Delta_U^{\varepsilon_k})^2 + (\Delta_U^{\varepsilon_k})^3 - 2(\Delta_U^{\varepsilon_k})^4 &\leq -((\Sigma_U^{\varepsilon_k})^{-1} - I_{\varepsilon_k}) \leq \Delta_U^{\varepsilon_k} - (\Delta_U^{\varepsilon_k})^2 + (\Delta_U^{\varepsilon_k})^3 \\ \frac{1}{2} \text{tr}(\Delta_U^{\varepsilon_k})^2 - \frac{1}{3} \text{tr}(\Delta_U^{\varepsilon_k})^3 &\leq -\log \det(\Sigma_U^{\varepsilon_k}) \leq \frac{1}{2} \text{tr}(\Delta_U^{\varepsilon_k})^2 - \frac{1}{3} \text{tr}(\Delta_U^{\varepsilon_k})^3 + \frac{1}{2} \text{tr}(\Delta_U^{\varepsilon_k})^4, \end{aligned}$$

for φ small enough such that $\|\Delta_U^{\varepsilon_k}\| \leq 1/2$. Thus we can bound $L_{n,p}$,

$$\begin{aligned} \underline{L}_{n,p} &\leq L_{n,p} \leq \bar{L}_{n,p} \text{ where} \\ \underline{L}_{n,p} &:= \log \mathbb{E}_U \exp \left(\frac{1}{2} \sum_{k=1}^n X_{\varepsilon_k}^\top (\Delta_U^{\varepsilon_k} - (\Delta_U^{\varepsilon_k})^2 + (\Delta_U^{\varepsilon_k})^3 - 2(\Delta_U^{\varepsilon_k})^4) X_{\varepsilon_k} \right. \\ &\quad \left. + \frac{1}{2} \sum_{k=1}^n \left(\frac{1}{2} \text{tr}(\Delta_U^{\varepsilon_k})^2 - \frac{1}{3} \text{tr}(\Delta_U^{\varepsilon_k})^3 \right) \right), \text{ and} \\ \bar{L}_{n,p} &:= \log \mathbb{E}_U \exp \left(\frac{1}{2} \sum_{k=1}^n X_{\varepsilon_k}^\top (\Delta_U^{\varepsilon_k} - (\Delta_U^{\varepsilon_k})^2 + (\Delta_U^{\varepsilon_k})^3) X_{\varepsilon_k} \right. \\ &\quad \left. + \frac{1}{2} \sum_{k=1}^n \left(\frac{1}{2} \text{tr}(\Delta_U^{\varepsilon_k})^2 - \frac{1}{3} \text{tr}(\Delta_U^{\varepsilon_k})^3 + \frac{1}{2} \text{tr}(\Delta_U^{\varepsilon_k})^4 \right) \right). \end{aligned}$$

Now we develop the terms of $\bar{L}_{n,p}$

$$\begin{aligned} \text{tr}(\Delta_U^{\varepsilon_k})^2 &= 2\sigma^2 \sum_{\substack{i < j \\ 1 < |i-j| < T}} \varepsilon_{k,i} \varepsilon_{k,j}, \\ \text{tr}(\Delta_U^{\varepsilon_k})^3 &= 3!\sigma^3 \sum_{\substack{i < i_1 < i_2 \\ 1 < |i-i_1|, |i_1-i_2|, |i_2-i| < T}} u_{ii_1} u_{i_1 i_2} u_{i_2 i} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2}. \end{aligned}$$

and

$$\text{tr}(\Delta_U^{\varepsilon_k})^4 = \sigma^4 \sum_{\substack{i, i_1, i_2, i_3 \\ 1 < |i-i_1|, \dots, |i_3-i| < T}} u_{ii_1} u_{i_1 i_2} u_{i_2 i_3} u_{i_3 i} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \varepsilon_{k,i_3}$$

$$\begin{aligned}
 &= \sigma^4 \sum_{1 < |i-i_1| < T} \varepsilon_{k,i} \varepsilon_{k,i_1} + 2\sigma^4 \sum_{\substack{i_1, i_1, i_2 \\ 1 < |i-i_1|, |i_1-i_2| < T}} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \\
 &+ 4! \sigma^4 \sum_{\substack{i < i_1 < i_2 < i_3 \\ 1 < |i-i_1|, \dots, |i_3-i| < T}} u_{ii_1} u_{i_1 i_2} u_{i_2 i_3} u_{i_3 i} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \varepsilon_{k,i_3}
 \end{aligned}$$

Moreover, we have (recall that $u_{ij}^2 = 1$ and $\varepsilon_{ij}^2 = \varepsilon_{ij}$)

$$\begin{aligned}
 \sum_{k=1}^n X_{\varepsilon_k}^\top \Delta_U^{\varepsilon_k} X_{\varepsilon_k} &= 2\sigma \cdot \sum_{\substack{i < j \\ 1 < |i-j| < T}} u_{ij} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,j} X_{k,i} X_{k,j} \\
 \sum_{k=1}^n X_{\varepsilon_k}^\top (\Delta_U^{\varepsilon_k})^2 X_{\varepsilon_k} &= \sigma^2 \sum_{k=1}^n \sum_{i,j} X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \sum_{\substack{i_1 \\ 1 < |i_1-i|, |i_1-j| < T}} u_{ii_1} u_{i_1 j} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,j} \\
 &= \sigma^2 \sum_{k=1}^n \sum_i X_{\varepsilon_{k,i}}^2 \sum_{\substack{i_1 \\ 1 < |i_1-i| < T}} \varepsilon_{k,i} \varepsilon_{k,i_1} \\
 &+ 2\sigma^2 \sum_{k=1}^n \sum_{i < j} X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \sum_{\substack{i_1 \\ 1 < |i_1-i|, |i_1-j| < T}} u_{ii_1} u_{i_1 j} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,j},
 \end{aligned}$$

and

$$\begin{aligned}
 &\sum_{k=1}^n X_{\varepsilon_k}^\top (\Delta_U^{\varepsilon_k})^3 X_{\varepsilon_k} \\
 &= \sigma^3 \sum_{k=1}^n \sum_{i,j} X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \sum_{\substack{i_1, i_2 \\ 1 < |i-i_1|, |i_1-i_2|, |i_2-j| < T}} u_{ii_1} u_{i_1 i_2} u_{i_2 j} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \varepsilon_{k,j} \\
 &= 2\sigma^3 \sum_{k=1}^n \sum_i X_{\varepsilon_{k,i}}^2 \sum_{\substack{i_1 < i_2 \\ 1 < |i-i_1|, |i_1-i_2|, |i_2-i| < T}} u_{ii_1} u_{i_1 i_2} u_{i_2 i} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \\
 &+ 2\sigma^3 \sum_{k=1}^n \sum_{\substack{i < j \\ 1 < |i-j| < T}} X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \left(u_{ij}^3 \varepsilon_{k,i} \varepsilon_{k,j} + 2 \sum_{\substack{i_1 \\ 1 < |i_1-i| < T}} u_{ij} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,j} \right) \\
 &+ 2\sigma^3 \sum_{k=1}^n \sum_{i < j} X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \sum_{\substack{j \neq i_1 \neq i_2 \neq i \\ 1 < |i-i_1|, |i_1-i_2|, |i_2-j| < T}} u_{ii_1} u_{i_1 i_2} u_{i_2 j} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \varepsilon_{k,j}
 \end{aligned}$$

In consequence, $\bar{L}_{n,p}$ can be written as follows:

$$\begin{aligned}
 &\bar{L}_{n,p} \\
 &= \log \mathbb{E}_U \exp \left\{ \sum_{\substack{i < j \\ 1 < |i-j| < T}} u_{ij} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,j} X_{k,i} X_{k,j} \left(\sigma + \sigma^3 (1 + 2 \sum_{\substack{i_1 \\ 1 < |i_1-i| < T}} \varepsilon_{k,i_1}) \right) \right\}
 \end{aligned}$$

$$\begin{aligned}
& - \sum_{\substack{i, i_1, j \\ i < j \\ 1 < |i_1 - i|, |i_1 - j| < T}} u_{ii_1} u_{i_1 j} \sum_{k=1}^n X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,j} \sigma^2 \\
& + \sum_{\substack{i < i_1 < i_2 \\ 1 < |i - i_1|, |i_1 - i_2|, |i_2 - i| < T}} u_{ii_1} u_{i_1 i_2} u_{i_2 i} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \sigma^3 \left(3X_{\varepsilon_{k,i}}^2 - 1 \right) \\
& + \sum_{\substack{i, i_1, i_2, j \\ i < j \\ j \neq i_1 \neq i_2 \neq i \\ 1 < |i - i_1|, |i_1 - i_2|, |i_2 - j| < T}} u_{ii_1} u_{i_1 i_2} u_{i_2 j} \sum_{k=1}^n X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \varepsilon_{k,j} \sigma^3 \\
& + 6 \left. \sum_{\substack{i < i_1 < i_2 < i_3 \\ 1 < |i - i_1|, \dots, |i_3 - i| < T}} u_{ii_1} u_{i_1 i_2} u_{i_2 i_3} u_{i_3 i} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \varepsilon_{k,i_3} \sigma^4 \right\} \\
& + \frac{\sigma^2}{4} \sum_{\substack{i < j \\ 1 < |i - j| < T}} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,j} - \frac{\sigma^2}{2} \sum_{\substack{i < i_1 \\ 1 < |i_1 - i| < T}} \sum_{k=1}^n X_{\varepsilon_{k,i}}^2 \varepsilon_{k,i} \varepsilon_{k,i_1} \\
& + \frac{\sigma^4}{4} \sum_{\substack{i, i_1 \\ 1 < |i - i_1| < T}} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,i_1} + \frac{\sigma^4}{2} \sum_{\substack{i, i_1, i_2 \\ 1 < |i - i_1|, \dots, |i_2 - i| < T}} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2}
\end{aligned}$$

We have that $\{u_{ij}\}_{1 \leq i < j \leq p}$ is a sequence of i.i.d Rademacher random variables. Note that sequences composed of finite products of i.i.d Rademacher random variables, for example the sequences $\{u_{ir}u_{rj}\}_{1 \leq i \neq r \neq j \leq p}$ and $\{u_{ir}u_{rs}u_{sj}\}_{1 \leq i \neq r \neq s \neq j \leq p}$, form sequences of i.i.d Rademacher random variables. Moreover they are mutually independent and independent from the initial sequence $\{u_{ij}\}_{i < j}$. Now we explicit in $\bar{L}_{n,p}$ the expected value with respect to the i.i.d Rademacher random variables and get

$$\begin{aligned}
\bar{L}_{n,p} & = \sum_{\substack{i < j \\ 1 < |i - j| < T}} \log \cosh \left(\sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,j} X_{k,i} X_{k,j} \left(\sigma + \sigma^3 \left(1 + 2 \sum_{\substack{i_1 \\ 1 < |i_1 - i| < T}} \varepsilon_{k,i_1} \right) \right) \right) \\
& + \frac{\sigma^2}{2} \sum_{\substack{i < j \\ 1 < |i - j| < T}} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,j} - \sigma^2 \sum_{\substack{i < i_1 \\ 1 < |i_1 - i| < T}} \sum_{k=1}^n X_{\varepsilon_{k,i}}^2 \varepsilon_{k,i} \varepsilon_{k,i_1} \\
& + \sum_{\substack{i, i_1, j \\ i < j \\ 1 < |i_1 - i|, |i_1 - j| < T}} \log \cosh \left(\sum_{k=1}^n X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,j} \sigma^2 \right) \\
& + \sum_{\substack{i < i_1 < i_2 \\ 1 < |i - i_1|, |i_1 - i_2|, |i_2 - i| < T}} \log \cosh \left(3\sigma^3 \sum_{k=1}^n X_{\varepsilon_{k,i}}^2 \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} - \sigma^3 \right)
\end{aligned}$$

$$\begin{aligned}
 &+ \sum_{\substack{i, i_1, i_2, j \\ i < j \\ j \neq i_1 \neq i_2 \neq i \\ 1 < |i - i_1|, |i_1 - i_2|, |i_2 - j| < T}} \log \cosh \left(\sum_{k=1}^n X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \varepsilon_{k,j} \sigma^3 \right) \\
 &+ \sum_{\substack{i < i_1 < i_2 < i_3 \\ 1 < |i - i_1|, \dots, |i_3 - i| < T}} \log \cosh \left(6 \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \varepsilon_{k,i_3} \sigma^4 \right) \\
 &+ \frac{\sigma^4}{4} \sum_{\substack{i, i_1 \\ 1 < |i - i_1| < T}} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,i_1} + \frac{\sigma^4}{2} \sum_{\substack{i, i_1, i_2 \\ 1 < |i - i_1|, \dots, |i_2 - i| < T}} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2}.
 \end{aligned}$$

We use the inequality $\frac{x^2}{2} - \frac{x^4}{12} \leq \log \cosh(x) \leq \frac{x^2}{2}$ for all $x \in \mathbb{R}$. Thus,

$$\begin{aligned}
 \bar{L}_{n,p,1} &:= \sum_{\substack{i < j \\ 1 < |i - j| < T}} \log \cosh \left(\sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,j} X_{k,i} X_{k,j} \left(\sigma + \sigma^3 \left(1 + 2 \sum_{\substack{i_1 \\ 1 < |i_1 - i| < T}} \varepsilon_{k,i_1} \right) \right) \right) \\
 &+ \frac{\sigma^2}{2} \sum_{\substack{i < j \\ 1 < |i - j| < T}} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,j} - \sigma^2 \sum_{\substack{i < i_1 \\ 1 < |i_1 - i| < T}} \sum_{k=1}^n X_{\varepsilon_{k,i}}^2 \varepsilon_{k,i} \varepsilon_{k,i_1} \\
 &\leq \frac{1}{2} \sum_{\substack{i < j \\ 1 < |i - j| < T}} \left(\sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,j} X_{k,i} X_{k,j} \left(\sigma + \sigma^3 (1 + 2T) \right) \right)^2 \\
 &+ \frac{\sigma^2}{2} \sum_{\substack{i < j \\ 1 < |i - j| < T}} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,j} - \sigma^2 \sum_{\substack{i < i_1 \\ 1 < |i_1 - i| < T}} \sum_{k=1}^n X_{\varepsilon_{k,i}}^2 \varepsilon_{k,i} \varepsilon_{k,i_1}.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 &\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \left(\bar{L}_{n,p,1} \right) \\
 &\leq \frac{na^2}{2} \sum_{\substack{i < j \\ 1 < |i - j| < T}} \left(\sigma + \sigma^3 (1 + 2T) \right)^2 + \frac{\sigma^2}{2} \sum_{\substack{i < j \\ 1 < |i - j| < T}} na^2 - \sigma^2 \sum_{\substack{i < i_1 \\ 1 < |i_1 - i| < T}} na^2 \\
 &= \frac{na^2}{2} \sum_{\substack{i < j \\ 1 < |i - j| < T}} \left(2\sigma^4 (1 + 2T) + \sigma^6 (1 + 2T)^2 \right) \\
 &= O(na^2 p T^2 \sigma^4) + O(na^2 p T^3 \sigma^6) = O(a^2 n p \varphi^4) + O(a^2 n p \varphi^6) \\
 &= O\left((a^2 n)^{\frac{-4\alpha+1}{4\alpha+1}} p^{\frac{1}{4\alpha+1}} \right) = o(1),
 \end{aligned}$$

as soon as $p = o(1)(a^2 n)^{4\alpha-1}$. Similarly, we show that

$$\begin{aligned}
 &\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \left(\bar{L}_{n,p,1} \right) \\
 &\geq -\frac{1}{12} \mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \sum_{\substack{i < j \\ 1 < |i - j| < T}} \left(\sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,j} X_{k,i} X_{k,j} \left(\sigma + \sigma^3 \left(1 + 2 \sum_{\substack{i_1 \\ 1 < |i_1 - i| < T}} \varepsilon_{k,i_1} \right) \right) \right)^4 + o(1)
 \end{aligned}$$

$$\begin{aligned}
&= -\frac{a^2 n}{12} \cdot \mathbb{E}_I^{(\varepsilon_1)}(X_{k,i}^4 X_{k,j}^4) \sum_{\substack{i < j \\ 1 < |i-j| < T}} \left(\sigma + \sigma^3(1+2T) \right)^4 + o(1) \\
&- \frac{a^4 n(n-1)}{12} \cdot 3\mathbb{E}_I^{(\varepsilon_1)}(X_{1,i}^2 X_{1,j}^2) \mathbb{E}_I^{(\varepsilon_2)}(X_{2,i}^2 X_{2,j}^2) \sum_{\substack{i < j \\ 1 < |i-j| < T}} \left(\sigma + \sigma^3(1+2T) \right)^4.
\end{aligned}$$

See that the first term was already bounded from above in the previous display and that

$$\begin{aligned}
&\frac{a^4 n(n-1)}{12} \cdot 3\mathbb{E}_I^{(\varepsilon_1)}(X_{1,i}^2 X_{1,j}^2) \mathbb{E}_I^{(\varepsilon_2)}(X_{2,i}^2 X_{2,j}^2) \sum_{\substack{i < j \\ 1 < |i-j| < T}} \left(\sigma + \sigma^3(1+2T) \right)^4 \\
&\leq O(a^4 n^2) \cdot \left(pT\sigma^4 + pT^2\sigma^6 + pT^3\sigma^8 + pT^4\sigma^{10} + pT^5\sigma^{12} \right) \\
&= O(a^4 n^2 p\varphi^{4+\frac{1}{\alpha}}) = o(1),
\end{aligned}$$

as soon as $a^4 n^2 p\varphi^{4+\frac{1}{\alpha}} \rightarrow 0$ and $\alpha > 1/2$. We deduce that $\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)}(\bar{L}_{n,p,1}) \geq o(1)$.

As consequence $\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)}(\bar{L}_{n,p,1}) = o(1)$. Now we treat the second term of $\bar{L}_{n,p}$:

$$\begin{aligned}
\bar{L}_{n,p,2} &:= \sum_{\substack{i, i_1, j \\ i < j \\ 1 < |i_1-i|, |i_1-j| < T}} \log \cosh \left(\sum_{k=1}^n X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,j} \sigma^2 \right) \\
&\leq \sum_{\substack{i, i_1, j \\ i < j \\ 1 < |i_1-i|, |i_1-j| < T}} \left(\sum_{k=1}^n X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,j} \sigma^2 \right)^2.
\end{aligned}$$

So,

$$\begin{aligned}
\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)}(\bar{L}_{n,p,2}) &\leq \mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \sum_{\substack{i, i_1, j \\ i < j \\ 1 < |i_1-i|, |i_1-j| < T}} \left(\sum_{k=1}^n X_{\varepsilon_{k,i}} X_{\varepsilon_{k,j}} \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,j} \sigma^2 \right)^2 \\
&= \sum_{\substack{i, i_1, j \\ i < j \\ 1 < |i_1-i|, |i_1-j| < T}} na^3 \sigma^4 \leq a^3 npT\sigma^4 = O(a^3 np\varphi^{4+\frac{1}{\alpha}}) = o(1).
\end{aligned}$$

Using the bound from below of log cosh inequality, we show that $\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)}(\bar{L}_{n,p,2})$ is bounded from below by a quantity that tends to zero. Therefore we get

$$\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)}(\bar{L}_{n,p,2}) = o(1).$$

In a similar way we show that the expected value of the remaining terms with log cosh in $\bar{L}_{n,p}$ tend to 0. Finally we have

$$\begin{aligned}
&\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \left(\frac{\sigma^4}{4} \sum_{\substack{i, i_1 \\ 1 < |i-i_1| < T}} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,i_1} + \frac{\sigma^4}{2} \sum_{\substack{i, i_1, i_2 \\ 1 < |i-i_1|, \dots, |i_2-i| < T}} \sum_{k=1}^n \varepsilon_{k,i} \varepsilon_{k,i_1} \varepsilon_{k,i_2} \right) \\
&= O(a^2 \sigma^4 pTn) + O(a^3 \sigma^4 pT^2n) = O(a^2 np\varphi^{4+\frac{1}{\alpha}}) + O(a^3 np\varphi^4) = o(1),
\end{aligned}$$

under the previous conditions. Consequently, if $a^4 n^2 p \varphi^{4+\frac{1}{\alpha}} = o(1)$ and if $p = o(1)(a^2 n)^{4\alpha-1}$, then

$$\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \left(-\bar{L}_{n,p} \right) = o(1).$$

To achieve the proof, we show in a similar way that $\mathbb{E}_\varepsilon \mathbb{E}_I^{(\varepsilon)} \left(-\underline{L}_{n,p} \right) = o(1)$. \square

Proof of Theorem 3.4: To control the type I error probability, we derive an inequality of Berry-Essen type for $\mathcal{A}_{n,p,2^l}$. For any fixed l in \mathbb{N}^* we denote by $v_{n,p,l} := \text{Var}_{a,I}(\mathcal{A}_{n,p,2^l})$, which gives $v_{n,p,l} \sim a^4/(n^2(p-2^l)^2)$ by Proposition 3.1. Next, we rewrite $\mathcal{A}_{n,p,2^l}$ as follows :

$$\mathcal{A}_{n,p,2^l} = \sum_{1 \leq k < \ell \leq n} H(Y_k, Y_\ell)$$

where,

$$H(Y_k, Y_\ell) = \frac{\sqrt{2}}{n(n-1)(p-2^l)^2} \cdot \frac{1}{\sqrt{2^l}} \sum_{j=1}^{2^l} \sum_{2^l+1 \leq i_1, i_2 \leq p} Y_{k,i_1} Y_{k,i_1-j} Y_{\ell,i_2} Y_{\ell,i_2-j}.$$

For $2 \leq k, h \leq n$, define

$$Z_k = \frac{1}{\sqrt{v_{n,p,l}}} \sum_{\ell=1}^{k-1} H(Y_k, Y_\ell), \quad \text{and} \quad S_h = \sum_{k=2}^h Z_k.$$

Remark that $\{S_h\}_{h \geq 2}$ is a centered martingale with respect to the filtration $\{\mathcal{F}_h\}_{h \geq 2}$ where \mathcal{F}_h is the σ -field generated by the random vectors $\{X_1, \dots, X_h\}$. Note that $\mathcal{A}_{n,p,2^l} = \sqrt{v_{n,p,l}} \cdot S_n$ and let $V_n^2 = \sum_{k=2}^n \mathbb{E}_{a,I}(Z_k^2/\mathcal{F}_{k-1})$. We fix $0 < \delta \leq 1$ and define

$$J_n = \sum_{k=2}^n \mathbb{E}_{a,I}(Z_k)^{2+2\delta} + \mathbb{E}_{a,I}|V_n^2 - 1|^{1+\delta}.$$

We use the Skorokhod representation and Lemma 3.3 in Hall and Heyde (1980) to obtain that, for any $0 < \varepsilon < 1/2$ and any $x \in \mathbb{R}$, there exists a positive constant C depending only on δ such that

$$\begin{aligned} \left| P_{a,I}(\mathcal{A}_{n,p,2^l} \leq x) - \Phi\left(\frac{x}{\sqrt{v_{n,p,l}}}\right) \right| &= \left| P_{a,I}\left(S_n \leq \frac{x}{\sqrt{v_{n,p,l}}}\right) - \Phi\left(\frac{x}{\sqrt{v_{n,p,l}}}\right) \right| \\ &\leq 16\varepsilon^{1/2} \exp\left(-\frac{x^2}{4v_{n,p,l}}\right) + C \cdot \varepsilon^{-1-\delta} J_n. \end{aligned}$$

Then using that $1 - \Phi(u) \leq (1/u) \exp(-u^2/2)$ for all $u > 0$, we obtain

$$\begin{aligned} P_{a,I}(\mathcal{A}_{n,p,2^l} > x) &\leq \left(1 - \Phi\left(\frac{x}{\sqrt{v_{n,p,l}}}\right)\right) + 16\varepsilon^{1/2} \exp\left(-\frac{x^2}{4v_{n,p,l}}\right) + C \cdot \varepsilon^{-1-\delta} J_n \\ &\leq \left(\frac{1}{x/\sqrt{v_{n,p,l}}} + 16\varepsilon^{1/2}\right) \exp\left(-\frac{x^2}{4v_{n,p,l}}\right) + C \cdot \varepsilon^{-1-\delta} J_n. \end{aligned} \tag{4.3}$$

Choose $\delta = 1$, then

$$J_n = \sum_{k=2}^n \mathbb{E}_{a,I}(Z_k)^4 + \mathbb{E}_{a,I}|V_n^2 - 1|^2.$$

We can show that

$$\sum_{k=2}^n \mathbb{E}_{a,I}(Z_k)^4 = O\left(\frac{1}{n}\right) \quad \text{and} \quad \mathbb{E}_{a,I}|V_n^2 - 1|^2 = O\left(\frac{1}{n}\right) + O\left(\frac{1}{2^l}\right) \quad (4.4)$$

Take $t_l = a^2 \frac{\sqrt{\mathcal{C}^* \ln l}}{n(p-2^l)}$, we use (4.3) and (4.4) to bound from above the type I error probability:

$$\begin{aligned} & P_{a,I}(\Delta_{ad} = 1) \\ &= P_{a,I}(\exists l \in \mathbb{N}, L_* \leq l \leq L^*; \mathcal{A}_{n,p,2^l} > t_l) \leq \sum_{L_* \leq l \leq L^*} P_{a,I}(\mathcal{A}_{n,p,2^l} > t_l) \\ &\leq \sum_{L_* \leq l \leq L^*} \left(\left(\frac{a^2}{n(p-2^l)t_l} + 16\varepsilon^{1/2} \right) \exp\left(-\frac{t_l^2}{4v_{n,p}}\right) + \frac{O(1)}{\varepsilon^2} \left(\frac{1}{n} + \frac{1}{2^l}\right) \right) \\ &\leq \sum_{l \geq L_*} \left(\frac{1}{\sqrt{\mathcal{C}^* \ln l}} + 16\varepsilon^{1/2} \right) \exp\left(-\frac{\mathcal{C}^* \ln l}{4}\right) + O(1) \frac{L^* - L_*}{n\varepsilon^2} + \frac{O(1)}{\varepsilon^2} \sum_{l \geq L_*} \frac{1}{2^l} \\ &\leq \sum_{l \geq L_*} \left(\frac{1}{\sqrt{\mathcal{C}^* \ln l}} + 16\varepsilon^{1/2} \right) l^{-\mathcal{C}^*/4} + O(1) \frac{L^*}{n\varepsilon^2} + \frac{O(1)2^{-L_*}}{\varepsilon^2} = o(1), \end{aligned}$$

for $\mathcal{C}^* > 4$ and since L_* and L^* both tend to infinity, such that $\ln(a^2 n \sqrt{p})/n$ tends to 0.

Now, we control the type II error probability. Assume that $\Sigma \in \mathcal{T}(\alpha)$ and that α is such that there exists $l_0 \in \{L_*, \dots, L^*\}$ such that $2^{l_0-1} \leq (\psi_\alpha)^{-\frac{1}{\alpha}} < 2^{l_0}$, thus

$$\begin{aligned} \mathbb{E}_{a,\Sigma}(\widehat{\mathcal{D}}_{n,p,2^{l_0}}) &= \frac{a^4}{\sqrt{2} \cdot 2^{l_0}} \left(\sum_{1 \leq j < p} \sigma_j^2 - \sum_{2^{l_0} < j < p} \sigma_j^2 \right) \\ &\geq \frac{a^4}{2(\psi_\alpha)^{-\frac{1}{\alpha}}} \left(\mathcal{C}^2 \psi_\alpha^2 - \sum_j \frac{j^{2\alpha}}{(2^{l_0})^{2\alpha}} \sigma_j^2 \right) \geq (\psi_\alpha)^{2+\frac{1}{2\alpha}} \cdot \frac{a^4}{2} (\mathcal{C}^2 - 1). \end{aligned}$$

We assumed that $a^2 n p (\psi_\alpha)^{2+\frac{1}{2\alpha}} = \sqrt{\ln \ln(a^2 n p)}$. Moreover, we have

$$t_{l_0} \leq \frac{a^2 \sqrt{\mathcal{C}^* \ln L^*}}{n(p-2^{l_0})} \leq \frac{a^2 \sqrt{\mathcal{C}^* \ln \ln(a^2 n p)}}{n(p-2^{l_0})} \leq 2\sqrt{\mathcal{C}^*} a^4 (\psi_\alpha)^{2+\frac{1}{2\alpha}}.$$

Thus, we have $\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0} \geq a^4 (\psi_\alpha)^{2+1/(2\alpha)} (\mathcal{C}^2 - 1 - 4\sqrt{\mathcal{C}^*})/2$ by our assumption that $\mathcal{C}^2 > 1 + 4\sqrt{\mathcal{C}^*}$. Therefore we get

$$\begin{aligned} P_{a,\Sigma}(\Delta_{ad} = 0) &= P_{a,\Sigma}(\forall l \in \{L_*, \dots, L^*\}; \mathcal{A}_{n,p,2^l} < t_l) \leq P_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}} < t_{l_0}) \\ &\leq P_{a,\Sigma}(|\mathcal{A}_{n,p,2^{l_0}} - \mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})| > \mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0}). \end{aligned}$$

Now, we use Markov inequality and get :

$$\begin{aligned} P_{a,\Sigma}(\Delta_{ad} = 0) &\leq \frac{\text{Var}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})}{(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} \\ &\leq \frac{R_1 + (n-1)(p-2^{l_0})^2 R_2}{n(n-1)(p-2^{l_0})^4 (\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2}. \end{aligned} \quad (4.5)$$

First, we bound from above the first term of (4.5), using Proposition 3.1

$$S_1 := \frac{R_1}{n(n-1)(p-2^{l_0})^4 (\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2}$$

$$\begin{aligned}
 &= \frac{a^4(1 + o(1)) + \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_{n,p,2^{l_0}}) \cdot (O(a^2\sqrt{2^{l_0}}) + O(a^3(2^{l_0})^{3/2-2\alpha}))}{n(n-1)(p-2^{l_0})^2(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} \\
 &+ \frac{\mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_{n,p,2^{l_0}}) \cdot O(m^2/a)}{n(n-1)(p-2^{l_0})^2(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2}
 \end{aligned}$$

We decompose S_1 as sum of three terms: the first one

$$\begin{aligned}
 S_{1,1} &:= \frac{a^4(1 + o(1))}{n(n-1)(p-2^{l_0})^2(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} \\
 &\leq \frac{a^4(1 + o(1))}{n(n-1)(p-2^{l_0})^2 a^8 (\psi_\alpha)^{4+\frac{1}{\alpha}} (\mathcal{C}^2 - 1 - 4\sqrt{\mathcal{C}^*})^2} \\
 &= O\left(\frac{1}{\ln \ln(a^2 np)}\right) = o(1), \text{ as soon as } a^2 np \rightarrow +\infty.
 \end{aligned}$$

Now we show that the second term of S_1 also tends to 0. Recall that $2^{l_0} \asymp (\psi_\alpha)^{-\frac{1}{\alpha}}$, therefore

$$\begin{aligned}
 S_{1,2} &:= \frac{O(a^2 2^{l_0}) + O(a^3(2^{l_0})^{3/2-2\alpha})}{n(n-1)(p-2^{l_0})^2 \mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}})(1 - t_{l_0}/\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}))^2} \\
 &\leq \frac{(O(a^2 2^{l_0}) + O(a^3(2^{l_0})^{3/2-2\alpha}))}{n(n-1)(p-2^{l_0})^2 \mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) \left(1 - \frac{4\sqrt{\mathcal{C}^*}}{\mathcal{C}^2 - 1}\right)^2} \\
 &\leq \frac{O(2^{l_0}) + O(a(2^{l_0})^{3/2-2\alpha})}{n(n-1)(p-2^{l_0})^2 a^2 (\psi_\alpha)^{2+\frac{1}{2\alpha}} \left(1 - \frac{4\sqrt{\mathcal{C}^*}}{\mathcal{C}^2 - 1}\right)^2} \\
 &\leq \frac{O(\sqrt{2^{l_0}} \cdot (\psi_\alpha)^{2+\frac{1}{2\alpha}}) + O((2^{l_0})^{3/2-2\alpha} \cdot (\psi_\alpha)^{2+\frac{1}{2\alpha}})}{\ln \ln(a^2 np)} = o(1).
 \end{aligned}$$

since $2^{l_0} \cdot (\psi_\alpha)^{2+\frac{1}{2\alpha}} \asymp (\psi_\alpha)^{2-\frac{1}{2\alpha}} = o(1)$ and $(2^{l_0})^{3/2-2\alpha} \cdot (\psi_\alpha)^{2+\frac{1}{2\alpha}} \asymp (\psi_\alpha)^{4-\frac{1}{\alpha}} = o(1)$ for all $\alpha > 1/4$. Finally,

$$\begin{aligned}
 S_{1,3} &:= \frac{\mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_{n,p,2^{l_0}}) \cdot O((2^{l_0})^2)}{n(n-1)(p-2^{l_0})^2(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} \\
 &= \frac{O((2^{l_0})^2)}{n(n-1)p^2} = o(1).
 \end{aligned}$$

Now, we bound from above the second term of (4.5).

$$\begin{aligned}
 S_2 &= \frac{R_2}{n(p-2^{l_0})^2(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} \\
 &= \frac{a^2 \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_{n,p,2^{l_0}}) \cdot o(1)}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} + \frac{\mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_{n,p,2^{l_0}}) \cdot O(2^{l_0})}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} \\
 &+ \frac{\mathbb{E}_\Sigma^{3/2}(\widehat{\mathcal{A}}_{n,p,2^{l_0}}) \cdot (O(a \cdot (2^{l_0})^{1/4}) + O(a^2(2^{l_0})^{3/4-\alpha}))}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2}.
 \end{aligned}$$

We bound from above each term of S_2 . For the first term,

$$\begin{aligned} S_{2,1} &:= \frac{a^2 \cdot \mathbb{E}_\Sigma(\widehat{\mathcal{A}}_{n,p,2^{l_0}}) \cdot o(1)}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} \leq \frac{o(1)}{n(p-2^{l_0})a^2(\psi_\alpha)^{2+\frac{1}{2\alpha}}} \\ &= \frac{o(1)}{\sqrt{\ln \ln(a^2 np)}} = o(1). \end{aligned}$$

For the second term we have,

$$S_{2,2} := \frac{\mathbb{E}_\Sigma^2(\widehat{\mathcal{A}}_{n,p,2^{l_0}}) \cdot O(2^{l_0})}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} \leq \frac{O(2^{l_0})}{np} = o(1)$$

Finally for the last term,

$$\begin{aligned} S_{2,3} &:= \frac{\mathbb{E}_\Sigma^{3/2}(\widehat{\mathcal{A}}_{n,p,2^{l_0}}) \cdot (O(a \cdot (2^{l_0})^{1/4}) + O(a^2(2^{l_0})^{3/4-\alpha}))}{n(p-2^{l_0})(\mathbb{E}_{a,\Sigma}(\mathcal{A}_{n,p,2^{l_0}}) - t_{l_0})^2} \\ &\leq \frac{O((2^{l_0})^{1/4}) + O(a \cdot (2^{l_0})^{3/4-\alpha})}{n(p-2^{l_0})a\psi_\alpha^{1+\frac{1}{4\alpha}}} \\ &\leq \frac{O((2^{l_0})^{1/4})}{\sqrt{n(p-2^{l_0})}(\ln \ln(a^2 np))^{\frac{1}{4}}} + \frac{O(a^2 \cdot \psi_\alpha^{1+\frac{1}{4\alpha}} \cdot (2^{l_0})^{3/4-\alpha})}{\sqrt{\ln \ln(a^2 np)}} = o(1), \end{aligned}$$

as $a^2 \cdot \psi^{1+\frac{1}{4\alpha}} \cdot (2^{l_0})^{3/4-\alpha} \asymp \psi_\alpha^{2-\frac{1}{2\alpha}} = o(1)$. □

References

- Z. Bai, D. Jiang, J.-F. Yao and S. Zheng. Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** (6B), 3822–3840 (2009). [MR2572444](#).
- C. Butucea and R. Zgheib. Sharp minimax tests for large covariance matrices and adaptation. *Electron. J. Stat.* **10** (2), 1927–1972 (2016a). [MR3522665](#).
- C. Butucea and R. Zgheib. Sharp minimax tests for large Toeplitz covariance matrices with repeated observations. *J. Multivariate Anal.* **146**, 164–176 (2016b). [MR3477657](#).
- T. T. Cai and Z. Ma. Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** (5B), 2359–2388 (2013). [MR3160557](#).
- S. X. Chen, L.-X. Zhang and P.-S. Zhong. Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** (490), 810–819 (2010). [MR2724863](#).
- M. S. Ermakov. Asymptotically minimax criteria for testing nonparametric hypotheses on the density of a distribution. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **166** (Issled. po Mat. Statist. 8), 44–53, 187 (1988). [MR955253](#).
- P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London (1980). ISBN 0-12-319350-8. [MR624435](#).
- Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Statist.* **2** (4), 249–268 (1993). [MR1259685](#).

- Yu. I. Ingster and I. A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169 of *Lecture Notes in Statistics*. Springer-Verlag, New York (2003). ISBN 0-387-95531-3. [MR1991446](#).
- D. Jiang, T. Jiang and F. Yang. Likelihood ratio tests for covariance matrices of high-dimensional normal distributions. *J. Statist. Plann. Inference* **142** (8), 2241–2256 (2012). [MR2911842](#).
- K. Jurczak and A. Rohde. Spectral analysis of high-dimensional sample covariance matrices with missing observations. *ArXiv Mathematics e-prints* (2015). [arXiv: 1507.01615](#).
- O. Ledoit and M. Wolf. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.* **30** (4), 1081–1102 (2002). [MR1926169](#).
- K. Lounici. High-dimensional covariance matrix estimation with missing observations. *Bernoulli* **20** (3), 1029–1058 (2014). [MR3217437](#).
- T. Schneider. Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Climate* **14** (5), 853–871 (2001). DOI: [10.1175/1520-0442\(2001\)014<0853:AOICDE>2.0.CO;2](#).
- Muni S. Srivastava. Some tests concerning the covariance matrix in high dimensional data. *J. Japan Statist. Soc.* **35** (2), 251–272 (2005). [MR2328427](#).
- A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York (2009). ISBN 978-0-387-79051-0. [MR2724359](#).
- C. Wang, L. Cao and B. Miao. Asymptotic power of likelihood ratio tests for high dimensional data. *ArXiv Mathematics e-prints* (2013). [arXiv: 1302.3302](#).