Journal of Multivariate Analysis 131 (2014) 174–196

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

High-dimensional sparse MANOVA

T. Tony Cai [a], Yin Xia [b,*]

[a] Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, United States

[b] Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, NC 27599, United States

A R T I C L E I N F O

Article history:
Received 1 November 2013
Available online 15 July 2014

AMS subject classifications:
62H15
62G32

Keywords:
Extreme value distribution
High dimensional test
Limiting null distribution
MANOVA
Precision matrix
Testing equality of mean vectors

A B S T R A C T

This paper considers testing the equality of multiple high-dimensional mean vectors under dependency. We propose a test that is based on a linear transformation of the data by the precision matrix which incorporates the dependence structure of the variables. The limiting null distribution of the test statistic is derived and is shown to be the extreme value distribution of type I. The convergence to the limiting distribution is, however, slow when the number of groups is relatively large. An intermediate correction factor is introduced which significantly improves the accuracy of the test. It is shown that the test is particularly powerful against sparse alternatives and enjoys certain optimality. A simulation study is carried out to examine the numerical performance of the test and compare with other tests given in the literature. The numerical results show that the proposed test significantly outperforms those tests against sparse alternatives.

1. Introduction

An interesting testing problem in multivariate analysis is that of testing the equality of K population means $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$, based on K independent random samples, each from a distribution with mean $\boldsymbol{\mu}_i$ and a common covariance matrix $\boldsymbol{\Sigma}$, where $1 \le i \le K$ and $K \ge 2$ is a fixed constant. This testing problem arises in many scientific applications, including genetics, medical imaging and biology. See, for example, [21,14,17]. In the Gaussian setting where one observes $\{\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{in_i}\} \overset{iid}{\sim} N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ for $1 \le i \le K$, the problem can be formulated as testing the hypotheses

$$H_0: \ \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_K \quad \text{versus} \quad H_1: \ \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j \text{ for some } i \neq j.$$

A classical procedure is the likelihood ratio test with the test statistic given by

$$\lambda = \sum_{i=1}^{K} (\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})^T \widehat{\boldsymbol{\Sigma}}_w^{-1} (\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}}), \tag{1}$$

where $\bar{\boldsymbol{X}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{X}_{ij}$, $\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{K}\sum_{j=1}^{n_i} \boldsymbol{X}_{ij}$ with $n = n_1 + \cdots + n_K$ and $\widehat{\boldsymbol{\Sigma}}_w = \sum_{i=1}^{K}\sum_{j=1}^{n_i} (\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i)(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i)^T$ is the within-class sample covariance matrix up to a constant. The likelihood ratio test has been well studied. See, for example, [1].

In many contemporary applications, high dimensional data, whose dimension is often much larger than the sample size, are commonly available. In such a setting, the classical methods which are designed for the low-dimensional case either

* Corresponding author.
E-mail addresses: tcai@wharton.upenn.edu (T.T. Cai), xiayin@email.unc.edu (Y. Xia).

http://dx.doi.org/10.1016/j.jmva.2014.07.002

perform poorly or are no longer applicable. For example, the likelihood ratio test is unsatisfactory when the dimension is high relative to the sample sizes. The two-sample case, i.e. $K = 2$, has been relatively well studied recently in the high-dimensional setting and several alternatives to the likelihood ratio test have been proposed. For example, Bai and Saranadasa [2], Srivastava and Du [20], Srivastava [18], and Chen and Qin [9] proposed tests, which are based on the sum of squares type statistics, that perform well under the dense alternatives where the difference of the two means spreads out. But these tests are known to suffer from low power under the sparse alternatives where the two mean vectors differ only in a small number of coordinates. Cai, Liu and Xia [7] introduced a test, which is based on the maximum type statistic, that is shown to be particularly powerful against sparse alternatives and enjoys certain optimality.

In comparison, the multiple-sample case is much less studied in the high-dimensional setting, although several proposals for correcting the likelihood ratio test have also been introduced. Fujikoshi, Himeno and Wakaki [11] considered the Dempster trace test, which is based on the ratio of the trace of between-class sample covariance matrix $\widehat{\boldsymbol{\Sigma}}_b$ and the trace of the within-class sample covariance matrix $\widehat{\boldsymbol{\Sigma}}_w$, where $\widehat{\boldsymbol{\Sigma}}_b = \sum_{i=1}^K n_i(\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})^T$. Instead of the ratio, Schott [16] proposed a test statistic based on the difference of two traces. Srivastava [19] constructed a test statistic by replacing the inverse of the within-class sample covariance matrix by its Moore–Penrose inverse. All of these test statistics are based on an estimator of $\sum_{1\le i\le K}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})^T \boldsymbol{A}^{(i)}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})$ for some positive definite matrices $\boldsymbol{A}^{(i)}$. We call these sum of squares type statistics as they all aim to estimate the squared Euclidean norm $\sum_{1\le i\le K}\|(\boldsymbol{A}^{(i)})^{\frac{1}{2}}(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})\|_2^2$. In genomics and many other applications, the means of the populations are typically either identical or are quite similar in the sense that they only possibly differ in a small number of coordinates. As in the two-sample case, the above mentioned sum of squares type tests in the multiple-sample case suffer from low power under sparse alternatives.

The goal of the present paper is to develop a test that is powerful against sparse alternatives for multiple samples in the high dimensional setting under dependency. To explore the sparsity in the mean differences and the dependence between the variables, the test is based on the linear transformation of the observations by the precision matrix $\boldsymbol{\Omega}$: $\{\boldsymbol{\Omega}\boldsymbol{X}_{i1}, \ldots, \boldsymbol{\Omega}\boldsymbol{X}_{in_i}\}$, for $1 \le i \le K$. The new test statistic is then defined to be the maximum of the sum of squares of all possible two sample t-statistics of the transformed observations $\{\boldsymbol{\Omega}\boldsymbol{X}_{i1}, \ldots, \boldsymbol{\Omega}\boldsymbol{X}_{in_i}\}$ and $\{\boldsymbol{\Omega}\boldsymbol{X}_{j1}, \ldots, \boldsymbol{\Omega}\boldsymbol{X}_{jn_j}\}$ for $1 \le i < j \le K$. The limiting null distribution of the test statistic is derived and is shown to be the extreme value distribution of type I. The convergence of the distribution of the test statistic under the null to the limiting distribution is, however, slow when the number of groups is relatively large. We further introduced an intermediate correction factor which significantly improves the accuracy of the test. Although the basic idea underlying the construction of the test statistic is similar to the one for the two-sample case in [7], the techniques and the intermediate correction procedure are new and are much more involved than the two-sample case.

Both theoretical and numerical properties of the test are studied. It is shown that the test is particularly powerful against sparse alternatives and enjoys certain optimality. A simulation study is carried out to examine the numerical performance of the test and compare with other tests given in the literature. The numerical results show that the proposed test significantly outperforms those tests against sparse alternatives. We also illustrate the improvement after using the correction factor by comparing its cumulative distribution with the type I extreme value distribution as well as the empirical limiting distribution. The limiting distribution after using the correction is a much better approximation to the empirical distribution, as illustrated in Fig. 2 in Section 3.2. As a direct consequence, numerical results show that the size of the resulting test is close to the nominal level.

The rest of the paper is organized as follows. After reviewing basic notation and definitions, Section 2 introduces the new test statistics. Theoretical properties of the proposed tests are investigated in Section 3. Limiting null distributions of the test statistics and the power of the tests, both for the case the precision matrix $\boldsymbol{\Omega}$ is known and the case $\boldsymbol{\Omega}$ is unknown, are analyzed. A simulation study is carried out in Section 4 to investigate the numerical performance of the tests. Discussions of the results and other related work are given in Section 5. The proofs of main results are presented in Section 6.

2. Methodology

We first construct a testing procedure in the oracle setting in Section 2.1 where the covariance matrix $\boldsymbol{\Sigma}$ is assumed to be known. In addition, another natural testing procedure is introduced in this setting. A data-driven procedure is given in Section 2.2 for the general case of unknown covariance matrix $\boldsymbol{\Sigma}$.

We begin with basic notation and definitions. For a vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{T^T} \in \mathbb{R}^p$, define the ℓ_q norm by $|\boldsymbol{\beta}|_q = (\sum_{i=1}^p |\beta_i|^q)^{1/q}$ for $1 \le q \le \infty$ with the usual modification for $q = \infty$. A vector $\boldsymbol{\beta}$ is called k-sparse if it has at most k nonzero entries. For a matrix $\boldsymbol{A} = (a_{ij})_{p\times p}$, the matrix 1-norm is the maximum absolute column sum, $\|\boldsymbol{A}\|_{L_1} = \max_{1\le j\le p}\sum_{i=1}^p |a_{ij}|$, the matrix elementwise infinity norm is defined to be $|\boldsymbol{A}|_\infty = \max_{1\le i,j\le p}|a_{ij}|$ and the elementwise ℓ_1 norm is $\|\boldsymbol{A}\|_1 = \sum_{i=1}^p\sum_{j=1}^p |a_{ij}|$. For a matrix $\boldsymbol{A}$, we say $\boldsymbol{A}$ is k-sparse if each row/column has at most k nonzero entries. We shall denote $(\sqrt{\frac{n_1 n_2}{n_1+n_2}}(\mu_{1i} - \mu_{2i}), \sqrt{\frac{n_1 n_3}{n_1+n_3}}(\mu_{1i} - \mu_{3i}), \ldots, \sqrt{\frac{n_{K-1} n_K}{n_{K-1}+n_K}}(\mu_{K-1i} - \mu_{Ki}))^T =: \boldsymbol{\delta}_i = (\delta_i^{(12)}, \ldots, \delta_i^{(K-1K)})^T$ so the null hypothesis can be equivalently written as $H_0 : |\boldsymbol{\delta}_i|_2 = 0$ for $i = 1, \ldots, p$. Let $\boldsymbol{\delta}^{(jl)} := (\delta_1^{(jl)}, \ldots, \delta_p^{(jl)})^T = \sqrt{\frac{n_j n_l}{n_j+n_l}}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_l)$, then the alternative is called k-sparse if $\boldsymbol{\delta}^{(jl)}$ is k-sparse for all $1 \le j < l \le K$. For two sequences of real numbers $\{a_n\}$ and $\{b_n\}$, write $a_n = O(b_n)$ if there exists a constant C such that $|a_n| \le C|b_n|$ holds for all sufficiently large n, write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$, and write $a_n \asymp b_n$ if there are positive constants c and C such that $c \le a_n/b_n \le C$ for all $n \ge 1$.

2.1. Oracle procedure

Suppose we observe independent p-dimensional random samples

$$\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n_1} \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}), \boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{2n_2} \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}), \ldots, \boldsymbol{X}_{K1}, \ldots \boldsymbol{X}_{Kn_K} \sim N(\boldsymbol{\mu}_K, \boldsymbol{\Sigma}),$$

where the covariance matrix $\boldsymbol{\Sigma} := (\sigma_{ij})$ is known. In this case, the null hypothesis $H_0 : |\boldsymbol{\delta}_i|_2 = 0$, for $i = 1, \ldots, p$, is equivalent to $H_0 : |\boldsymbol{\eta}_i|_2 = 0$, for $i = 1, \ldots, p$, where $\boldsymbol{\eta}_i = ((\boldsymbol{A}\delta^{(12)})_i, \ldots, (\boldsymbol{A}\delta^{(K-1K)})_i)^T$ for any $p \times p$ positive definite matrix $\boldsymbol{A} := (a_{ij})$. An unbiased estimator of $\boldsymbol{\eta}_i$ is the sample mean vector $(\sqrt{\frac{n_1 n_2}{n_1+n_2}}(\boldsymbol{A}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2))_i, \ldots, \sqrt{\frac{n_{K-1}n_K}{n_{K-1}+n_K}}(\boldsymbol{A}(\bar{\boldsymbol{X}}_{K-1} - \bar{\boldsymbol{X}}_K))_i)^T$, where $(\bar{X}_{j1}, \ldots, \bar{X}_{jp}) =: \bar{\boldsymbol{X}}_j = \frac{1}{n_j}\sum_{t=1}^{n_j} \boldsymbol{X}_{jt}$, $1 \le j \le K$. For testing the null hypothesis $H_0 : |\boldsymbol{\delta}_i|_2 = 0$, for $i = 1, \ldots, p$, a natural class of test statistics is

$$M_{\boldsymbol{A}} = \max_{1\le i\le p} \sum_{1\le j<l\le K} \frac{n_j n_l}{n_j + n_l} \frac{(\boldsymbol{A}(\bar{\boldsymbol{X}}_j - \bar{\boldsymbol{X}}_l))_i^2}{b_{ii}}, \tag{2}$$

where $(b_{ij}) =: \boldsymbol{B} = \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}$. In the present paper, we are particularly interested in the choice of $\boldsymbol{A} = \boldsymbol{\Sigma}^{-1} =: \boldsymbol{\Omega} := (\omega_{ij})$,

$$M_{\boldsymbol{\Omega}} = \max_{1\le i\le p} \sum_{1\le j<l\le K} \frac{n_j n_l}{n_j + n_l} \frac{(\boldsymbol{\Omega}(\bar{\boldsymbol{X}}_j - \bar{\boldsymbol{X}}_l))_i^2}{\omega_{ii}}. \tag{3}$$

In the two-sample case, [7] showed that the choice of precision matrix works well and the resulting test enjoys certain optimality against sparse alternatives. The motivation on the linear transformation of the data by the precision matrix $\boldsymbol{\Omega}$ in the multiple-sample case is similar as in [7]. Under a sparse alternative, the power of a test mainly depends on the magnitudes of the signals (nonzero coordinates of $(|\boldsymbol{\delta}_1|_2, \ldots, |\boldsymbol{\delta}_p|_2)^T$) and the number of the signals. It will be shown in Section 6 that $|\boldsymbol{\eta}_i|_2$ is approximately equal to $\omega_{ii}|\boldsymbol{\delta}_i|_2$ for all i such that $|\boldsymbol{\delta}_i|_2 \ne 0$. The magnitudes of the nonzero signals $|\boldsymbol{\delta}_i|_2$ are then transformed to $\omega_{ii}^{\frac{1}{2}}|\boldsymbol{\delta}_i|_2$ after normalized by the standard deviation of the transformed variable $(\boldsymbol{\Omega}\boldsymbol{X})_i$. In comparison, the magnitudes of the signals in the original data are $|\boldsymbol{\delta}_i|_2/\sigma_{ii}^{\frac{1}{2}}$. It can be seen from the inequality $\omega_{ii}\sigma_{ii} \ge 1$ for $i = 1, \ldots, p$ that $\omega_{ii}^{\frac{1}{2}}|\boldsymbol{\delta}_i|_2 \ge |\boldsymbol{\delta}_i|_2/\sigma_{ii}^{\frac{1}{2}}$. That is, such a linear transformation magnifies the signals and the number of the signals due to the dependence in the data. The transformation thus helps to distinguish the null and alternative hypothesis. The advantage of this linear transformation will be discussed in Section 5. Similar transformations are also studied in, for example, the detection problem through the innovated higher criticism in [13]. A similar innovated thresholding method is also considered in [10] for an optimal classification procedure.

A natural choice of $\boldsymbol{A}$ is $\boldsymbol{A} = \boldsymbol{I}$. That is, the test is directly based on the sample means $\bar{\boldsymbol{X}}_j - \bar{\boldsymbol{X}}_l$ for $1 \le j < l \le K$. Define the test statistic

$$M_{\boldsymbol{I}} = \max_{1\le i\le p} \sum_{1\le j<l\le K} \frac{n_j n_l}{n_j + n_l} \frac{(\bar{\boldsymbol{X}}_j - \bar{\boldsymbol{X}}_l)_i^2}{\sigma_{ii}}, \tag{4}$$

where σ_{ii} are the diagonal elements of $\boldsymbol{\Sigma}$. It will be shown in Section 5 that the test based on $M_{\boldsymbol{I}}$ is uniformly outperformed by the test based on $M_{\boldsymbol{\Omega}}$ for testing against sparse alternatives.

2.2. Data-driven procedure

We have so far focused on the oracle case in which the covariance matrix is known. For testing the hypothesis $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_K$ in the case of unknown covariance matrix, motivated by the oracle procedure $M_{\boldsymbol{A}}$ given in Section 2.1, the general test statistic is $M_{\widehat{\boldsymbol{A}}}$, where $\widehat{\boldsymbol{A}}$ is an estimator for $\boldsymbol{A}$, defined by

$$M_{\widehat{\boldsymbol{A}}} = \max_{1\le i\le p} \sum_{1\le j<l\le K} \frac{n_j n_l}{n_j + n_l} \frac{(\widehat{\boldsymbol{A}}(\bar{\boldsymbol{X}}_j - \bar{\boldsymbol{X}}_l))_i^2}{\hat{b}_{ii}}, \tag{5}$$

where $(\hat{b}_{ij}) =: \widehat{\boldsymbol{B}} = \frac{1}{\sum_{l=1}^K n_l - K}\{\sum_{l=1}^K \sum_{t=1}^{n_l} (\widehat{\boldsymbol{A}}(\boldsymbol{X}_{lt} - \bar{\boldsymbol{X}}_l))(\widehat{\boldsymbol{A}}(\boldsymbol{X}_{lt} - \bar{\boldsymbol{X}}_l))^T\}$. For the specific choice of $\boldsymbol{A} = \boldsymbol{\Omega}$, we use the constrained ℓ_1 minimization method given in [6] to estimate $\boldsymbol{\Omega}$. Other good estimators of the precision matrix can also be used. See more discussions in Remark 2 in Section 3.3.2. Then our final test statistic is

$$M_{\widehat{\boldsymbol{\Omega}}} = \max_{1\le i\le p} \sum_{1\le j<l\le K} \frac{n_j n_l}{n_j + n_l} \frac{(\widehat{\boldsymbol{\Omega}}(\bar{\boldsymbol{X}}_j - \bar{\boldsymbol{X}}_l))_i^2}{\hat{b}_{ii}}, \tag{6}$$

with $(\hat{b}_{ij}) =: \widehat{\boldsymbol{B}} = \frac{1}{\sum_{l=1}^{K} n_l - K}\{\sum_{l=1}^{K}\sum_{t=1}^{n_l}(\widehat{\boldsymbol{\Omega}}(\boldsymbol{X}_{lt} - \bar{\boldsymbol{X}}_l))(\widehat{\boldsymbol{\Omega}}(\boldsymbol{X}_{lt} - \bar{\boldsymbol{X}}_l))^T\}$. The simulation results in Section 4 show that the numerical performance of the test based on $M_{\widehat{\boldsymbol{\Omega}}}$ is similar to that of the test based on $M_{\boldsymbol{\Omega}}$.

3. Theoretical analysis

We now turn to the analysis of the properties of $M_{\boldsymbol{\Omega}}$ and $M_{\widehat{\boldsymbol{\Omega}}}$ including the limiting null distribution and the power of the corresponding tests. An intermediate correction for the limiting distribution is introduced. We will show that the test based on $M_{\boldsymbol{\Omega}}$ enjoys certain optimality when testing against sparse alternatives. Moreover, under suitable conditions the test based on $M_{\widehat{\boldsymbol{\Omega}}}$ performs as well as that based on $M_{\boldsymbol{\Omega}}$ and thus shares the same optimality. The asymptotic null distribution of $M_{\boldsymbol{I}}$ is also derived.

3.1. Asymptotic distributions of the oracle test statistics

We first establish the asymptotic null distributions for the oracle test statistics $M_{\boldsymbol{\Omega}}$ and $M_{\boldsymbol{I}}$. Let $\boldsymbol{D}_1 = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{pp})$ and $\boldsymbol{D}_2 = \mathrm{diag}(\omega_{11}, \ldots, \omega_{pp})$, where σ_{kk} and ω_{kk} are the diagonal entries of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ respectively. The correlation matrix of $\boldsymbol{X}$ is then $\boldsymbol{\Gamma} = (\gamma_{ij}) = \boldsymbol{D}_1^{-1/2}\boldsymbol{\Sigma}\boldsymbol{D}_1^{-1/2}$ and the correlation matrix of $\boldsymbol{\Omega}\boldsymbol{X}$ is $\boldsymbol{R} = (r_{ij}) = \boldsymbol{D}_2^{-1/2}\boldsymbol{\Omega}\boldsymbol{D}_2^{-1/2}$. To obtain the limiting null distributions, we assume that the eigenvalues of the covariance matrix $\boldsymbol{\Sigma}$ are bounded from above and below, and the correlations in $\boldsymbol{\Gamma}$ and $\boldsymbol{R}$ are bounded away from -1 and 1. More specifically we assume the following:

(C1) : $C_0^{-1} \le \lambda_{\min}(\boldsymbol{\Sigma}) \le \lambda_{\max}(\boldsymbol{\Sigma}) \le C_0$ for some constant $C_0 > 0$;
(C2) : $\max_{1\le i<j\le p}|\gamma_{ij}| \le r_1 < 1$ for some constant $0 < r_1 < 1$;
(C3) : $\max_{1\le i<j\le p}|r_{ij}| \le r_2 < 1$ for some constant $0 < r_2 < 1$.

Condition (C1) on the eigenvalues is a common assumption in the high-dimensional setting. Conditions (C2) and (C3) are also mild. For example, if $\max_{1\le i<j\le p}|r_{ij}| = 1$, then $\boldsymbol{\Sigma}$ is singular.

Let $\boldsymbol{Y}_i = \frac{1}{\sigma_{ii}}(\sqrt{\frac{n_1n_2}{n_1+n_2}}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)_i, \sqrt{\frac{n_1n_3}{n_1+n_3}}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_3)_i, \ldots, \sqrt{\frac{n_{K-1}n_K}{n_{K-1}+n_K}}(\bar{\boldsymbol{X}}_{K-1} - \bar{\boldsymbol{X}}_K)_i)^T$, Let $\boldsymbol{\Sigma}_0$ be the $b \times b$ covariance matrix of $\boldsymbol{Y}_i := (Y_{1i}, \ldots, Y_{bi})$ for $i = 1, \ldots, p$, where $b = \frac{K(K-1)}{2}$. Let σ^2 be the largest eigenvalue of $\boldsymbol{\Sigma}_0$ and d be the dimension of the corresponding eigenspace. Let σ_i^2, $1 \le i < d'$, be the positive eigenvalues of $\boldsymbol{\Sigma}_0$ arranged in a nonincreasing order and taking into account the multiplicities. Further, if $d' < \infty$, put $\sigma_i^2 = 0, i \ge d'$. Let $H(\boldsymbol{\Sigma}) := \prod_{i=d+1}^{\infty}(1 - \sigma_i^2/\sigma^2)^{-1/2}$. Then the following theorem states the asymptotic null distributions for the oracle statistics $M_{\boldsymbol{\Omega}}$ and $M_{\boldsymbol{I}}$.

Theorem 1. *Let the test statistics $M_{\boldsymbol{\Omega}}$ and $M_{\boldsymbol{I}}$ be defined as in (3) and (4), respectively.*
(i) *Suppose* (C1) *and* (C3) *hold. Then for any $x \in \mathbb{R}$, as $p \to \infty$,*

$$P_{H_0}\left(M_{\boldsymbol{\Omega}} - 2\sigma^2\log p - (d-2)\sigma^2\log\log p \le x\right) \to \exp\left(-\Gamma^{-1}\left(\frac{d}{2}\right)H(\boldsymbol{\Sigma})\exp\left(-\frac{x}{2\sigma^2}\right)\right)$$

where $\Gamma(\cdot)$ is the gamma function.
(ii) *Suppose* (C1) *and* (C2) *hold. Then for any $x \in \mathbb{R}$, as $p \to \infty$,*

$$P\left(M_{\boldsymbol{I}} - 2\sigma^2\log p - (d-2)\sigma^2\log\log p \le x\right) \to \exp\left(-\Gamma^{-1}\left(\frac{d}{2}\right)H(\boldsymbol{\Sigma})\exp\left(-\frac{x}{2\sigma^2}\right)\right).$$

When the sample sizes are equal, that is, $n_1 = n_2 = \cdots = n_K$, it is easy to check that $\sigma^2 = \frac{K}{2}$, $d = K-1$ and $H(\boldsymbol{\Sigma}) = 1$. Thus, we have the following simple expression for the asymptotic limiting distribution.

Corollary 1. *Let the test statistics $M_{\boldsymbol{\Omega}}$ and $M_{\boldsymbol{I}}$ be defined as in (2) and (4), respectively.*
(i) *Suppose* (C1) *and* (C3) *hold and $n_1 = n_2 = \cdots = n_K$. Then for any $x \in \mathbb{R}$, as $p \to \infty$,*

$$P_{H_0}\left(M_{\boldsymbol{\Omega}} - K\log p - \frac{K(K-3)}{2}\log\log p \le x\right) \to \exp\left(-\Gamma^{-1}\left(\frac{K-1}{2}\right)\exp\left(-\frac{x}{K}\right)\right).$$

(ii) *Suppose* (C1) *and* (C2) *hold and $n_1 = n_2 = \cdots = n_K$. Then for any $x \in \mathbb{R}$, as $p \to \infty$,*

$$P\left(M_{\boldsymbol{I}} - K\log p - \frac{K(K-3)}{2}\log\log p \le x\right) \to \exp\left(-\Gamma^{-1}\left(\frac{K-1}{2}\right)\exp\left(-\frac{x}{K}\right)\right).$$

Theorem 1 holds for any fixed sample sizes n_j for $1 \le j \le K$ and it shows that $M_{\boldsymbol{\Omega}}$ and $M_{\boldsymbol{I}}$ have the same asymptotic null distribution. Based on the limiting null distribution, we propose the asymptotically α-level test

$$\Phi_\alpha(\boldsymbol{\Omega}) = I\{M_{\boldsymbol{\Omega}} \ge 2\sigma^2\log p + (d-2)\sigma^2\log\log p + q_\alpha\} \quad (7)$$

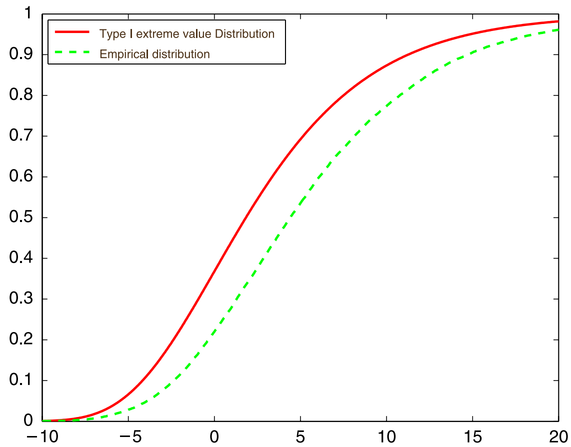


Fig. 1. Comparison of the empirical cumulative distribution and the limiting cumulative distributions with $p = 200$, $n_1 = \cdots = n_5 = 100$ and $K = 5$.

where q_α is the $1-\alpha$ quantile of the type I extreme value distribution with cumulative distribution function $\exp\left(-\Gamma^{-1}\left(\frac{d}{2}\right) H(\mathbf{\Sigma}) \exp\left(-\frac{x}{2\sigma^2}\right)\right)$, i.e.,

$$q_\alpha = -2\sigma^2 \log\left(\Gamma\left(\frac{d}{2}\right)\right) + 2\sigma^2 \log(H(\mathbf{\Sigma})) - 2\sigma^2 \log\log(1-\alpha)^{-1}.$$

The null hypothesis H_0 is rejected if and only if $\Phi_\alpha(\cdot) = 1$. Similarly, we define

$$\Phi_\alpha(\boldsymbol{I}) = I\{M_{\boldsymbol{I}} \geq 2\sigma^2 \log p + (d-2)\sigma^2 \log\log p + q_\alpha\}.$$

Although the asymptotic null distribution of the test statistics $M_{\mathbf{\Omega}}$ and $M_{\boldsymbol{I}}$ are the same, the power of the tests $\Phi_\alpha(\mathbf{\Omega})$ and $\Phi_\alpha(\boldsymbol{I})$ are quite different. It is shown in Section 5 that the power of $\Phi_\alpha(\mathbf{\Omega})$ uniformly dominates the power of $\Phi_\alpha(\boldsymbol{I})$ when testing against sparse alternatives.

3.2. *Intermediate correction factor for large K*

When the number of groups is larger than 3, the test $\Phi_\alpha(\mathbf{\Omega})$ given in (7) based on the asymptotic distribution under the null hypothesis summarized in Theorem 1 has serious size distortion because the convergence rate in distribution of the extreme value type statistics is slow. See, for example, [12,15,4]. Fig. 1 illustrates the size distortion of the limiting distribution in Theorem 1 by comparing its cumulative distribution with the empirical distribution when the data are generated from $\mathrm{N}(\mathbf{0}, \boldsymbol{I})$, with $p = 200$, $n_1 = \cdots = n_K = 100$ and $K = 5$.

It can be seen from Fig. 1 that there is a noticeable difference between the two cumulative distributions, and directly applying the limiting distribution in Theorem 1 would lead to a test whose true size is significantly different from the nominal level. This distortion mainly comes from the accumulation of the normal approximation error when K is relatively large. Thus, instead of directly calculating the approximated normal tails, we derive the following intermediate correction for the asymptotic limiting null distribution.

Proposition 1. *Define the test statistics $M_{\mathbf{\Omega}}$ and $M_{\boldsymbol{I}}$ as in (3) and (4), respectively.*

(i) *Suppose* (C1) *and* (C3) *hold. Then for any $x \in \mathbb{R}$,*

$$P_{H_0}\left(M_{\mathbf{\Omega}} \leq x_p\right) / \exp\left(-p \cdot P(\|\boldsymbol{Y}\|_2^2 \geq x_p)\right) \to 1$$

as $p \to \infty$, where $x_p = 2\sigma^2 \log p + (d-2)\sigma^2 \log\log p + x$ and $\boldsymbol{Y}$ is a Gaussian random variable with mean zero and covariance matrix $\mathbf{\Sigma}_0$, where $\mathbf{\Sigma}_0$ is the $b \times b$ covariance matrix as defined in Section 3.1.

(ii) *Suppose* (C1) *and* (C2) *hold. Then for any $x \in \mathbb{R}$*

$$P_{H_0}\left(M_{\boldsymbol{I}} \leq x_p\right) / \exp\left(-p \cdot P(\|\boldsymbol{Y}\|_2^2 \geq x_p)\right) \to 1$$

as $p \to \infty$, where $x_p = 2\sigma^2 \log p + (d-2)\sigma^2 \log\log p + x$ and $\boldsymbol{Y}$ is a Gaussian random variable with mean zero and covariance matrix $\mathbf{\Sigma}_0$.

In light of the results given in Proposition 1, for any $p \times p$ positive definite matrix $\boldsymbol{A}$, based on the test statistic $M_{\boldsymbol{A}}$ given in (2), a corrected α-level test can be defined by $\Psi_\alpha(\boldsymbol{A}) = I\{M_{\boldsymbol{A}} \geq t_{\alpha,p}\}$, where $t_{\alpha,p}$ satisfies $P(\|\boldsymbol{Y}\|_2 \geq t_{\alpha,p}) = -1/p \log(1-\alpha)$

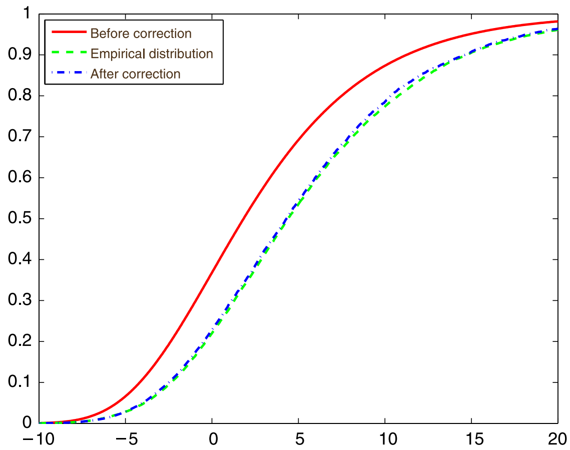


Fig. 2. Comparison of three cumulative distributions with $p = 200$, $n_1 = \cdots = n_5 = 100$ and $K = 5$.

and $\boldsymbol{Y}$ is a Gaussian random variable with mean zero and covariance matrix $\boldsymbol{\Sigma}_0$. In particular, we propose the corrected α-level test

$$\Psi_\alpha(\boldsymbol{\Omega}) = I\{M_{\boldsymbol{\Omega}} \geq t_{\alpha,p}\}. \tag{8}$$

Similarly, we define $\Psi_\alpha(\boldsymbol{I}) = I\{M_{\boldsymbol{I}} \geq t_{\alpha,p}\}$.

As an illustration of the accuracy of the corrected distribution in Proposition 1, we compare its cumulative distribution with the empirical distribution under the same setting as in Fig. 1, as well as the limiting distribution derived in Theorem 1. We can see from Fig. 2 that the corrected asymptotic distribution is much closer to the empirical distribution and as a result will provide a much more precise cutoff value for a given nominal level. Simulation results in Section 4 show that the actual size of $\Psi_\alpha(\boldsymbol{\Omega})$ is close to the pre-specified nominal level. We recommend to use the test $\Phi_\alpha(\boldsymbol{\Omega})$ given in (7) for $K \leq 3$ and use the test $\Psi_\alpha(\boldsymbol{\Omega})$ given in (8) for $K \geq 4$.

3.3. The asymptotic properties of $\Phi_\alpha(\boldsymbol{\Omega})$ and $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$

In this section, we analyze the asymptotic power of the test $\Phi_\alpha(\boldsymbol{\Omega})$ and show that it is minimax rate optimal against sparse alternatives. For a given positive definite matrix A, the corrected test $\Psi_\alpha(\boldsymbol{A})$ shares the same asymptotic properties as $\Phi_\alpha(\boldsymbol{A})$ since it is derived from the intermediate correction term of the limiting distribution in Theorem 1 instead of directly calculating the tail probability. Thus in this section we focus the discussion on the asymptotic properties of $\Phi_\alpha(\boldsymbol{A})$.

In practice, $\boldsymbol{\Omega}$ is unknown and the test statistic $M_{\widehat{\boldsymbol{\Omega}}}$ should be used instead of $M_{\boldsymbol{\Omega}}$. Define the set of k_p-sparse vectors by

$$\mathcal{S}(k_p) = \left\{\boldsymbol{\delta}^{(jl)}, 1 \leq j < l \leq K : \max_{1\leq j<l\leq K} \sum_{i=1}^{p} I\{\delta_i^{(jl)} \neq 0\} \leq k_p\right\},$$

where $\boldsymbol{\delta}^{(jl)} = \sqrt{\frac{n_j n_l}{n_j+n_l}}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_l)$. Throughout the section, we analyze the power of $M_{\boldsymbol{\Omega}}$ and $M_{\widehat{\boldsymbol{\Omega}}}$ under the alternative

$$H_1 : \{\boldsymbol{\delta}^{(jl)}, 1 \leq j < l \leq K\} \in \mathcal{S}(k_p) \text{ with } k_p = p^r \text{ and the nonzero locations of } \boldsymbol{\delta}^{(jl)},$$
$$\text{for every } 1 \leq j < l \leq K, \text{ are randomly uniformly drawn from } \{1, \ldots, p\}.$$

As discussed in [7], the condition on the nonzero coordinates in H_1 is mild. The same condition has been imposed in [13]. We show that, under some suitable assumptions, $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ performs as well as $\Phi_\alpha(\boldsymbol{\Omega})$ asymptotically.

3.3.1. The asymptotic power of $\Phi_\alpha(\boldsymbol{\Omega})$ and its optimality

The asymptotic power of $\Phi_\alpha(\boldsymbol{\Omega})$ is analyzed under certain conditions on the separation among $\boldsymbol{\mu}_j$ and $\boldsymbol{\mu}_l$ for $1 \leq j < l \leq K$. Furthermore, a lower bound is derived to show that this condition is minimax rate optimal in order to distinguish H_1 and H_0 with probability tending to 1.

Theorem 2. *Suppose that* (C1) *holds. If* $r < 1/4$ *and* $\max_i |\boldsymbol{\delta}_i|_2/\sigma_{ii}^{\frac{1}{2}} \geq \sqrt{2\sigma^2\beta\log p}$ *with* $\beta \geq 1/(\min_i \sigma_{ii}\omega_{ii}) + \varepsilon$ *for some constant* $\varepsilon > 0$, *then we have, as* $p \to \infty$,

$$P_{H_1}(\Phi_\alpha(\boldsymbol{\Omega}) = 1) \to 1.$$

We shall show that the condition $\max_i |\boldsymbol{\delta}_i|_2/\sigma_{ii}^{\frac{1}{2}} \geq \sqrt{2\sigma^2\beta\log p}$ is minimax rate optimal for testing against sparse alternatives, which is a direct result of Theorem 3 in [7]. First we introduce some conditions as in [7].

(C4) $k_p = p^r$ for some $r < 1/2$ and $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ is s_p-sparse with $s_p = O((p/k_p^2)^\gamma)$ for some $0 < \gamma < 1$.
(**C4**′) $k_p = p^r$ for some $r < 1/4$.
(C5) $\|\boldsymbol{\Omega}\|_{L_1} \leq M$ for some constant $M > 0$.

Define the class of α-level tests by

$$\mathcal{T}_\alpha = \{\Phi_\alpha : P_{H_0}(\Phi_\alpha = 1) \leq \alpha\}.$$

Let $\mathcal{A}_{\delta,c} = \mathcal{S}(k_p) \cap \{\max_{1\leq i\leq p} |\boldsymbol{\delta}_i|_2 \geq c\sqrt{\log p}\}$ be a set of k_p-sparse vectors $\{\boldsymbol{\delta}^{(jl)}, 1 \leq j < l \leq K\}$ with the ℓ_∞ norm of $(|\boldsymbol{\delta}_1|_2, \ldots, |\boldsymbol{\delta}_p|_2)$ having the magnitude greater than or equal to $c\sqrt{\log p}$ for some constant $c > 0$. The following theorem shows that the condition $\max_i |\boldsymbol{\delta}_i|_2/\sigma_{ii}^{\frac{1}{2}} \geq \sqrt{2\sigma^2\beta\log p}$ is minimax rate optimal.

Theorem 3. *Assume that* (C4) *(or* (C4′)*) and* (C5) *hold. Let* $\alpha, \nu > 0$ *and* $\alpha + \nu < 1$*. Then there exists a constant* $c > 0$ *such that for all sufficiently large* n_i *and* p, $i = 1, \ldots, K$,

$$\inf_{\{\boldsymbol{\delta}^{(jl)}, 1\leq j<l\leq K\}\in\mathcal{A}_{\delta,c}} \sup_{\Phi_\alpha\in\mathcal{T}_\alpha} P(\Phi_\alpha = 1) \leq 1 - \nu.$$

Remark 1. The lower bound result follows directly from Theorem 3 in [7]. We construct $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ exactly the same as the worst case in the proof of lower bound result in [7] and let $\boldsymbol{\mu}_j = \mathbf{0}$ for $j = 3, \ldots, K$. Then the result of above theorem follows.

3.3.2. The asymptotic properties of $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ and its optimality

We now analyze the properties of $M_{\widehat{\boldsymbol{\Omega}}}$ and the corresponding test including the limiting null distribution and the asymptotic power. We shall show that $M_{\widehat{\boldsymbol{\Omega}}}$ has the same limiting null distribution as $M_{\boldsymbol{\Omega}}$ and define the corresponding test $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ by

$$\Phi_\alpha(\widehat{\boldsymbol{\Omega}}) = I\{M_{\widehat{\boldsymbol{\Omega}}} \geq 2\sigma^2\log p + (d-2)\sigma^2\log\log p + q_\alpha\}.$$

Under some suitable assumptions, the asymptotic properties of $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ are similar to those of $\Phi_\alpha(\boldsymbol{\Omega})$. Define the following class of matrices that belong to an ℓ_q ball with $0 \leq q < 1$:

$$\mathcal{U}_q(s_p, M_p) = \left\{\boldsymbol{\Omega} \succ 0 : \|\boldsymbol{\Omega}\|_{L_1} \leq M_p, \ \max_{1\leq j\leq p}\sum_{i=1}^p |\omega_{ij}|^q \leq s_p\right\}.$$

We assume that $\boldsymbol{\Omega} \in \mathcal{U}_q(s_p, M_p)$ so $\boldsymbol{\Omega}$ can be well estimated by the CLIME estimator $\widehat{\boldsymbol{\Omega}}$ under some conditions on s_p and M_p; see [6].

Theorem 4. *Suppose that* (C1) *and* (C3) *hold and* $\boldsymbol{\Omega} \in \mathcal{U}_q(s_p, M_p)$ *with*

$$s_p = o\left(\frac{n^{(1-q)/2}}{M_p^{1-q}(\log p)^{(3-q)/2}}\right). \tag{9}$$

(i)*Then under the null hypothesis* H_0*, for any* $x \in \mathbb{R}$,

$$P_{H_0}\left(M_{\widehat{\boldsymbol{\Omega}}} - 2\sigma^2\log p - (d-2)\sigma^2\log\log p \leq x\right) \to \exp\left(-\Gamma^{-1}\left(\frac{d}{2}\right)H(\boldsymbol{\Sigma}_0)\exp\left(-\frac{x}{K}\right)\right),$$

as $n_j, p \to \infty$ *for* $j = 1, \ldots, K$*. Furthermore, for any* $x \in \mathbb{R}$,

$$P_{H_0}\left(M_{\widehat{\boldsymbol{\Omega}}} \leq x_p\right) / \exp\left(-p \cdot P(\|\boldsymbol{Y}\|_2^2 \geq x_p)\right) \to 1$$

as $n_j, p \to \infty$*, where* $x_p = 2\sigma^2\log p + (d-2)\sigma^2\log\log p + x$ *and* $\boldsymbol{Y}$ *is a Gaussian mean zero r.v. with covariance matrix* $\boldsymbol{\Sigma}_0$.
(ii)*Under the alternative hypothesis* H_1 *with* $r < 1/6$*, we have*

$$\frac{P_{H_1}\left(\Phi_\alpha(\widehat{\boldsymbol{\Omega}}) = 1\right)}{P_{H_1}\left(\Phi_\alpha(\boldsymbol{\Omega}) = 1\right)} \to 1,$$

as $n_j, p \to \infty$ *for* $j = 1, \ldots, K$*. Furthermore, if* $\max_i |\boldsymbol{\delta}_i|_2/\sigma_{ii}^{\frac{1}{2}} \geq \sqrt{2\sigma^2\beta\log p}$ *with* $\beta \geq 1/(\min_i \sigma_{ii}\omega_{ii}) + \varepsilon$ *for some constant* $\varepsilon > 0$*, then we have, for* $j = 1, \ldots, K$,

$$P_{H_1}\left(\Phi_\alpha(\widehat{\boldsymbol{\Omega}}) = 1\right) \to 1, \quad \text{as } n_j, p \to \infty.$$

By Theorem 4, we see that $M_{\widehat{\Omega}}$ and M_{Ω} have the same asymptotic distribution and power, and so the test $\Phi_\alpha(\widehat{\mathbf{\Omega}})$ is also minimax rate optimal.

Remark 2. The CLIME estimator in [6] is considered in this section. As in the two-sample case, other "good" estimators of the precision matrix can also be used. In general, Theorem 4 still holds if $\log p = o(n)$ and the estimator $\widehat{\mathbf{\Omega}}$ satisfies the following conditions:

$$\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}\|_{L_1} = o_P\left(\frac{1}{\log p}\right) \quad \text{and} \quad \max_{1\le i\le p} |\hat{b}_{ii} - b_{ii}| = o_P\left(\frac{1}{\log p}\right), \tag{10}$$

where $(b_{ij}) =: \boldsymbol{B} = \mathbf{\Omega}\mathbf{\Sigma}\mathbf{\Omega}$ and $(\hat{b}_{ij}) =: \widehat{\boldsymbol{B}} = \widehat{\mathbf{\Omega}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{\Omega}}$.

3.3.3. Comparison with $\Phi_\alpha(\boldsymbol{I})$

It is interesting to compare the power of the new test with the maximum test based on the original observations. More specifically, we compare the power of the test $\Phi_\alpha(\mathbf{\Omega})$ with that of $\Phi_\alpha(\boldsymbol{I})$ under the same alternative H_1 as in Section 3.3.2. We show in the following Proposition that the power of $\Phi_\alpha(\mathbf{\Omega})$ dominates the power of $\Phi_\alpha(\boldsymbol{I})$ under suitable conditions.

Proposition 2. *Suppose* (C1)–(C3) *hold. Then under H_1 with $r < 1/6$, we have*

$$\liminf_{p\to\infty} \frac{P_{H_1}(\Phi_\alpha(\mathbf{\Omega}) = 1)}{P_{H_1}(\Phi_\alpha(\boldsymbol{I}) = 1)} \ge 1. \tag{11}$$

Proposition 2 shows that, under some sparsity conditions on $\{\boldsymbol{\delta}^{(jl)}, 1 \le j < l \le K\}$, $\Phi_\alpha(\mathbf{\Omega})$ is uniformly at least as powerful as $\Phi_\alpha(\boldsymbol{I})$. The test $\Phi_\alpha(\mathbf{\Omega})$ can be strictly more powerful than $\Phi_\alpha(\boldsymbol{I})$. Assume that

$$\begin{aligned} &H_1' : \max_{1\le j<l\le K} \sum_{i=1}^{p} I\{\delta_i^{(jl)} \neq 0\} = k_p = p^r, \quad r < \frac{1}{2}, \\ &\text{with nonzero elements } \sqrt{2\sigma^2\beta_0 \log p} \le \frac{|\boldsymbol{\delta}_i|_2}{\sqrt{\sigma_{ii}}} \le \sqrt{2\sigma^2\beta_1 \log p}. \end{aligned} \tag{12}$$

The nonzero locations of $\boldsymbol{\delta}^{(jl)}$, for every $1 \le j < l \le K$, are randomly and uniformly drawn from $\{1, \dots, p\}$.

Proposition 3. *Suppose that* (C1)–(C3) *hold and* $\min_{1\le i\le p} \sigma_{ii}\omega_{ii} \ge 1 + \varepsilon_1$ *for some* $\varepsilon_1 > 0$. *Then, under* H_1' *with*

$$\frac{(1-\sqrt{r})^2}{\min\limits_{1\le i\le p} \sigma_{ii}\omega_{ii}} + \epsilon \le \beta_0 < \beta_1 < (1-\sqrt{r})^2$$

for some $\epsilon > 0$, *we have*

$$\lim_{p\to\infty} P_{H_1'}(\Phi_\alpha(\mathbf{\Omega}) = 1) = 1$$

and

$$\limsup_{p\to\infty} P_{H_1'}(\Phi_\alpha(\boldsymbol{I}) = 1) \le \alpha.$$

When the variables are correlated, ω_{ii} can be strictly larger than $1/\sigma_{ii}$. For example, let $\mathbf{\Sigma} = (\phi^{|i-j|})$ with $|\phi| < 1$. Then $\min_{1\le i\le p} \sigma_{ii}\omega_{ii} \ge (1-\phi^2)^{-1} > 1$. That is, $\Phi_\alpha(\mathbf{\Omega})$ is strictly more powerful than $\Phi_\alpha(\boldsymbol{I})$ under H_1'. For reasons of space, we omit the proofs of these two propositions.

4. Simulation study

In this section, we consider the numerical performance of the tests $\Phi_\alpha(\mathbf{\Omega})$ and $\Phi_\alpha(\widehat{\mathbf{\Omega}})$ and compare these tests with a number of other tests, including the oracle test $\Phi_\alpha(\boldsymbol{I})$, the tests based on the sum of squares type statistics in [11,16,19], and the commonly used likelihood ratio test. These last four tests are denoted respectively by FHW, Sc, Sr and LRT respectively in the tables below.

In the simulations, we consider two settings on the number of the groups: $K = 3$ and $K = 5$. We follow the recommendations made in Section 3.2 by using the test $\Phi_\alpha(\mathbf{\Omega})$ given in (7) for $K = 3$ and using the test $\Psi_\alpha(\mathbf{\Omega})$ given in (8) for $K = 5$. We shall always take $\boldsymbol{\mu}_1 = \mathbf{0}$. Under the null hypothesis, $\boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_K = \mathbf{0}$, while under the alternative hypothesis, we take $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{ip})^T$, for $i = 2, \dots, K$, to have m nonzero entries with the support $S_i = \{l_{1i}, \dots, l_{im} : l_{i1} < l_{i2} < \cdots < l_{im}\}$

uniformly and randomly drawn from $\{1, \ldots, p\}$. For any $l_{ij} \in S_i$, $\mu_{il_{ij}}$ has magnitude randomly uniformly drawn from the interval $[-\sqrt{2\log p/n}, \sqrt{2\log p/n}]$. We take $\mu_{ik} = 0$ for $k \in S_i^c$.

For both $K = 3$ and $K = 5$, we consider three different values of m: the extreme sparse alternative, moderate sparse alternative and non-sparse alternative. In this simulation study, the dimension p takes values 50, 100, 200 and 400, and the corresponding values of m for each dimension are as follows. Under the extreme sparse alternative, let $m = 2$ for $p = 50$ and 100 and $m = 5$ for $p = 200$ and 400. We select $m = 5$ for $p = 50$, $m = 10$ for $p = 100$, $m = 15$ for $p = 200$ and $m = 20$ for $p = 400$ when the alternative is moderate sparse. In the scenario when the alternative is non-sparse, we choose $m = 20$ for $p = 50$, $m = 30$ for $p = 100$, $m = 40$ for $p = 200$ and $m = 50$ for $p = 400$. We consider $m = 50$ as a non-sparse alternative when $p = 400$, because in this case the number of nonzero entries of the difference of any pair of mean vectors can be as large as 100, and the value k_p as defined in Section 3.3 is equal to 150 when $K = 3$ and is equal to 250 when $K = 5$.

Three different settings of the precision matrix $\boldsymbol{\Omega}$ are considered in the simulation: $\boldsymbol{\Omega}$ is known, $\boldsymbol{\Omega}$ is unknown but sparse and the case where the covariance matrix $\boldsymbol{\Sigma}$ is unknown but sparse. In the case when $\boldsymbol{\Omega}$ is known, we compare the oracle performance of the three tests based on the maximum-type statistics with the tests based on the sum of squares type statistics. When $\boldsymbol{\Omega}$ is unknown, we use the CLIME estimator in [6] to estimate it when $\boldsymbol{\Omega}$ is sparse, while the inverse of the adaptive thresholding estimator in [5] is used to estimate when $\boldsymbol{\Sigma}$ is sparse.

Let $\boldsymbol{D} = (d_{ij})$ be a diagonal matrix with diagonal elements $d_{ii} = \text{Unif}(1, 3)$ for $i = 1, \ldots, p$. Denote by $\lambda_{\min}(\boldsymbol{A})$ the minimum eigenvalue of a symmetric matrix $\boldsymbol{A}$. In the case when the precision matrix $\boldsymbol{\Omega}$ is known, the following two models for $\boldsymbol{\Sigma}$ are considered:

- Model 1: $\boldsymbol{\Sigma}^* = (\sigma_{ij}^*)$ where $\sigma_{ii}^* = 1$, $\sigma_{ij}^* = 0.5$ for $i \neq j$. $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{\Sigma}^*\boldsymbol{D}^{1/2}$.
- Model 2: $\boldsymbol{\Sigma}^* = (\sigma_{ij}^*)$ where $\sigma_{ii}^* = 1$, $\sigma_{ij}^* = \text{Unif}(0, 1)$ for $i < j$ and $\sigma_{ji}^* = \sigma_{ij}^*$. $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}(\boldsymbol{\Sigma}^* + \delta\boldsymbol{I})/(1 + \delta)\boldsymbol{D}^{1/2}$ with $\delta = |\lambda_{\min}(\boldsymbol{\Sigma}^*)| + 0.05$.

In the case when the precision matrix $\boldsymbol{\Omega}$ is sparse, we consider the following two models:

- Model 3: $\boldsymbol{\Sigma} = (\sigma_{ij})$ where $\sigma_{ii} = 1$, $\sigma_{ij} = 0.8$ for $2(k-1)+1 \le i \neq j \le 2k$, where $k = 1, \ldots, [p/2]$ and $\sigma_{ij} = 0$ otherwise.
- Model 4: $\boldsymbol{\Sigma} = (\sigma_{ij})$ where $\sigma_{ij} = 0.6^{|i-j|}$ for $1 \le i, j \le p$.

The following two models are considered when the covariance matrix $\boldsymbol{\Sigma}$ is sparse:

- Model 5: $\boldsymbol{\Sigma}^* = (\sigma_{ij}^*)$ where $\sigma_{ii}^* = 1$, $\sigma_{ij}^* = 0.8$ for $2(k-1)+1 \le i \neq j \le 2k$, where $k = 1, \ldots, [p/2]$ and $\sigma_{ij}^* = 0$ otherwise. $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{\Sigma}^*\boldsymbol{D}^{1/2}$.
- Model 6: $\boldsymbol{\Omega} = (\omega_{ij})$ where $\omega_{ij} = 0.6^{|i-j|}$ for $1 \le i, j \le p$. $\boldsymbol{\Sigma} = \boldsymbol{D}^{1/2}\boldsymbol{\Omega}^{-1}\boldsymbol{D}^{1/2}$.

Under each model, two independent random samples $\{\boldsymbol{X}_k\}$ and $\{\boldsymbol{Y}_l\}$ are generated with the same sample size $n = 100$ and $n = 60$ for $K = 3$ and $K = 5$ respectively from two multivariate normal distributions with the means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ respectively and a common covariance matrix $\boldsymbol{\Sigma}$. The size and power are calculated from 1000 replications. The numerical results are summarized in Tables 1–4.

It can be seen from Table 1 that the estimated sizes are close to the nominal level 0.05 for all the tests. Tables 2–4 summarize the power results under various alternatives. Under the extreme sparsity alternative, Table 2 shows that the tests based on the sum of squares test statistics have trivial power, while the oracle test $\Phi_\alpha(\boldsymbol{\Omega})$ has the highest power in all six models over all dimensions ranging from 50 to 400, and the performance of the test $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ based on either the CLIME estimator or the inverse of the adaptive thresholding estimator is close to that of the oracle test $\Phi_\alpha(\boldsymbol{\Omega})$ in Models 3–6. Under the moderate sparsity alternative, similar phenomena are observed, the tests $\Phi_\alpha(\boldsymbol{\Omega})$ and $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ are significantly more powerful in comparison to the other tests.

When the number of nonzero entries increases, the powers of all tests increase as well. Under the non-sparse alternative, as can be seen from Table 4, the sum of squares type tests also enjoy high power in Models 3 and 4. In other models, the tests $\Phi_\alpha(\boldsymbol{\Omega})$ and $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ still significantly outperform the other tests though the alternative is non-sparse. In summary, The tests $\Psi_\alpha(\boldsymbol{\Omega})$ and $\Psi_\alpha(\widehat{\boldsymbol{\Omega}})$ perform similarly and significantly outperform the other tests against a full range of alternatives in the simulation study. Similar phenomena are observed for the corrected tests as shown in Tables 2–4.

As a graphical illustration, we also summarize the power comparison results in Fig. 3 for $K = 3$ and $p = 400$. The horizontal axis represents each model and the vertical axis represents the powers of the four tests. We do not include LRT, Sr and $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ because, when $p = 400$ LRT is not well defined and Sr has trivial power, and $\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$ has similar power as $\Phi_\alpha(\boldsymbol{\Omega})$. It can be easily seen from Fig. 3 that the test $\Phi_\alpha(\boldsymbol{\Omega})$ significantly outperforms the other tests.

5. Discussion

We introduced in this paper the data-driven testing procedure $\Phi_\alpha(\hat{\boldsymbol{\Omega}})$ and showed that it performs particularly well against sparse alternatives. This procedure requires a good estimate of the precision matrix $\boldsymbol{\Omega}$. We have mainly focused in this paper on the sparse precision matrices for which the CLIME estimator is known to perform well. The test $\Phi_\alpha(\hat{\boldsymbol{\Omega}})$ can be used with a much wider range of covariance/precision matrices. As mentioned in Section 3.3, one only needs an estimate $\hat{\boldsymbol{\Omega}}$ satisfying the ℓ_1 condition (10) and then the result given in Theorem 4 extends directly. For example, when the covariance matrix $\boldsymbol{\Sigma}$ is either sparse or bandable, Condition (10) can be achieved by inverting thresholding or tapering estimators of

Table 1
Empirical sizes of tests with $\alpha = 0.05$. $n = 100$ when $K = 3$ and $n = 60$ when $K = 5$. Based on 1000 replications.

p	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400
	$K = 3$																							
	Model 1				Model 2				Model 3				Model 4				Model 5				Model 6			
LRT	0.06	0.05	0.05	–	0.05	0.05	0.06	–	0.06	0.06	0.06	–	0.04	0.06	0.06	–	0.05	0.06	0.07	–	0.06	0.06	0.06	–
FHW	0.09	0.06	0.07	0.06	0.06	0.05	0.05	0.02	0.04	0.04	0.03	0.02	0.06	0.06	0.03	0.02	0.05	0.05	0.03	0.02	0.05	0.05	0.03	0.01
Sc	0.09	0.06	0.08	0.06	0.06	0.06	0.07	0.05	0.05	0.05	0.05	0.05	0.06	0.07	0.06	0.06	0.06	0.06	0.05	0.06	0.06	0.06	0.06	0.06
Sr	0.06	0.05	0.05	0.00	0.05	0.05	0.06	0.00	0.06	0.06	0.06	0.00	0.04	0.06	0.06	0.00	0.05	0.06	0.07	0.00	0.06	0.06	0.06	0.00
$\Phi_\alpha(\boldsymbol{I})$	0.03	0.03	0.03	0.02	0.04	0.05	0.04	0.06	0.04	0.04	0.04	0.03	0.05	0.04	0.05	0.05	0.04	0.06	0.04	0.03	0.04	0.05	0.05	0.05
$\Phi_\alpha(\boldsymbol{\Omega})$	0.05	0.06	0.04	0.05	0.04	0.03	0.05	0.03	0.04	0.04	0.06	0.05	0.05	0.05	0.05	0.06	0.04	0.05	0.04	0.05	0.05	0.04	0.06	0.05
$\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$	–	–	–	–	–	–	–	–	0.06	0.06	0.07	0.07	0.06	0.07	0.07	0.08	0.05	0.05	0.05	0.06	0.06	0.04	0.06	0.06
	$K = 5$																							
LRT	0.06	0.06	0.07	–	0.05	0.06	0.08	–	0.06	0.06	0.05	–	0.06	0.06	0.06	–	0.05	0.05	0.08	–	0.07	0.06	0.05	–
FHW	0.09	0.08	0.07	0.06	0.06	0.06	0.05	0.03	0.05	0.06	0.03	0.02	0.06	0.06	0.04	0.01	0.05	0.04	0.03	0.03	0.05	0.04	0.02	0.01
Sc	0.09	0.08	0.07	0.06	0.07	0.06	0.06	0.06	0.06	0.07	0.06	0.05	0.06	0.07	0.06	0.04	0.06	0.05	0.05	0.06	0.06	0.06	0.05	0.05
Sr	0.06	0.06	0.07	0.00	0.05	0.06	0.08	0.00	0.06	0.06	0.05	0.00	0.06	0.06	0.06	0.00	0.05	0.05	0.08	0.00	0.07	0.06	0.05	0.00
$\Psi_\alpha(\boldsymbol{I})$	0.04	0.03	0.03	0.03	0.05	0.04	0.05	0.05	0.04	0.06	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.04	0.05	0.04	0.05	0.05	0.05	0.05
$\Psi_\alpha(\boldsymbol{\Omega})$	0.04	0.06	0.05	0.04	0.04	0.05	0.06	0.04	0.03	0.05	0.05	0.06	0.05	0.06	0.06	0.04	0.06	0.05	0.05	0.05	0.05	0.05	0.04	0.05
$\Psi_\alpha(\widehat{\boldsymbol{\Omega}})$	–	–	–	–	–	–	–	–	0.06	0.07	0.07	0.08	0.06	0.06	0.06	0.07	0.05	0.07	0.07	0.07	0.06	0.06	0.06	0.06

Table 2
Powers of tests under extreme sparse alternative with $\alpha = 0.05$. Based on 1000 replications.

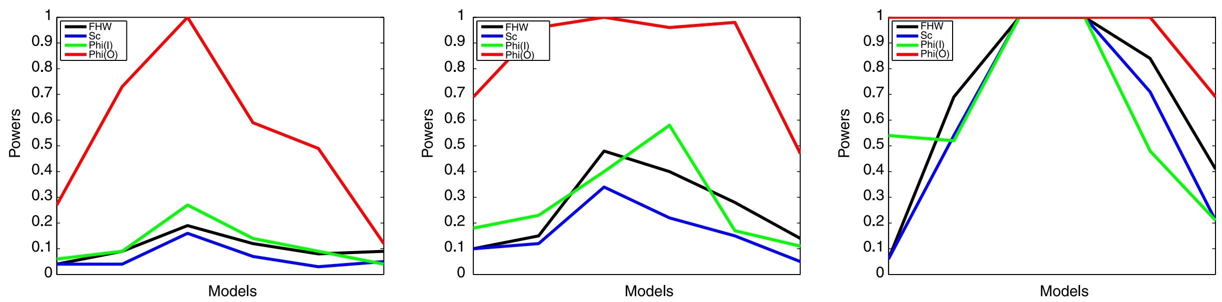
p	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400
	$K = 3$ with extreme sparse alternative																							
	Model 1				Model 2				Model 3				Model 4				Model 5				Model 6			
LRT	0.11	0.06	0.28	–	0.23	0.14	0.15	–	0.26	0.34	0.34	–	0.26	0.22	0.33	–	0.19	0.14	0.21	–	0.04	0.08	0.11	–
FHW	0.08	0.11	0.05	0.04	0.14	0.06	0.09	0.09	0.18	0.10	0.11	0.19	0.08	0.12	0.19	0.12	0.13	0.05	0.08	0.08	0.00	0.05	0.12	0.09
Sc	0.11	0.11	0.05	0.04	0.14	0.06	0.05	0.04	0.16	0.10	0.11	0.16	0.08	0.11	0.19	0.07	0.13	0.03	0.06	0.03	0.00	0.05	0.07	0.05
Sr	0.11	0.06	0.28	0.00	0.23	0.14	0.15	0.00	0.26	0.34	0.34	0.00	0.26	0.22	0.33	0.00	0.19	0.14	0.21	0.00	0.04	0.08	0.11	0.00
$\Phi_\alpha(\boldsymbol{I})$	0.07	0.05	0.11	0.06	0.16	0.07	0.09	0.09	0.14	0.13	0.26	0.27	0.17	0.16	0.25	0.14	0.08	0.07	0.11	0.09	0.05	0.09	0.11	0.04
$\Phi_\alpha(\boldsymbol{\Omega})$	0.07	0.12	0.48	0.27	0.29	0.15	0.45	0.73	0.48	0.78	0.82	1.00	0.28	0.43	0.69	0.59	0.19	0.17	0.45	0.49	0.12	0.11	0.18	0.12
$\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$	–	–	–	–	–	–	–	–	0.47	0.80	0.84	0.99	0.32	0.44	0.72	0.60	0.19	0.19	0.47	0.48	0.14	0.11	0.21	0.13
	$K = 5$ with extreme sparse alternative																							
LRT	0.07	0.32	0.33	–	0.54	0.15	0.52	–	0.79	0.72	0.71	–	0.44	0.36	0.19	–	0.21	0.35	0.28	–	0.29	0.13	0.13	–
FHW	0.09	0.13	0.07	0.07	0.10	0.09	0.08	0.10	0.18	0.14	0.27	0.25	0.26	0.20	0.11	0.21	0.11	0.14	0.12	0.13	0.18	0.11	0.09	0.11
Sc	0.09	0.13	0.07	0.08	0.08	0.07	0.07	0.05	0.18	0.11	0.21	0.22	0.26	0.20	0.10	0.13	0.11	0.13	0.07	0.08	0.18	0.09	0.07	0.02
Sr	0.07	0.32	0.33	0.00	0.54	0.15	0.52	0.00	0.79	0.72	0.71	0.00	0.44	0.36	0.19	0.00	0.21	0.35	0.28	0.00	0.29	0.13	0.13	0.00
$\Phi_\alpha(\boldsymbol{I})$	0.06	0.26	0.11	0.18	0.17	0.14	0.19	0.18	0.34	0.24	0.43	0.21	0.26	0.26	0.17	0.31	0.03	0.11	0.10	0.15	0.13	0.06	0.07	0.05
$\Phi_\alpha(\boldsymbol{\Omega})$	0.12	0.57	0.59	0.74	0.89	0.34	1.00	1.00	0.86	0.94	0.98	1.00	0.56	0.75	0.53	0.92	0.28	0.43	0.65	0.79	0.25	0.09	0.19	0.23
$\Phi_\alpha(\widehat{\boldsymbol{\Omega}})$	–	–	–	–	–	–	–	–	0.85	0.96	0.98	1.00	0.58	0.75	0.54	0.93	0.30	0.42	0.66	0.81	0.25	0.11	0.17	0.27

Table 3
Powers of tests under moderate sparse alternative with $\alpha = 0.05$. Based on 1000 replications.

p	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400
	$K = 3$ with moderate sparse alternative																							
	Model 1				Model 2				Model 3				Model 4				Model 5				Model 6			
LRT	0.36	0.37	0.28	–	0.54	0.79	0.69	–	0.81	1.00	1.00	–	0.70	0.83	0.68	–	0.65	0.78	0.76	–	0.35	0.22	0.21	–
FHW	0.15	0.08	0.11	0.10	0.05	0.20	0.13	0.15	0.40	0.61	0.68	0.48	0.28	0.42	0.41	0.40	0.15	0.36	0.32	0.28	0.11	0.09	0.09	0.14
Sc	0.15	0.08	0.11	0.10	0.05	0.19	0.10	0.12	0.40	0.58	0.59	0.34	0.28	0.38	0.32	0.22	0.15	0.32	0.21	0.15	0.10	0.08	0.05	0.05
Sr	0.36	0.37	0.28	0.00	0.54	0.79	0.69	0.00	0.81	1.00	1.00	0.00	0.70	0.83	0.68	0.00	0.65	0.78	0.76	0.00	0.35	0.22	0.21	0.00
$\Psi_\alpha(\boldsymbol{I})$	0.19	0.04	0.17	0.18	0.11	0.15	0.17	0.23	0.28	0.52	0.54	0.40	0.34	0.44	0.38	0.58	0.13	0.26	0.27	0.17	0.06	0.11	0.07	0.11
$\Psi_\alpha(\boldsymbol{\Omega})$	0.25	0.32	0.53	0.69	0.91	1.00	1.00	0.96	0.85	1.00	1.00	1.00	0.76	0.88	0.87	0.96	0.57	0.81	0.95	0.98	0.38	0.23	0.22	0.47
$\Psi_\alpha(\widehat{\boldsymbol{\Omega}})$	–	–	–	–	–	–	–	–	0.84	1.00	1.00	1.00	0.78	0.87	0.91	0.96	0.57	0.82	0.95	0.99	0.37	0.24	0.25	0.49
	$K = 5$ with moderate sparse alternative																							
LRT	0.91	0.61	0.76	–	0.97	0.96	0.93	–	1.00	1.00	1.00	–	0.86	0.79	0.52	–	0.93	0.99	0.96	–	0.31	0.83	0.29	–
FHW	0.12	0.10	0.09	0.08	0.17	0.21	0.37	0.53	0.56	0.96	0.86	0.95	0.31	0.26	0.20	0.85	0.22	0.35	0.53	0.51	0.16	0.34	0.19	0.17
Sc	0.12	0.10	0.09	0.08	0.16	0.16	0.31	0.32	0.56	0.94	0.82	0.92	0.30	0.26	0.20	0.82	0.19	0.31	0.45	0.33	0.15	0.29	0.11	0.07
Sr	0.91	0.61	0.76	0.00	0.97	0.96	0.93	0.00	1.00	1.00	1.00	0.00	0.86	0.79	0.52	0.00	0.93	0.99	0.96	0.00	0.31	0.83	0.29	0.00
$\Psi_\alpha(\boldsymbol{I})$	0.15	0.20	0.26	0.43	0.18	0.12	0.45	0.28	0.54	0.78	0.68	0.74	0.36	0.37	0.40	0.76	0.20	0.37	0.33	0.19	0.11	0.24	0.13	0.13
$\Psi_\alpha(\boldsymbol{\Omega})$	0.65	0.63	0.81	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.92	0.94	1.00	0.80	1.00	1.00	1.00	0.32	0.66	0.28	0.36
$\Psi_\alpha(\widehat{\boldsymbol{\Omega}})$	–	–	–	–	–	–	–	–	1.00	1.00	1.00	1.00	0.96	0.93	0.92	1.00	0.81	1.00	1.00	1.00	0.36	0.68	0.27	0.37

Table 4
Powers of tests under non-sparse alternative with $\alpha = 0.05$. Based on 1000 replications.

p	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400	50	100	200	400
	$K = 3$ with non-sparse alternative																							
	Model 1				Model 2				Model 3				Model 4				Model 5				Model 6			
LRT	0.99	0.99	1.00	–	1.00	1.00	0.99	–	1.00	1.00	1.00	–	1.00	1.00	1.00	–	1.00	1.00	0.93	–	0.69	0.71	0.71	–
FHW	0.17	0.12	0.15	0.06	0.42	0.63	0.62	0.69	0.94	0.98	1.00	1.00	0.82	0.96	1.00	0.98	0.41	0.63	0.65	0.84	0.30	0.43	0.38	0.41
Sc	0.16	0.12	0.14	0.06	0.38	0.57	0.52	0.54	0.94	0.98	0.98	1.00	0.82	0.96	1.00	0.92	0.41	0.55	0.54	0.71	0.30	0.37	0.28	0.21
Sr	0.99	0.99	1.00	0.00	1.00	1.00	0.99	0.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00	0.93	0.00	0.69	0.71	0.71	0.00
$\Psi_\alpha(\boldsymbol{I})$	0.34	0.35	0.47	0.54	0.38	0.39	0.44	0.52	0.74	0.64	0.86	0.86	0.68	0.66	0.84	0.84	0.31	0.34	0.42	0.48	0.22	0.17	0.24	0.21
$\Psi_\alpha(\boldsymbol{\Omega})$	0.88	0.89	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.73	0.73	0.88	0.69
$\Psi_\alpha(\widehat{\boldsymbol{\Omega}})$	–	–	–	–	–	–	–	–	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.77	0.72	0.90	0.75
	$K = 5$ with non-sparse alternative																							
LRT	1.00	1.00	1.00	–	1.00	1.00	1.00	–	1.00	1.00	1.00	–	1.00	1.00	1.00	–	1.00	1.00	1.00	–	0.95	0.96	0.92	–
FHW	0.59	0.21	0.27	0.10	0.93	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	0.94	1.00	0.99	1.00	0.46	0.53	0.66	0.69
Sc	0.59	0.21	0.27	0.10	0.91	0.96	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.96	1.00	0.93	1.00	0.97	0.97	0.41	0.50	0.54	0.47
Sr	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	0.00	0.95	0.96	0.92	0.00
$\Psi_\alpha(\boldsymbol{I})$	0.83	0.45	0.63	0.62	0.59	0.66	0.87	0.75	1.00	0.98	0.98	1.00	0.84	0.82	0.86	0.94	0.59	0.78	0.73	0.85	0.23	0.30	0.26	0.21
$\Psi_\alpha(\boldsymbol{\Omega})$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.87	0.97	0.87
$\Psi_\alpha(\widehat{\boldsymbol{\Omega}})$	–	–	–	–	–	–	–	–	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.90	0.88	0.97	0.86



(a) Extreme sparse. (b) Moderate sparse. (c) Non-sparse.

Fig. 3. Plots of the comparisons of powers for all models.

the covariance matrix $\boldsymbol{\Sigma}$. The simulation results showed that the data-driven test $\Phi_\alpha(\hat{\boldsymbol{\Omega}})$ performs well when $\boldsymbol{\Sigma}$ is sparse. See [8] for further details on estimating covariance matrices and their inverse under the matrix ℓ_1 norm.

In the present paper, it is shown that the test $\Phi_\alpha(\boldsymbol{\Omega})$ outperforms $\Phi_\alpha(\boldsymbol{I})$ when testing against sparse alternatives. Similar comparison can be made between $\Phi_\alpha(\boldsymbol{\Omega})$ and $\Phi_\alpha(\boldsymbol{\Omega}^{1/2})$ as in [7]. The power of $\Phi_\alpha(\boldsymbol{\Omega})$ can be proved to dominate the power of $\Phi_\alpha(\boldsymbol{\Omega}^{1/2})$ as in Proposition 2, but under stronger conditions. For reasons of space, we omit the discussion in this paper.

We have focused on the Gaussian case in this paper. The results can be extended to non-Gaussian distributions. Let $\boldsymbol{X}_j$, $j = 1, \ldots, K$, be p-dimensional random vectors satisfying

$$\boldsymbol{X}_j = \boldsymbol{\mu}_j + \boldsymbol{U}_j,$$

where $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_K$ are independent and identical distributed random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})_{p\times p}$. Let $\boldsymbol{V}_j = \boldsymbol{\Omega}\boldsymbol{U}_j =: (V_{1j}, \ldots, V_{pj})^T$ for $j = 1, \ldots, K$. The results in Theorem 1, Proposition 1 and Theorem 4 still hold with the Gaussian assumption replaced by either of the following moment conditions.

- **(C6). (Sub-Gaussian-type tails)** Suppose that $\log p = o(n^{1/4})$. There exist some constants $\eta > 0$ and $C > 0$ such that
$$\mathsf{E}\exp(\eta V_{ij}^2/\omega_{ii}) \le C \quad \text{for } 1 \le i \le p, 1 \le j \le K.$$
- **(C7). (Polynomial-type tails)** Suppose that for some constants $\gamma_0, c_1 > 0$, $p \le c_1 n^{\gamma_0}$, and for some constants $\epsilon > 0$ and $C > 0$
$$\mathsf{E}|V_{ij}/\omega_{ii}^{\frac{1}{2}}|^{2\gamma_0+2+\epsilon} \le C \quad \text{for } 1 \le i \le p, 1 \le j \le K.$$

6. Proof of main results

We prove the main results in this section. We begin by collecting and proving in Section 6.1 a few technical lemmas that will be used in the proofs of the main theorems.

6.1. Technical lemmas

Lemma 1 (*Bonferroni Inequality*). *Let $A = \cup_{t=1}^{p} A_t$. For any $k < [p/2]$, we have*

$$\sum_{t=1}^{2k} (-1)^{t-1} E_t \le P(A) \le \sum_{t=1}^{2k-1} (-1)^{t-1} E_t,$$

where $E_t = \sum_{1\le i_1<\cdots<i_t\le p} P(A_{i_1} \cap \cdots \cap A_{i_t})$.

Lemma 2 (*Berman [3]*). *If X and Y have a bivariate normal distribution with expectation zero, unit variance and correlation coefficient ρ, then*

$$\lim_{c\to\infty} \frac{P(X > c, Y > c)}{[2\pi(1-\rho)^{1/2}c^2]^{-1} \exp\left(-\frac{c^2}{1+\rho}\right)(1+\rho)^{3/2}} = 1,$$

uniformly for all ρ such that $|\rho| \le \delta$, for any δ, $0 < \delta < 1$.

Lemma 3 (*Zolotarev [22]*). *Let $\boldsymbol{Y}$ be a nondegenerate Gaussian mean zero r.v. with covariance operator $\boldsymbol{\Sigma}$. Let σ^2 be the largest eigenvalue of $\boldsymbol{\Sigma}$ and d be the dimension of the corresponding eigenspace. Let σ_i^2, $1 \le i < d'$, be the positive eigenvalues of $\boldsymbol{\Sigma}$ arranged in a nonincreasing order and taking into account the multiplicities. Further, if $d' < \infty$, put $\sigma_i^2 = 0$, $i \ge d'$. Let $H(\boldsymbol{\Sigma}) := \prod_{i=d+1}^{\infty}(1-\sigma_i^2/\sigma^2)^{-1/2}$. Then for $y > 0$,*

$$P\{\|\boldsymbol{Y}\| > y\} \sim 2A\sigma^2 y^{d-2} \exp(-y^2/(2\sigma^2)), \quad \text{as } y \to \infty,$$

where $A := (2\sigma^2)^{-d/2}\Gamma^{-1}(d/2)H(\boldsymbol{\Sigma})$ with $\Gamma(\cdot)$ the gamma function.

Lemma 4. *For general positive definite matrix $\boldsymbol{A}$ and $(b_{i,j}) =: \boldsymbol{B} = \boldsymbol{A\Sigma A}$, suppose $C^{-1} \le \lambda_{\min}(\boldsymbol{A}) \le \lambda_{\max}(\boldsymbol{A}) \le C$ and $C^{-1} \le \lambda_{\min}(\boldsymbol{B}) \le \lambda_{\max}(\boldsymbol{B}) \le C$ for some constant $C > 0$ and $\boldsymbol{\Sigma}$ has all diagonal elements equal to 1. Then for p^r-sparse $\{\boldsymbol{\delta}^{(jl)}, 1 \le j < l \le K\}$, with $r < 1/4$ and nonzero locations of $\boldsymbol{\delta}^{(jl)}$ randomly and uniformly drawn from $\{1,\ldots,p\}$ for every $1 \le j < l \le K$, we have*

$$P\left(\max_{i\in H}\left|\frac{|\boldsymbol{\eta}_i|_2}{\sqrt{b_{ii}}} - \frac{a_{ii}}{\sqrt{b_{ii}}}|\boldsymbol{\delta}_i|_2\right| = O(p^{r-a/2})\max_{i\in H}|\boldsymbol{\delta}_i|_2\right) \to 1, \tag{13}$$

and

$$P\left(\max_{i\in H}\left|\frac{\eta_i^{(jl)}}{\sqrt{b_{ii}}}\frac{a_{ii}}{\sqrt{b_{ii}}}\delta_i^{(jl)}\right| = O(p^{r-a/2})\max_{i\in H}\delta_i^{(jl)}\right) \to 1, \tag{14}$$

for $1 \le j < l \le K$ and for any $2r < a < 1-2r$, as $p \to \infty$, where $\boldsymbol{\delta}_i = (\delta_i^{(12)}, \delta_i^{(13)}, \ldots, \delta_i^{(K-1K)})^T$ and $\boldsymbol{\eta}_i = ((\boldsymbol{A}\boldsymbol{\delta}^{(12)})_i, \ldots, (\boldsymbol{A}\boldsymbol{\delta}^{(K-1K)})_i)^T$ for $i \in H := \{1 \le i \le p : \delta_i^{(jl)} \ne 0 \text{ for some } 1 \le j < l \le K\} = \{i_1, \ldots, i_m\}$.

Proof of Lemma 4. We only need to prove (13) because the proof of (14) is similar. We re-order $a_{i1}, \ldots, a_{ip}$ as $|a_{i(1)}| \ge \cdots \ge |a_{i(p)}|$. Let a satisfy $2r < a < 1-2r$ with $r < 1/4$. Define $\mathcal{I} = \{1 \le i_1 < \cdots < i_m \le p\}$ and

$$\mathcal{I}_0 = \{1 \le i_1 < \cdots < i_m \le p : \text{ there exist some } 1 \le k \le m \text{ and some } j \ne k \text{ and } 1 \le j \le m, \text{ such that } |a_{i_k i_j} \ge |a_{i_k(p^a)}|\}.$$

We can show that

$$|\mathcal{I}_0|/|\mathcal{I}| = O\left(p \cdot p^a \binom{p}{p^r - 2}\right) \Big/ \binom{p}{p^r}.$$

So we have $\frac{|\mathcal{I}_0|}{|\mathcal{I}|} = O(p^{a+2r-1}) = o(1)$. For $1 \le t \le m$, write

$$\sum_{1\le j<l\le K} (A\boldsymbol{\delta}^{(jl)})_{i_t}^2 = \sum_{1\le j<l\le K}\left(\sum_{k=1}^{p} a_{i_t k}\delta_k^{(jl)}\right)^2 = \sum_{1\le j<l\le K}\left(a_{i_t i_t}\delta_{i_t}^{(jl)} + \sum_{q=1,q\ne t}^{m} a_{i_t i_q}\delta_{i_q}^{(jl)}\right)^2.$$

So we have

$$|\boldsymbol{\eta}_{i_t}|_2 = \left|a_{i_t i_t}\boldsymbol{\delta}_{i_t} + \sum_{q=1,q\ne t}^{m} a_{i_t i_q}\boldsymbol{\delta}_{i_q}\right|_2 \ge |a_{i_t i_t}\boldsymbol{\delta}_{i_t}|_2 - \left|\sum_{q=1,q\ne t}^{m} a_{i_t i_q}\boldsymbol{\delta}_{i_q}\right|_2,$$

and $|\boldsymbol{\eta}_{i_t}|_2 \le |a_{i_t i_t}\boldsymbol{\delta}_{i_t}|_2 + |\sum_{q=1,q\neq t}^m a_{i_t i_q}\boldsymbol{\delta}_{i_q}|_2$. Note that for any $(i_1,\dots,i_m) \in \mathcal{I}_0^c$,

$$\sum_{q=1,q\neq t}^m |a_{i_t i_q}| \le p^r\sqrt{\frac{C_1}{p^a}}.$$

It follows that for $H \in \mathcal{I}_0$ and $i \in H$,

$$\frac{|\boldsymbol{\eta}_i|_2}{\sqrt{b_{ii}}} = \frac{a_{ii}}{\sqrt{b_{ii}}}|\boldsymbol{\delta}_i|_2 + O(p^{r-a/2})\max_{i\in H}|\boldsymbol{\delta}_i|_2.$$

So the lemma is proved. ■

6.2. Proof of Theorem 1

Without loss of generality, we assume $\sigma_{ii} = 1$ for $i = 1,\dots,p$ throughout the proof. Let $\boldsymbol{Y}_i = (\sqrt{\frac{n_1n_2}{n_1+n_2}}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2)_i, \sqrt{\frac{n_1n_3}{n_1+n_3}}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_3)_i, \dots, \sqrt{\frac{n_{K-1}n_K}{n_{K-1}+n_K}}(\bar{\boldsymbol{X}}_{K-1} - \bar{\boldsymbol{X}}_K)_i)^T$. Let $\boldsymbol{\Sigma}_0$ be the $b \times b$ covariance matrix of $\boldsymbol{Y}_i := (Y_{1i},\dots,Y_{bi})^T$ for $i = 1,\dots,p$, where $b = \frac{K(K-1)}{2}$. Let $M_n = \max_{1\le i\le p}|\boldsymbol{Y}_i|_2^2$. Then it is enough to prove the following lemma.

Lemma 5. *Suppose that* $\max_{1\le i\neq j\le p}|\sigma_{ij}| \le r < 1$ *and* $C_0^{-1} \le \lambda_{\min}(\boldsymbol{\Sigma}) \le \lambda_{\max}(\boldsymbol{\Sigma}) \le C_0$. *We have*

$$P\left(M_n - 2\sigma^2\log p - (d-2)\sigma^2\log\log p \le x\right) \to \exp\left(-\Gamma^{-1}\left(\frac{d}{2}\right)H(\boldsymbol{\Sigma})\exp(-x/2\sigma^2)\right). \tag{15}$$

Proof. Set $x_p = \sqrt{2\sigma^2\log p + (d-2)\sigma^2\log\log p + x}$. By Lemma 1, we have for any fixed $m \le [p/2]$,

$$\sum_{t=1}^{2m}(-1)^{t-1}E_t \le P\left(\max_{1\le i\le p}|\boldsymbol{Y}_i|_2 \ge x_p\right) \le \sum_{t=1}^{2m-1}(-1)^{t-1}E_t, \tag{16}$$

where

$$E_t = \sum_{1\le i_1<\cdots<i_t\le p} P\left(|\boldsymbol{Y}_{i_1}|_2 \ge x_p,\dots,|\boldsymbol{Y}_{i_t}|_2 \ge x_p\right) =: \sum_{1\le i_1<\cdots<i_t\le p} P_{i_1,\dots,i_t}.$$

Then it suffices to show that

$$\sum_{1\le i_1<\cdots<i_t\le p} P_{i_1,\dots,i_t} = (1+o(1))\frac{1}{t!}\Gamma^{-t}\left(\frac{d}{2}\right)H^t(\boldsymbol{\Sigma})\exp\left(-\frac{tx}{2\sigma^2}\right). \tag{17}$$

When $t = 1$, by Lemma 3, we have

$$\sum_{1\le i_1\le p} P_{i_1} = (1+o(1))\Gamma^{-1}\left(\frac{d}{2}\right)H(\boldsymbol{\Sigma})\exp\left(-\frac{x}{2\sigma^2}\right).$$

This implies (17). It remains to prove the lemma when $t \ge 2$. Let $\gamma > 0$ be a sufficiently small number which will be specified later. Define

$$\mathcal{I} = \left\{1 \le i_1 < \cdots < i_t \le p : \max_{1\le k<l\le t}|\sigma_{i_k i_l}| \ge p^{-\gamma}\right\}.$$

For $d = 1$, define

$$\mathcal{I}_1 = \left\{1 \le i_1 < \cdots < i_t \le p : |\sigma_{i_k i_l}| \ge p^{-\gamma} \text{ for every } 1 \le k < l \le t\right\}.$$

So when $t = 2$, we have $\mathcal{I} = \mathcal{I}_1$. For $2 \le d \le t-1$ and $t \ge 3$, define

$$\mathcal{I}_d = \{1 \le i_1 < \cdots < i_t \le p : \text{the cardinality of } S \text{ is } d, \text{ where } S \text{ is the largest subsetof } \{i_1,\dots,i_t\} \text{ such that } \forall k \neq l \in S, |\sigma_{i_k i_l}| < p^{-\gamma}\right\}.$$

So we have $\mathcal{I} = \cup_{d=1}^{t-1}\mathcal{I}_d$ for $t \ge 2$. Let $\text{Card}(\mathcal{I}_d)$ denote the total number of the vectors $(i_1,\dots,i_t)$ in $\mathcal{I}_d$. We can show that $\text{Card}(\mathcal{I}_d) \le Cp^{d+2\gamma t}$. In fact, the total number of the subsets of $\{i_1,\dots,i_t\}$ with cardinality d is $\binom{p}{d}$. For a fixed subset

S with cardinality d, the number of i such that $|\sigma_{i_k i_l}| \geq p^{-\gamma}$ for some $j \in S$ is no more than $Cdp^{2\gamma}$. This implies that $\text{Card}(\mathcal{I}_d) \leq Cp^{d+2\gamma t}$. Define $\mathcal{I}^c = \{1 \leq i_1 < \cdots < i_t \leq p\} \setminus \mathcal{I}$. Then the number of elements in the sum $\sum_{(i_1,\ldots,i_t)\in\mathcal{I}^c} P_{i_1,\ldots,i_t}$ is $\binom{p}{t} - O(\sum_{d=1}^{t-1} p^{d+2\gamma t}) = \binom{p}{t} - O(p^{t-1+2\gamma t}) = (1+o(1))\binom{p}{t}$.

To prove Lemma 5, it suffices to show that

$$P_{i_1,\ldots,i_t} = (1+o(1))\Gamma^{-t}\left(\frac{d}{2}\right) H^t(\mathbf{\Sigma}) p^{-t} \exp\left(-\frac{tx}{2\sigma^2}\right) \tag{18}$$

uniformly in $(i_1,\ldots,i_t) \in \mathcal{I}^c$, and for $1 \leq d \leq t-1$,

$$\sum_{(i_1,\ldots,i_t)\in\mathcal{I}_d} P_{i_1,\ldots,i_t} \to 0. \tag{19}$$

By submitting (18) and (19) into (16), we obtain that

$$(1+o(1))S_{2m} \leq P\left(\max_{1\leq i\leq p} |\mathbf{Y}_i|_2 \geq x_p\right) \leq (1+o(1))S_{2m-1}, \tag{20}$$

where $S_m = \sum_{t=1}^m (-1)^{t-1}\frac{1}{t!}\Gamma^{-t}(\frac{d}{2})H^t(\mathbf{\Sigma})\exp(-\frac{tx}{2\sigma^2})$. Note that

$$\lim_{m\to\infty} S_m = 1 - \exp\left(-\Gamma^{-1}\left(\frac{d}{2}\right) H(\mathbf{\Sigma}) \exp(-x/2\sigma^2)\right).$$

By letting $p \to \infty$ first and then $m \to \infty$ in (20), we prove Lemma 5.

First we prove (18). Let $\tilde{\mathbf{Y}} = (\mathbf{Y}_{i_1}^T, \ldots, \mathbf{Y}_{i_t}^T)^T$ and $(\mathbf{Z}_{i_1}^T, \ldots, \mathbf{Z}_{i_t}^T)^T =: \mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_{bt\times bt})$, where $b = \frac{K(K-1)}{2}$, $\mathbf{Z}_{i_j} = (Z_{1i_j}, \ldots, Z_{bi_j})^T$ for $j = 1, \ldots, t$ and $\tilde{\mathbf{Y}}$ and $\mathbf{Z}$ are independent. Let $|\tilde{\mathbf{Y}}|_t = \min_{1\leq j\leq t} |\mathbf{Y}_{i_j}|_2$ and let $\lambda_p = Cp^{-\gamma/4}$ for some constant $C > 0$. Then we have

$$\begin{aligned}
P_{i_1,\ldots,i_t} &= P(|\tilde{\mathbf{Y}}|_t \geq x_p) \\
&\leq P\left(|\tilde{\mathbf{Y}} + \lambda_p \mathbf{Z}|_t \geq x_p - \lambda_p \max_{1\leq j\leq t} |\mathbf{Z}_{i_j}|_2\right) \\
&\leq \frac{1}{(2\pi)^{bt/2}\det(\mathbf{\Sigma}_1 + \lambda_p \mathbf{I})^{1/2}} \int_{|\mathbf{z}|_t \geq x_p - Cp^{-\gamma/8}} \exp\left(-\frac{1}{2}\mathbf{z}^T(\mathbf{\Sigma}_1 + \lambda_p\mathbf{I})^{-1}\mathbf{z}\right) d\mathbf{z} \\
&\quad + P\left(\lambda_p \max_{1\leq j\leq t} |\mathbf{Z}_{i_j}|_2 \geq Cp^{-\gamma/8}\right) \\
&\leq \frac{1}{(2\pi)^{bt/2}\det(\mathbf{\Sigma}_1 + \lambda_p \mathbf{I})^{1/2}} \int_{|\mathbf{z}|_t \geq x_p - Cp^{-\gamma/8}} \exp\left(-\frac{1}{2}\mathbf{z}^T(\mathbf{\Sigma}_1 + \lambda_p\mathbf{I})^{-1}\mathbf{z}\right) d\mathbf{z} + O(p^{-2t}),
\end{aligned} \tag{21}$$

where $\mathbf{z} \in \mathbb{R}^{bt}$ and $\mathbf{\Sigma}_1$ is the covariance matrix of $\tilde{\mathbf{Y}}$ and C is a constant. Let $\widetilde{\mathbf{\Sigma}}$ be a $bt \times bt$ matrix with $\widetilde{\mathbf{\Sigma}}_{jb+1:(j+1)b,jb+1:(j+1)b} = \mathbf{\Sigma}_0$ for $j = 0, \ldots, t-1$ and $\widetilde{\mathbf{\Sigma}}_{ij} = 0$ otherwise. For $(i_1,\ldots,i_t) \in \mathcal{I}^c$, we have $\mathbf{\Sigma}_{1jb+1:(j+1)b,jb+1:(j+1)b} = \mathbf{\Sigma}_0$ for $j = 0, \ldots, t-1$ and $|\mathbf{\Sigma}_{1ij}| < p^{-\gamma}$ otherwise. Write

$$\begin{aligned}
\int_{|\mathbf{z}|_t \geq x_p - Cp^{-\gamma/8}} \exp\left(-\frac{1}{2}\mathbf{z}^T(\mathbf{\Sigma}_1 + \lambda_p\mathbf{I})^{-1}\mathbf{z}\right) d\mathbf{z} &= \int_{|\mathbf{z}|_t \geq x_p - Cp^{-\gamma/8}, \|\mathbf{z}\|^2 \geq (\log p)^2} \exp\left(-\frac{1}{2}\mathbf{z}^T(\mathbf{\Sigma}_1 + \lambda_p\mathbf{I})^{-1}\mathbf{z}\right) d\mathbf{z} \\
&\quad + \int_{|\mathbf{z}|_t \geq x_p - Cp^{-\gamma/8}, \|\mathbf{z}\|^2 \leq (\log p)^2} \exp\left(-\frac{1}{2}\mathbf{z}^T(\mathbf{\Sigma}_1 + \lambda_p\mathbf{I})^{-1}\mathbf{z}\right) d\mathbf{z}.
\end{aligned} \tag{22}$$

Because $\lambda_{\max}(\mathbf{\Sigma}_1 + \lambda_p\mathbf{I}) \leq \lambda_{\max}(\mathbf{\Sigma}_0) + O(p^{-\gamma/4}) \leq M$ by some constant $M > 0$, we can get

$$\int_{|\mathbf{z}|_t \geq x_p - Cp^{-\gamma/8}, \|\mathbf{z}\|^2 \geq (\log p)^2} \exp\left(-\frac{1}{2}\mathbf{z}^T(\mathbf{\Sigma}_1 + \lambda_p\mathbf{I})^{-1}\mathbf{z}\right) d\mathbf{z} \leq C\exp(-(\log p)^2/2bt) \leq Cp^{-2bt}, \tag{23}$$

uniformly in $(i_1,\ldots,i_t) \in \mathcal{I}^c$. For the second part of the sum in (22), note that

$$\|(\mathbf{\Sigma}_1 + \lambda_p\mathbf{I})^{-1} - (\widetilde{\mathbf{\Sigma}} + \lambda_p\mathbf{I})^{-1}\|_2 \leq C\lambda_p^{-2}p^{-\gamma} \leq Cp^{-\gamma/2}, \tag{24}$$

we can obtain that

$$\begin{aligned}
&\int_{|\boldsymbol{z}|_t \geq x_p - Cp^{-\gamma/8}, \|\boldsymbol{z}\|^2 \leq (\log p)^2} \exp\left(-\frac{1}{2}\boldsymbol{z}^T(\boldsymbol{\Sigma}_1 + \lambda_p \boldsymbol{I})^{-1}\boldsymbol{z}\right) d\boldsymbol{z} \\
&= \int_{|\boldsymbol{z}|_t \geq x_p - Cp^{-\gamma/8}, \|\boldsymbol{z}\|^2 \leq (\log p)^2} \exp\left(-\frac{1}{2}\boldsymbol{z}^T((\boldsymbol{\Sigma}_1 + \lambda_p \boldsymbol{I})^{-1} - (\widetilde{\boldsymbol{\Sigma}} + \lambda_p \boldsymbol{I})^{-1})\boldsymbol{z} - \frac{1}{2}\boldsymbol{z}^T(\widetilde{\boldsymbol{\Sigma}} + \lambda_p \boldsymbol{I})^{-1}\boldsymbol{z}\right) d\boldsymbol{z} \\
&= (1 + O(p^{-\gamma/2}(\log p)^2)) \int_{|\boldsymbol{z}|_t \geq x_p - Cp^{-\gamma/8}, \|\boldsymbol{z}\|^2 \leq (\log p)^2} \exp\left(-\frac{1}{2}\boldsymbol{z}^T(\widetilde{\boldsymbol{\Sigma}} + \lambda_p \boldsymbol{I})^{-1}\boldsymbol{z}\right) d\boldsymbol{z} \\
&= (1 + O(p^{-\gamma/2}(\log p)^2)) \int_{|\boldsymbol{z}|_t \geq x_p - Cp^{-\gamma/8}} \exp\left(-\frac{1}{2}\boldsymbol{z}^T(\widetilde{\boldsymbol{\Sigma}} + \lambda_p \boldsymbol{I})^{-1}\boldsymbol{z}\right) d\boldsymbol{z} + O(p^{-2bt}) \\
&= (1 + O(p^{-\gamma/2}(\log p)^2)) \left(\int_{\|\boldsymbol{z}_{i_1}\|_2 \geq x_p - Cp^{-\gamma/8}} \exp\left(-\frac{1}{2}\boldsymbol{z}_{i_1}^T(\boldsymbol{\Sigma}_0 + \lambda_p \boldsymbol{I})^{-1}\boldsymbol{z}_{i_1}\right) d\boldsymbol{z}_{i_1}\right)^t + O(p^{-2bt}),
\end{aligned} \tag{25}$$

where $\boldsymbol{z}_{i_1} \in \mathbb{R}^b$. So for $(i_1, \ldots, i_t) \in \mathcal{I}^c$, we have

$$\begin{aligned}
P_{i_1,\ldots,i_t} &\leq (1 + O(p^{-\gamma/2}(\log p)^2)) \left(P(|\boldsymbol{Y}_{i_1} + \lambda_p \boldsymbol{Z}_{i_1}|_2 \geq x_p - Cp^{-\gamma/8})\right)^t + Cp^{-2t} \\
&= (1 + o(1)) \left(P(|\boldsymbol{Y}_{i_1}|_2 \geq x_p)\right)^t + Cp^{-2t} \\
&= (1 + o(1))\Gamma^{-t}\left(\frac{d}{2}\right) H^t(\boldsymbol{\Sigma}) p^{-t} \exp\left(-\frac{tx}{2\sigma^2}\right),
\end{aligned} \tag{26}$$

where the last equation comes from Lemma 3. Similarly, because

$$P(|\tilde{\boldsymbol{Y}}|_t \geq x_p) \geq P\left(|\tilde{\boldsymbol{Y}} + \lambda_p \boldsymbol{Z}|_t \geq x_p + \lambda_p \max_{1 \leq j \leq t} |\boldsymbol{Z}_{i_j}|_2\right), \tag{27}$$

we can get

$$P_{i_1,\ldots,i_t} \geq (1 - o(1))\Gamma^{-t}\left(\frac{d}{2}\right) H^t(\boldsymbol{\Sigma}) p^{-t} \exp\left(-\frac{tx}{2\sigma^2}\right). \tag{28}$$

So (18) is proved.

It remains to prove (19). For $S \subset \mathcal{I}_d$ with $d \geq 1$, without loss of generality, we can assume $S = \{i_{t-d+1}, \ldots, i_t\}$. By the definition of S and $\mathcal{I}_d$, for any $k \in \{i_1, \ldots, i_{t-d}\}$, there exists at least one $l \in S$ such that $|\sigma_{kl}| \geq p^{-\gamma}$. We divide $\mathcal{I}_d$ into two parts:

$$\begin{aligned}
\mathcal{I}_{d,1} &= \{1 \leq i_1 < \cdots < i_t \leq p : \text{there exists an } k \in \{i_1, \ldots, i_{t-d}\} \text{ such thatfor some } l_1, l_2 \in S \text{ with} \\
&\qquad l_1 \neq l_2, |\sigma_{kl_1}| \geq p^{-\gamma} \text{ and } |\sigma_{kl_2}| \geq p^{-\gamma}\}, \\
\mathcal{I}_{d,2} &= \mathcal{I}_d \setminus \mathcal{I}_{d,1}.
\end{aligned}$$

Clearly, $\mathcal{I}_{1,1} = \emptyset$ and $\mathcal{I}_{1,2} = \mathcal{I}_1$. Moreover, we can show that Card($\mathcal{I}_{d,1}$)$\leq Cp^{d-1+2\gamma t}$. Similarly as proved in (21) and (26)–(28), for any $(i_1, \ldots, i_t) \in \mathcal{I}_{d,1}$,

$$P\left(|\boldsymbol{Y}_{i_1}|_2 \geq x_p, \ldots, |\boldsymbol{Y}_{i_t}|_2 \geq x_p\right) \leq P\left(|\boldsymbol{Y}_{i_{t-d+1}}|_2 \geq x_p, \ldots, |\boldsymbol{Y}_{i_t}|_2 \geq x_p\right) = O(p^{-d}).$$

Hence by letting γ be sufficiently small,

$$\sum_{\mathcal{I}_{d,1}} P_{i_1,\ldots,i_t} \leq Cp^{-1+2\gamma t} = o(1). \tag{29}$$

For any $(i_1, \ldots, i_t) \in \mathcal{I}_{d,2}$, without loss of generality, we assume that $|\sigma_{i_1,i_{t-d+1}}| \geq p^{-\gamma}$. Note that

$$P\left(|\boldsymbol{Y}_{i_1}|_2 \geq x_p, \ldots, |\boldsymbol{Y}_{i_t}|_2 \geq x_p\right) \leq P\left(|\boldsymbol{Y}_{i_1}|_2 \geq x_p, |\boldsymbol{Y}_{i_{t-d+1}}|_2 \geq x_p, \ldots, |\boldsymbol{Y}_{i_t}|_2 \geq x_p\right).$$

Let $\boldsymbol{W}_l$ be the covariance matrix of $(\boldsymbol{Y}_{i_1}^T, \boldsymbol{Y}_{i_{t-d+1}}^T, \ldots, \boldsymbol{Y}_{i_t}^T)^T$. We can show that $\|\boldsymbol{W}_l - \bar{\boldsymbol{W}}_l\|_2 = O(p^{-\gamma})$, where $\bar{\boldsymbol{W}}_l = diag(\boldsymbol{D}, \widetilde{\boldsymbol{\Sigma}}_{(t-d)b+1:tb,(t-d)b+1:tb})$ and $\boldsymbol{D}$ is the covariance matrix of $(\boldsymbol{Y}_{i_1}^T, \boldsymbol{Y}_{i_{t-d+1}}^T)^T$. Using the similar arguments as in (22)–(25), we can get

$$P\left(|\boldsymbol{Y}_{i_1}|_2 \geq x_p, \ldots, |\boldsymbol{Y}_{i_t}|_2 \geq x_p\right) \leq (1 + o(1))P(|\boldsymbol{Y}_{i_1}|_2 \geq x_p, |\boldsymbol{Y}_{i_{t-d+1}}|_2 \geq x_p) \times O(p^{-d+1}) + O(p^{-2t}).$$

Define a set $A = \{-1, -1 + p^{-\alpha}, -1 + 2p^{-\alpha}, \dots, -1 + 2[p^{\alpha}]p^{-\alpha}, 1\}$, where α is a constant that will be specified later and $\lfloor p^{\alpha} \rfloor$ is the largest integer no larger than p^{α}. Because $|\boldsymbol{Y}_{i_1}|_2 = \sup_{|\boldsymbol{z}|_2=1} |\boldsymbol{Y}_{i_1}^T \boldsymbol{z}|$, we have

$$
\begin{aligned}
P(|\boldsymbol{Y}_{i_1}|_2 \ge x_p, |\boldsymbol{Y}_{i_{t-d+1}}|_2 \ge x_p) &= P\left(\sup_{|\boldsymbol{z}|_2=1} |\boldsymbol{Y}_{i_1}^T \boldsymbol{z}| \ge x_p, \sup_{|\boldsymbol{z}|_2=1} |\boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}| \ge x_p \right) \\
&\le P\left(\max_{z_i \in A, |\boldsymbol{z}|_2=1} |\boldsymbol{Y}_{i_1}^T \boldsymbol{z}| \ge x_p - C \max_{1 \le j \le b} |Y_{ji_1}| p^{-\alpha}, \max_{z_i \in A, |\boldsymbol{z}|_2=1} |\boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}| \ge x_p - C \max_{1 \le j \le b} |Y_{ji_{t-d+1}}| p^{-\alpha} \right) \\
&\le P\left(\max_{z_i \in A, |\boldsymbol{z}|_2=1} |\boldsymbol{Y}_{i_1}^T \boldsymbol{z}| \ge x_p - Cp^{-\alpha/2}, \max_{z_i \in A, |\boldsymbol{z}|_2=1} |\boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}| \ge x_p - Cp^{-\alpha/2} \right) + O(p^{-2t}) \\
&\le (1 + o(1))Cp^{b\alpha} \max_{z_i^{(j)} \in A, |\boldsymbol{z}^{(j)}|_2=1} P(|\boldsymbol{Y}_{i_1}^T \boldsymbol{z}^{(1)}| \ge x_p, |\boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}^{(2)}| \ge x_p) + O(p^{-2t}) \\
&\le (1 + o(1))Cp^{b\alpha} \max_{z_i^{(j)} \in A, |\boldsymbol{z}^{(j)}|_2=1} P\left(|x_1| \ge x_p / \sqrt{\mathrm{Var}(\boldsymbol{Y}_{i_1}^T \boldsymbol{z}^{(1)})}, |x_2| \ge x_p / \sqrt{\mathrm{Var}(\boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}^{(2)})} \right) + O(p^{-2t})
\end{aligned}
$$

for $i = 1, \dots, b$ and $j = 1, 2$, where $x_1 = \boldsymbol{Y}_{i_1}^T \boldsymbol{z}^{(1)} / \sqrt{\mathrm{Var}(\boldsymbol{Y}_{i_1}^T \boldsymbol{z}^{(1)})} \sim N(0, 1)$ and $x_2 = \boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}^{(2)} / \sqrt{\mathrm{Var}(\boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}^{(2)})} \sim N(0, 1)$ and

$$
\mathrm{Cov}(x_1, x_2) = \frac{\mathrm{Cov}(\boldsymbol{Y}_{i_1}^T \boldsymbol{z}^{(1)}, \boldsymbol{Y}_{t-d+1}^T \boldsymbol{z}^{(2)})}{\sqrt{\mathrm{Var}(\boldsymbol{Y}_{i_1}^T \boldsymbol{z}^{(1)}) \mathrm{Var}(\boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}^{(2)})}}.
$$

Because

$$
\mathrm{Var}(\boldsymbol{Y}_{i_1}^T \boldsymbol{z}^{(1)}) = \sum_{1 \le j,l \le b} \mathrm{Cov}(Y_{ji_1} z_j^{(1)}, Y_{li_1} z_l^{(1)}) = \sum_{1 \le j,l \le b} \xi_{jl} z_j^{(1)} z_l^{(1)},
$$

and

$$
\mathrm{Var}(\boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}^{(2)}) = \sum_{1 \le j,l \le b} \mathrm{Cov}(Y_{ji_{t-d+1}} z_j^{(2)}, Y_{li_1} z_l^{(2)}) = \sum_{1 \le j,l \le b} \xi_{jl} z_j^{(2)} z_l^{(2)},
$$

where $\xi_{jl} = \mathrm{Cov}(Y_{ji_1}, Y_{li_1})$, then we have

$$
\begin{aligned}
\sqrt{\mathrm{Var}(\boldsymbol{Y}_{i_1}^T \boldsymbol{z}^{(1)}) \mathrm{Var}(\boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}^{(2)})} &= \sqrt{\sum_{1 \le j,l \le b} \xi_{jl} z_j^{(1)} z_l^{(1)} \sum_{1 \le j,l \le b} \xi_{jl} z_j^{(2)} z_l^{(2)}} \\
&= \sqrt{\sum_{1 \le j,l \le b} \xi_{jl} z_j^{(1)} z_l^{(2)} \sum_{1 \le j,l \le b} \xi_{jl} z_l^{(1)} z_j^{(2)}} \\
&= \sum_{1 \le j,l \le b} \xi_{jl} z_j^{(1)} z_l^{(2)}.
\end{aligned}
$$

Also we have

$$
\mathrm{Cov}(\boldsymbol{Y}_{i_1}^T \boldsymbol{z}^{(1)}, \boldsymbol{Y}_{t-d+1}^T \boldsymbol{z}^{(2)}) = \sum_{1 \le j,l \le b} \mathrm{Cov}(Y_{ji_1} z_j^{(1)}, Y_{li_{t-d+1}} z_l^{(2)}) = \sum_{1 \le j,l \le b} r_{i_1 i_{t-d+1}} \xi_{jl} z_j^{(1)} z_l^{(2)},
$$

so we get $\mathrm{Cov}(x_1, x_2) = r_{i_1 i_{t-d+1}}$. In addition, $\mathrm{Var}(\boldsymbol{Y}_{i_1}^T \boldsymbol{z}^{(1)}) \le \lambda_{\max}(\boldsymbol{\Sigma}_0) = \sigma^2$ and $\mathrm{Var}(\boldsymbol{Y}_{i_{t-d+1}}^T \boldsymbol{z}^{(2)}) \le \lambda_{\max}(\boldsymbol{\Sigma}_0) = \sigma^2$, we have

$$
P(|\boldsymbol{Y}_{i_1}|_2 \ge x_p, |\boldsymbol{Y}_{i_{t-d+1}}|_2 \ge x_p) \le (1 + o(1))Cp^{b\alpha} P(|x_1| \ge x_p/\sigma, |x_2| \ge x_p/\sigma) + O(p^{-2t}).
$$

Thus, by Lemma 2 and the assumption $\max_{1 \le i \ne j \le p} |r_{ij}| \le r < 1$, for any $(i_1, \dots, i_t) \in \mathcal{I}_{d,2}$, we have

$$
P\left(|\boldsymbol{Y}_{i_1}|_2 \ge x_p, \dots, |\boldsymbol{Y}_{i_t}|_2 \ge x_p \right) \le (1 + o(1))4Cp^{b\alpha} p^{-\frac{2}{1+r}} \times O(p^{-d+1}).
$$

Thus by letting γ and α be sufficiently small,

$$
\sum_{\mathcal{I}_{d,2}} P_{i_1,\dots,i_t} \le (1 + o(1))4Cp^{d+2\gamma t+b\alpha-d+1-\frac{2}{1+r}} = o(1). \tag{30}
$$

Combining (29) and (30), we prove (19). The proof of Lemma 5 is complete. ■

6.3. Proof of Theorem 2

It suffices to prove $P\left(\max_{1\le i\le p}||\boldsymbol{\eta}_i|_2/\sqrt{b_{ii}}|\ge\sqrt{(2\sigma^2+\varepsilon/2)\log p}\right)\to 1$. By Lemma 4 and the condition $\max_i|\boldsymbol{\delta}_i|_2/\sigma_{ii}^{\frac{1}{2}}\ge\sqrt{2\sigma^2\beta\log p}$ with $\beta\ge 1/(\min_i\sigma_{ii}a_{ii})+\varepsilon$ for some constant $\varepsilon>0$, we can get $\max_{1\le i\le p}|\boldsymbol{\eta}_i|_2/\sqrt{\omega_{ii}}\ge\sqrt{(2\sigma^2+\varepsilon/2)\log p}$ with probability tending to one. So Theorem 2 follows. ■

6.4. Proof of Theorem 4

We only prove part (ii) of Theorem 4 in this section, part (i) follows from the proof of part (ii) directly. Without loss of generality, we assume that $\sigma_{ii}=1$ for $1\le i\le p$. Let $\boldsymbol{Y}_i=(\sqrt{\frac{n_1n_2}{n_1+n_2}}(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2)_i,\sqrt{\frac{n_1n_3}{n_1+n_3}}(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_3)_i,\ldots,\sqrt{\frac{n_{K-1}n_K}{n_{K-1}+n_K}}(\bar{\boldsymbol{X}}_{K-1}-\bar{\boldsymbol{X}}_K)_i)^T$, and let $\boldsymbol{Z}_i=\frac{1}{\sqrt{b_{i,i}}}\left(\sqrt{\frac{n_1n_2}{n_1+n_2}}(\boldsymbol{A}(\bar{\boldsymbol{X}}_1-\bar{\boldsymbol{X}}_2))_i,\ldots,\sqrt{\frac{n_{K-1}n_K}{n_{K-1}+n_K}}(\boldsymbol{A}(\bar{\boldsymbol{X}}_{K-1}-\bar{\boldsymbol{X}}_K))_i\right)^T$. Let $H=\{1\le i\le p:\ \delta_i^{(jl)}\ne 0$ for some $1\le j<l\le K\}=\{l_1,\ldots,l_m\}$. Define the event $\boldsymbol{G}=\{\max_{1\le i\le p}|\boldsymbol{\delta}_i|_2\le 8\sqrt{\sigma^2\log p}\}$. We first prove the following two lemmas.

Lemma 6. (i) *Suppose* (C1) *and* (C2) *hold. Then under* H_1 *with* $r<1/6$, *we have*

$$P(\Phi_\alpha(\boldsymbol{I})=1,\boldsymbol{G})=\alpha P(\boldsymbol{G})+(1-\alpha)P(\boldsymbol{E}^c,\boldsymbol{G})+o(1),\tag{31}$$

where $\boldsymbol{E}=\{\max_{i\in H}|\boldsymbol{Y}_i|_2<x_p\}$, *and*

$$P(\boldsymbol{E}^c,\boldsymbol{G})=I\{\boldsymbol{G}\}-I\{\boldsymbol{G}\}\prod_{i\in H}\left(1-P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}_i|_2\ge x_p\right)\right)+o(1).$$

(ii) *Suppose* (C1) *and* (C3) *hold. Then under* H_1 *with* $r<1/6$, *we have*

$$P(\Phi_\alpha(\boldsymbol{\Omega})=1,\boldsymbol{G})=\alpha P(\boldsymbol{G})+(1-\alpha)P(\tilde{\boldsymbol{E}}^c,\boldsymbol{G})+o(1),\tag{32}$$

where $\tilde{\boldsymbol{E}}=\{\max_{i\in H}|\boldsymbol{Z}_i|_2<x_p\}$, *and*

$$P(\tilde{\boldsymbol{E}}^c,\boldsymbol{G})=I\{\boldsymbol{G}\}-I\{\boldsymbol{G}\}\prod_{i\in H}\left(1-P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(|\boldsymbol{Z}_i|_2\ge x_p\right)\right)+o(1).$$

Lemma 7. *Let* $a_p=o((\log p)^{-1/2})$. *We have*

$$\max_{1\le k\le p^r}\left|P\left(\max_{1\le i\le k}|\boldsymbol{Y}_i|_2\ge x_p+a_n\right)-P\left(\max_{1\le i\le k}|\boldsymbol{Y}_i|_2\ge x_p\right)\right|=o(1)\tag{33}$$

uniformly in the means $\boldsymbol{\delta}_i$, $1\le i\le p$, *where* $x_p=\sqrt{2\sigma^2\log p+(d-2)\sigma^2\log\log p+q_\alpha}$, $r<1/6$ *and* $\boldsymbol{Y}_i, i\in H$ *are independent normal random vectors with covariance matrix* $\boldsymbol{\Sigma}_0$.

Proof of Lemma 6. To prove (31) and (32), we only need to prove

$$P(\Phi_\alpha(\boldsymbol{I})=1,\boldsymbol{G})\le\alpha P(\boldsymbol{G})+(1-\alpha)P(\boldsymbol{E}^c,\boldsymbol{G})+o(1),$$

under (C1) and (C2) and

$$P(\Phi_\alpha(\boldsymbol{\Omega})=1,\boldsymbol{G})\ge\alpha P(\boldsymbol{G})+(1-\alpha)P(\tilde{\boldsymbol{E}}^c,\boldsymbol{G})+o(1),$$

under (C1) and (C3). In the case when $\boldsymbol{A}=\boldsymbol{\Omega}$, by Lemma 4, we have

$$P\left(\max_{1\le i\le p}\frac{|\boldsymbol{\eta}_i|_2}{\sqrt{b_{ii}}}\ge(1-o(1))\max_{1\le i\le p}|\boldsymbol{\delta}_i|_2\right)\to 1.$$

Thus we have

$$\begin{aligned}P\left(\Phi_\alpha(\boldsymbol{A})=1,\boldsymbol{G}^c\right)&\ge P\left(8\sqrt{\sigma^2\log p}-\max_{1\le i\le p}\sum_{1\le j<l\le K}\frac{n_jn_l}{n_j+n_l}\frac{(\boldsymbol{A}(\bar{\boldsymbol{U}}_j-\bar{\boldsymbol{U}}_l))_i^2}{b_{ii}}\ge(1+\delta)\sqrt{2\sigma^2\log p},\boldsymbol{G}^c\right)-o(1)\\&=(1-o(1))P(\boldsymbol{G}^c)-o(1),\end{aligned}$$

where $\boldsymbol{U}_{j1}, \ldots, \boldsymbol{U}_{jn_j} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, $j = 1, \ldots, K$, for sufficiently small $\delta > 0$. We next consider $P(\Phi_\alpha(\boldsymbol{I}) = 1, \boldsymbol{G})$ and $P(\Phi_\alpha(\boldsymbol{A}) = 1, \boldsymbol{G})$. For notation briefness, we denote $P(\boldsymbol{LG}|\boldsymbol{\delta}_i)$ and $P(\boldsymbol{L}|\boldsymbol{\delta}_i)$ by $P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}(\boldsymbol{L})$ and $P_{\{\boldsymbol{\delta}_i\}}(\boldsymbol{L})$ respectively for any event $\boldsymbol{L}$ and $i = 1, \ldots, p$. Let $H^c = \{1, \ldots, p\} \setminus H$. We have

$$P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}(\Phi_\alpha(\boldsymbol{I}) = 1) = P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(\max_{i\in H} |\boldsymbol{Y}_i|_2 \geq x_p\right) + P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(\max_{i\in H} |\boldsymbol{Y}_i|_2 < x_p, \max_{j\in H^c} |\boldsymbol{Y}_j|_2 \geq x_p\right), \tag{34}$$

where $x_p = \sqrt{2\sigma^2 \log p + (d-2)\sigma^2 \log\log p + q_\alpha}$. Define

$$H_1^c = \{j \in H^c : |\sigma_{ij}| \leq p^{-\xi} \text{ for any } i \in H\}$$

for $2r < \xi < (1-r)/2$. It is easy to see that Card $(H_1) \leq Kp^{r+2\xi}$. It follows that

$$P\left(\max_{j\in H_1} |\boldsymbol{Y}_j|_2 \geq x_p\right) \leq Kp^{r+2\xi} P\left(|\boldsymbol{Y}_1|_2 \geq x_p\right) = O(p^{r+2\xi-1}) = o(1). \tag{35}$$

We claim that

$$P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(\max_{i\in H} |\boldsymbol{Y}_i|_2 < x_p, \max_{j\in H_1^c} |\boldsymbol{Y}_j|_2 \geq x_p\right) \leq (1+o(1))P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(\max_{i\in H} |\boldsymbol{Y}_i|_2 < x_p\right) P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(\max_{j\in H_1^c} |\boldsymbol{Y}_j|_2 \geq x_p\right) + o(1). \tag{36}$$

To prove (36), we set

$$\boldsymbol{E} = \left\{\max_{i\in H} |\boldsymbol{Y}_i|_2 < x_p\right\}, \qquad \boldsymbol{F}_j = \{|\boldsymbol{Y}_j|_2 \geq x_p\}, \quad j \in H_1^c.$$

Then by Bonferroni inequality, we have for any fixed integer $k > 0$,

$$P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(\bigcup_{j\in H_1^c} \{\boldsymbol{E} \cap \boldsymbol{F}_j\}\right) \leq \sum_{t=1}^{2k-1} (-1)^{t-1} \sum_{i_1<\cdots<i_t\in H_1^c} P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(\boldsymbol{E} \cap \boldsymbol{F}_{i_1} \cap \cdots \cap \boldsymbol{F}_{i_t}\right). \tag{37}$$

Let $\boldsymbol{Y}^* = (\boldsymbol{Y}_i^T, i \in H)^T$, $\boldsymbol{Y}^\star = (\boldsymbol{Y}_{i_1}^T, \ldots, \boldsymbol{Y}_{i_t}^T)^T$, and let $|\boldsymbol{Y}^*|_m = \max_{i\in H} |\boldsymbol{Y}_i|_2$ and $|\boldsymbol{Y}^\star|_t = \min_{1\leq j\leq t} |\boldsymbol{Y}_{i_j}|_2$. Let $(\boldsymbol{Z}_1^{*T}, \ldots, \boldsymbol{Z}_m^{*T})^T =: \boldsymbol{Z}^* \sim N(\boldsymbol{0}, \boldsymbol{I}_{bm\times bm})$, independent with $\boldsymbol{Y}^*$, and $(\boldsymbol{Z}_{i_1}^{\star T}, \ldots, \boldsymbol{Z}_{i_t}^{\star T})^T =: \boldsymbol{Z}^\star \sim N(\boldsymbol{0}, \boldsymbol{I}_{bt\times bt})$, independent with $\boldsymbol{Y}^\star$. Similarly as proved in Theorem 1, let $\lambda_p = Cp^{-\xi}$ for some constant $C > 0$, we have

$$\begin{aligned} P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}(\boldsymbol{E}) &= P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}(|\boldsymbol{Y}^*|_m < x_p) \\ &\leq P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}^* + \lambda_p \boldsymbol{Z}^*|_2 < x_p + \lambda_p \max_{i\in H} |\boldsymbol{Z}_i^*|_2\right) \\ &\leq P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}(|\boldsymbol{Y}^* + \lambda_p \boldsymbol{Z}^*|_2 < x_p + Cp^{-\xi/8}) + O(p^{-M}), \end{aligned} \tag{38}$$

for sufficiently large constant $M > 0$. We also have

$$\begin{aligned} P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(\bigcap_{1\leq j\leq t} \boldsymbol{F}_{i_j}\right) &= P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}(|\boldsymbol{Y}^\star|_t \geq x_p) \\ &\leq P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}^\star + \lambda_p \boldsymbol{Z}^\star|_t \geq x_p - \lambda_p \max_{1\leq j\leq t} |\boldsymbol{Z}_{i_j}^\star|_2\right) \\ &\leq P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}(|\boldsymbol{Y}^\star + \lambda_p \boldsymbol{Z}^\star|_t \geq x_p - Cp^{-\xi/8}) + O(p^{-2t}). \end{aligned} \tag{39}$$

Thus, we have

$$P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(\boldsymbol{E} \cap \boldsymbol{F}_{i_1} \cap \cdots \cap \boldsymbol{F}_{i_t}\right) \leq P_{\{\boldsymbol{\delta}_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}^* + \lambda_p \boldsymbol{Z}^*|_2 < x_p + Cp^{-\xi/8}, |\boldsymbol{Y}^\star + \lambda_p \boldsymbol{Z}^\star|_t \geq x_p - Cp^{-\xi/8}\right) + O(p^{-2t}). \tag{40}$$

Let $\boldsymbol{W} = (w_{ij})$ be the covariance matrix of the vector $((\boldsymbol{Y}^* + \lambda_p \boldsymbol{Z}^*)^T, (\boldsymbol{Y}^\star + \lambda_p \boldsymbol{Z}^\star)^T)^T$. Let $(\tilde{w}_{ij}) =: \widetilde{\boldsymbol{W}} = \text{diag}(\boldsymbol{W}_1, \boldsymbol{W}_2)$, where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are the covariance matrices of $\boldsymbol{Y}^* + \lambda_p \boldsymbol{Z}^*$ and $\boldsymbol{Y}^\star + \lambda_p \boldsymbol{Z}^\star$ respectively. So for $(i_1, \ldots, i_t) \in H_1^c$, we have $\|\boldsymbol{W} - \widetilde{\boldsymbol{W}}\|_2 = O(p^{r-\xi})$. Set $\boldsymbol{z} = (\boldsymbol{\delta}_i^T, i \in H, \boldsymbol{z}_{i_1}^T, \ldots, \boldsymbol{z}_{i_t}^T)^T$ and

$$\mathcal{R} = \{|\boldsymbol{u}_i + \boldsymbol{\delta}_i|_2 \leq x_p + Cp^{-\xi/8}, i \in H, |\boldsymbol{z}_{i_1}|_2 \geq x_p \ldots, |\boldsymbol{z}_{i_t}|_2 \geq x_p - Cp^{-\xi/8}\},$$

$$\mathcal{R}_1 = \mathcal{R} \cap \left\{\max_{1\leq j\leq t} |\boldsymbol{z}_{i_j}|_2 \leq 8b\sqrt{t\log p}\right\},$$

$$\mathcal{R}_2 = \mathcal{R} \cap \left\{\max_{1\leq j\leq t} |\boldsymbol{z}_{i_j}|_2 > 8b\sqrt{t\log p}\right\}.$$

We have

$$P_{\{\delta_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}^*+\lambda_p\boldsymbol{Z}^*|_2<x_p+Cp^{-\xi/8},|\boldsymbol{Y}^\star+\lambda_p\boldsymbol{Z}^\star|_t\geq x_p-Cp^{-\xi/8}\right)$$
$$=\frac{I\{\boldsymbol{G}\}}{(2\pi)^{(bm+bt)/2}|\boldsymbol{W}|^{1/2}}\int_{\mathcal{R}}\exp\left(-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{W}^{-1}\boldsymbol{z}\right)d\boldsymbol{z}. \tag{41}$$

Note that $|\boldsymbol{W}|=(1+O(p^{r-\xi}))^{bm+bt}|\widetilde{\boldsymbol{W}}|=(1+O(p^{2r-\xi}))|\widetilde{\boldsymbol{W}}|$ and

$$\|\boldsymbol{W}^{-1}-\widetilde{\boldsymbol{W}}^{-1}\|_2\leq C\lambda_p^{-2}p^{r-\xi}=O(p^{r-\xi/2}).$$

This implies that

$$\frac{1}{(2\pi)^{(bm+bt)/2}|\boldsymbol{W}|^{1/2}}\int_{\mathcal{R}_1}\exp\left(-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{W}^{-1}\boldsymbol{z}\right)d\boldsymbol{z}$$
$$=(1+O(p^{2r-\xi}\log p))\frac{1}{(2\pi)^{(bm+bt)/2}|\widetilde{\boldsymbol{W}}|^{1/2}}\int_{\mathcal{R}_1}\exp\left(-\frac{1}{2}\boldsymbol{z}^T\widetilde{\boldsymbol{W}}^{-1}\boldsymbol{z}\right)d\boldsymbol{z}. \tag{42}$$

Furthermore, it is easy to see that

$$\frac{1}{(2\pi)^{(bm+bt)/2}|\boldsymbol{W}|^{1/2}}\int_{\mathcal{R}_2}\exp\left(-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{W}^{-1}\boldsymbol{z}\right)d\boldsymbol{z}=O(p^{-16bt}),$$
$$\frac{1}{(2\pi)^{(bm+bt)/2}|\widetilde{\boldsymbol{W}}|^{1/2}}\int_{\mathcal{R}_2}\exp\left(-\frac{1}{2}\boldsymbol{z}^T\tilde{\boldsymbol{W}}^{-1}\boldsymbol{z}\right)d\boldsymbol{z}=O(p^{-16bt}). \tag{43}$$

Thus, it follows from (41) to (43) that

$$P_{\{\delta_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}^*+\lambda_p\boldsymbol{Z}^*|_2<x_p+Cp^{-\xi/8},|\boldsymbol{Y}^\star+\lambda_p\boldsymbol{Z}^\star|_t\geq x_p-Cp^{-\xi/8}\right)$$
$$=(1+O(p^{2r-\xi}\log p))P_{\{\delta_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}^*+\lambda_p\boldsymbol{Z}^*|_2<x_p+Cp^{-\xi/8}\right)P\left(|\boldsymbol{Y}^\star+\lambda_p\boldsymbol{Z}^\star|_t\geq x_p-Cp^{-\xi/8}\right)+O(p^{-16bt})$$
$$=(1+o(1))P_{\{\delta_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}^*|_2<x_p\right)P_{\{\delta_i\}}\left(|\boldsymbol{Y}^\star|_t\geq x_p\right)+O(p^{-16bt}).$$

So

$$P_{\{\delta_i\},\boldsymbol{G}}\left(\boldsymbol{E}\cap\boldsymbol{F}_{i_1}\cap\cdots\cap\boldsymbol{F}_{i_t}\right)\leq(1+o(1))P_{\{\delta_i\},\boldsymbol{G}}(\boldsymbol{E})P_{\{\delta_i\}}(\boldsymbol{F}_{i_1}\cap\cdots\cap\boldsymbol{F}_{i_t})+O(p^{-2t}).$$

Similarly as (40), we have

$$P_{\{\delta_i\},\boldsymbol{G}}\left(\boldsymbol{E}\cap\boldsymbol{F}_{i_1}\cap\cdots\cap\boldsymbol{F}_{i_t}\right)\geq P_{\{\delta_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}^*+\lambda_p\boldsymbol{Z}^*|_2<x_p-Cp^{-\xi/8},|\boldsymbol{Y}^\star+\lambda_p\boldsymbol{Z}^\star|_t\geq x_p+Cp^{-\xi/8}\right)+O(p^{-2t}). \tag{44}$$

Thus, by using the exact argument as above, we have

$$P_{\{\delta_i\},\boldsymbol{G}}\left(\boldsymbol{E}\cap\boldsymbol{F}_{i_1}\cap\cdots\cap\boldsymbol{F}_{i_t}\right)\geq(1+o(1))P_{\{\delta_i\},\boldsymbol{G}}(\boldsymbol{E})P_{\{\delta_i\}}(\boldsymbol{F}_{i_1}\cap\cdots\cap\boldsymbol{F}_{i_t})+O(p^{-2t}).$$

So we have

$$P_{\{\delta_i\},\boldsymbol{G}}\left(\boldsymbol{E}\cap\boldsymbol{F}_{i_1}\cap\cdots\cap\boldsymbol{F}_{i_t}\right)=(1+o(1))P_{\{\delta_i\},\boldsymbol{G}}(\boldsymbol{E})P_{\{\delta_i\}}(\boldsymbol{F}_{i_1}\cap\cdots\cap\boldsymbol{F}_{i_t})+O(p^{-2t}).$$

As the proof of Lemma 5, we can show that

$$\sum_{i_1<\cdots<i_t\in H_1^c}P_{\{\delta_i\}}\left(\boldsymbol{F}_{i_1}\cap\cdots\cap\boldsymbol{F}_{i_t}\right)=(1+o(1))\Gamma^{-t}\left(\frac{K-1}{2}\right)\frac{1}{t!}\exp\left(-\frac{tq_\alpha}{K}\right).$$

It follows from (37) that

$$P_{\{\delta_i\},\boldsymbol{G}}\left(\bigcup_{j\in H_1^c}\{\boldsymbol{E}\cap\boldsymbol{F}_j\}\right)\leq\alpha P_{\{\delta_i\},\boldsymbol{G}}(\boldsymbol{E})+o(1).$$

This, together with (34) and (35), implies that

$$P_{\{\delta_i\},\boldsymbol{G}}(\Phi_\alpha(\boldsymbol{I})=1)\leq\alpha I\{\boldsymbol{G}\}+(1-\alpha)P_{\{\delta_i\},\boldsymbol{G}}(\boldsymbol{E}^c)+o(1),$$

where $o(1)$ is uniformly for $\{\boldsymbol{\delta}^{(jl)},1\leq j<l\leq K\}$. Hence, we have

$$P(\Phi_\alpha(\boldsymbol{I})=1,\boldsymbol{G})\leq\alpha P(\boldsymbol{G})+(1-\alpha)P(\boldsymbol{E}^c,\boldsymbol{G})+o(1).$$

We next prove that

$$P(\Phi_\alpha(\boldsymbol{A}) = 1, \boldsymbol{G}) \geq \alpha P(\boldsymbol{G}) + (1-\alpha) P(\tilde{\boldsymbol{E}}^c, \boldsymbol{G}) + o(1), \tag{45}$$

where

$$\tilde{\boldsymbol{E}} = \left\{ \max_{i \in H} |\boldsymbol{Z}_i|_2 < x_p \right\}.$$

Define

$$\tilde{H}_1^c = \{j \in H^c : |a_{ij}| \leq p^{-\xi} \text{ and } |b_{ij}| \leq p^{-\xi} \text{ for any } i \in H\}$$

for $2r < \xi < (1-r)/2$. We can see that $\mathrm{Card}(\tilde{H}_1^c) \geq p - O(p^{r+2\xi})$. Then

$$\begin{aligned} P_{\{\delta_i\},\boldsymbol{G}}(\Phi_\alpha(\boldsymbol{A}) = 1) &= P_{\{\delta_i\},\boldsymbol{G}}\left(\max_{i \in H} |\boldsymbol{Z}_i|_2 \geq x_p\right) + P_{\{\delta_i\},\boldsymbol{G}}\left(\max_{i \in H} |\boldsymbol{Z}_i|_2 < x_p, \max_{k \in H^c} |\boldsymbol{Z}_k|_2 \geq x_p\right) \\ &\geq P_{\{\delta_i\},\boldsymbol{G}}\left(\max_{i \in H} |\boldsymbol{Z}_i|_2 \geq x_p\right) + P_{\{\delta_i\},\boldsymbol{G}}\left(\max_{i \in H} |\boldsymbol{Z}_i|_2 < x_p, \max_{k \in \tilde{H}_1^c} |\boldsymbol{Z}_k|_2 \geq x_p\right). \end{aligned}$$

Note that on $\boldsymbol{G}$,

$$\max_{k \in \tilde{H}_1^c} |\boldsymbol{\eta}_k|_2 = \max_{k \in \tilde{H}_1^c} \frac{1}{\sqrt{2}} \left(\sum_{1 \leq j < l \leq K} \left(\sum_{i \in H} a_{ki} \delta_i^{(jl)} \right)^2 \right)^{1/2} = O(p^{r-\xi}\sqrt{\log p}).$$

Following the exact arguments as above and using the left side Bonferroni inequality, we can show that

$$\begin{aligned} P_{\{\delta_i\},\boldsymbol{G}}\left(\max_{i \in H} |\boldsymbol{Z}_i|_2 < x_p, \max_{k \in \tilde{H}_1^c} |\boldsymbol{Z}_k|_2 \geq x_p\right) &\geq P_{\{\delta_i\},\boldsymbol{G}}\left(\max_{i \in H} |\boldsymbol{Z}_i|_2 < x_p, \max_{k \in \tilde{H}_1^c} |\boldsymbol{Z}_k - \boldsymbol{\eta}_k/\sqrt{b_{kk}}|_2 \geq x_p + O(p^{r-\xi}\sqrt{\log p})\right) \\ &\geq \alpha P_{\{\delta_i\},\boldsymbol{G}}(\tilde{\boldsymbol{E}}) - o(1). \end{aligned}$$

Hence (45) is proved. Now we prove

$$P_{\{\delta_i\},\boldsymbol{G}}\left(\max_{i \in H} |\boldsymbol{Y}_i|_2 \geq x_p\right) = I\{\boldsymbol{G}\} - I\{\boldsymbol{G}\} \prod_{i \in H} \left(1 - P_{\{\delta_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}_i|_2 \geq x_p\right)\right) + o(1). \tag{46}$$

Let $\mathcal{I}_0 = \{(i_1, \ldots, i_m) : \exists 1 \leq k < j \leq m, \text{ such that } |\sigma_{i_k i_l}| \geq p^{-\xi}\}$ for $2r < \xi < \frac{1}{2}(1-2r)$ and let $\mathcal{I} = \{(i_1, \ldots, i_m)\}$. We can show that

$$|\mathcal{I}_0|/|\mathcal{I}| \leq O\left(p \cdot p^{2\xi} \binom{p}{k_p - 2} \Big/ \binom{p}{k_p}\right).$$

So for $\xi < \frac{1}{2}(1-2r)$, $|\mathcal{I}_0|/|\mathcal{I}| = o(1)$. Let $\boldsymbol{Y}^* = (\boldsymbol{Y}_i^T, i \in H)^T$ and $|\boldsymbol{Y}^*|_m = \max_{i \in H} |\boldsymbol{Y}_i|_2$. Let $(\boldsymbol{Z}_{i_1}^T, \ldots, \boldsymbol{Z}_{i_m}^T)^T =: \boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I}_{bm \times bm})$, where $b = \frac{K(K-1)}{2}$ and $m = \mathrm{Card}(H)$. Let $\lambda = Cp^{-2\delta}$ for $\delta < \frac{1}{4}(\xi - 2r)$. So for $H \in \mathcal{I}_0^c$, we have

$$\begin{aligned} P_{\{\delta_i\},\boldsymbol{G}}(\max_{i \in H} |\boldsymbol{Y}_i|_2 \leq x_p) &\leq P_{\{\delta_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}^* + \lambda_p \boldsymbol{Z}|_m \leq x_p + \lambda_p \max_{1 \leq j \leq m} |\boldsymbol{Z}_{i_j}|_2\right) \\ &\leq \frac{I\{\boldsymbol{G}\}}{(2\pi)^{bm/2} \det(\boldsymbol{\Sigma}_1 + \lambda_p \boldsymbol{I})^{1/2}} \int_{|\boldsymbol{z}|_m \leq x_p + Cp^{-\delta}} \exp\left(-\frac{1}{2} \boldsymbol{z}^T (\boldsymbol{\Sigma}_1 + \lambda_p \boldsymbol{I})^{-1} \boldsymbol{z}\right) d\boldsymbol{z} + O(p^{-M}), \end{aligned}$$

where $\boldsymbol{z} \in \mathbb{R}^{bt}$ and $\boldsymbol{\Sigma}_1$ is the covariance matrix of $\boldsymbol{Y}^*$, C is a constant and M is a sufficiently large constant. Let $\widetilde{\boldsymbol{\Sigma}}$ be a $bm \times bm$ matrix with $\widetilde{\boldsymbol{\Sigma}}_{jb+1:(j+1)b, jb+1:(j+1)b} = \boldsymbol{\Sigma}_0$ for $j = 0, \ldots, m-1$ and $\widetilde{\boldsymbol{\Sigma}}_{ij} = 0$ otherwise. For $(i_1, \ldots, i_t) \in \mathcal{I}_0^c$, we have $\boldsymbol{\Sigma}_{1jb+1:(j+1)b, jb+1:(j+1)b} = \boldsymbol{\Sigma}_0$ for $j = 0, \ldots, m-1$ and $|\boldsymbol{\Sigma}_{1ij}| < p^{-\xi}$ otherwise. Write

$$\begin{aligned} \int_{|\boldsymbol{z}|_m \leq x_p + Cp^{-\delta}} \exp\left(-\frac{1}{2} \boldsymbol{z}^T (\boldsymbol{\Sigma}_1 + \lambda_p \boldsymbol{I})^{-1} \boldsymbol{z}\right) d\boldsymbol{z} &= \int_{|\boldsymbol{z}|_m \leq x_p + Cp^{-\delta}, \|\boldsymbol{z}\|^2 \geq m(\log p)^2} \exp\left(-\frac{1}{2} \boldsymbol{z}^T (\boldsymbol{\Sigma}_1 + \lambda_p \boldsymbol{I})^{-1} \boldsymbol{z}\right) d\boldsymbol{z} \\ &\quad + \int_{|\boldsymbol{z}|_m \leq x_p + Cp^{-\delta}, \|\boldsymbol{z}\|^2 \leq m(\log p)^2} \exp\left(-\frac{1}{2} \boldsymbol{z}^T (\boldsymbol{\Sigma}_1 + \lambda_p \boldsymbol{I})^{-1} \boldsymbol{z}\right) d\boldsymbol{z}. \end{aligned}$$

Because $\lambda_{\max}(\boldsymbol{\Sigma}_1+\lambda_p\boldsymbol{I})\le\lambda_{\max}(\boldsymbol{\Sigma}_0)+O(p^{-2\delta})\le M$ for some constant $M>0$, we can get

$$\int_{|\boldsymbol{z}|_m\le x_p+Cp^{-\delta},\|\boldsymbol{z}\|^2\ge m(\log p)^2}\exp\left(-\frac{1}{2}\boldsymbol{z}^T(\boldsymbol{\Sigma}_1+\lambda_p\boldsymbol{I})^{-1}\boldsymbol{z}\right)d\boldsymbol{z}\le C\exp(-(\log p)^2/2b)\le Cp^{-2b},$$

uniformly in $(i_1,\ldots,i_t)\in\mathcal{I}_0^c$. For the second part of the sum in (22), note that

$$\|(\boldsymbol{\Sigma}_1+\lambda_p\boldsymbol{I})^{-1}-(\widetilde{\boldsymbol{\Sigma}}+\lambda_p\boldsymbol{I})^{-1}\|_2\le C\lambda_p^{-2}p^{r-\xi}\le Cp^{r-\xi+4\delta},$$

we can obtain that

$$\begin{aligned}
&\int_{|\boldsymbol{z}|_m\le x_p+Cp^{-\delta},\|\boldsymbol{z}\|^2\le m(\log p)^2}\exp\left(-\frac{1}{2}\boldsymbol{z}^T(\boldsymbol{\Sigma}_1+\lambda_p\boldsymbol{I})^{-1}\boldsymbol{z}\right)d\boldsymbol{z}\\
&=\int_{|\boldsymbol{z}|_m\le x_p+Cp^{-\delta},\|\boldsymbol{z}\|^2\le m(\log p)^2}\exp\left(-\frac{1}{2}\boldsymbol{z}^T((\boldsymbol{\Sigma}_1+\lambda_p\boldsymbol{I})^{-1}-(\widetilde{\boldsymbol{\Sigma}}+\lambda_p\boldsymbol{I})^{-1})\boldsymbol{z}-\frac{1}{2}\boldsymbol{z}^T(\widetilde{\boldsymbol{\Sigma}}+\lambda_p\boldsymbol{I})^{-1}\boldsymbol{z}\right)d\boldsymbol{z}\\
&=(1+O(p^{2r-\xi+4\delta}(\log p)^2))\int_{|\boldsymbol{z}|_m\le x_p+Cp^{-\delta},\|\boldsymbol{z}\|^2\le m(\log p)^2}\exp\left(-\frac{1}{2}\boldsymbol{z}^T(\widetilde{\boldsymbol{\Sigma}}+\lambda_p\boldsymbol{I})^{-1}\boldsymbol{z}\right)d\boldsymbol{z}\\
&=(1+O(p^{2r-\xi+4\delta}(\log p)^2))\int_{|\boldsymbol{z}|_m\le x_p+Cp^{-\delta}}\exp\left(-\frac{1}{2}\boldsymbol{z}^T(\widetilde{\boldsymbol{\Sigma}}+\lambda_p\boldsymbol{I})^{-1}\boldsymbol{z}\right)d\boldsymbol{z}+O(p^{-2b})\\
&=(1+O(p^{2r-\xi+4\delta}(\log p)^2))\prod_{i\in H}\left(\int_{|\boldsymbol{z}_i|_2\le x_p+Cp^{-\delta}}\exp\left(-\frac{1}{2}\boldsymbol{z}_i^T(\boldsymbol{\Sigma}_0+\lambda_p\boldsymbol{I})^{-1}\boldsymbol{z}_i\right)d\boldsymbol{z}_i\right)+O(p^{-2b}),
\end{aligned}$$

where $\boldsymbol{z}_{i_1}\in\mathbb{R}^b$. Because $\|(\boldsymbol{\Sigma}_1+\lambda_p\boldsymbol{I})-(\widetilde{\boldsymbol{\Sigma}}+\lambda_p\boldsymbol{I})\|_2=O(p^{r-\xi})$, we have $\det(\boldsymbol{\Sigma}_1+\lambda_p\boldsymbol{I})=(1+O(p^{r-\xi}))^{bm}\det(\widetilde{\boldsymbol{\Sigma}}+\lambda_p\boldsymbol{I})=(1+O(p^{2r-\xi}))\det(\widetilde{\boldsymbol{\Sigma}}+\lambda_p\boldsymbol{I})$. So we have

$$\begin{aligned}
P_{\{\delta_i\},\boldsymbol{G}}(\max_{i\in H}|\boldsymbol{Y}_i|_2\le x_p)&\le(1+o(1))I\{\boldsymbol{G}\}\prod_{i\in H}P_{\{\delta_i\},\boldsymbol{G}}(|\boldsymbol{Y}_i+\lambda_p\boldsymbol{Z}_i|_2\le x_p)+o(1)\\
&\le(1+o(1))I\{\boldsymbol{G}\}\prod_{i\in H}P_{\{\delta_i\},\boldsymbol{G}}(|\boldsymbol{Y}_i|_2\le x_p+\lambda_p\max_{i\in H}|\boldsymbol{Z}_i|_2)+o(1)\\
&=(1+o(1))I\{\boldsymbol{G}\}\prod_{i\in H}P_{\{\delta_i\},\boldsymbol{G}}(|\boldsymbol{Y}_i|_2\le x_p)+o(1).
\end{aligned}$$

Similarly, because

$$P_{\{\delta_i\},\boldsymbol{G}}\left(\max_{i\in H}|\boldsymbol{Y}_i|_2\le x_p\right)\ge P_{\{\delta_i\},\boldsymbol{G}}\left(|\boldsymbol{Y}^*+\lambda_p\boldsymbol{Z}|_m\le x_p-\lambda\max_{1\le j\le m}|\boldsymbol{Z}_{i_j}|_2\right),$$

we can get

$$P_{\{\delta_i\},\boldsymbol{G}}\left(\max_{i\in H}|\boldsymbol{Y}_i|_2\le x_p\right)\ge(1-o(1))I\{\boldsymbol{G}\}\prod_{i\in H}P_{\{\delta_i\},\boldsymbol{G}}(|\boldsymbol{Y}_i|_2\le x_p)-o(1).$$

So (46) is proved. Similarly, let $\mathcal{I}_1=\{(i_1,\ldots,i_m):\ \exists 1\le k<j\le m,\ \text{s.t.}\ |b_{i_ki_l}|\ge p^{-\xi}\}$, then we can get $|\mathcal{I}_1|/|\mathcal{I}|=o(1)$, and for $H\in\mathcal{I}_1^c$,

$$P_{\{\delta_i\},\boldsymbol{G}}\left(\max_{i\in H}|\boldsymbol{Z}_i|_2\ge x_p\right)=I\{\boldsymbol{G}\}-I\{\boldsymbol{G}\}\prod_{i\in H}\left(1-P_{\{\delta_i\},\boldsymbol{G}}\left(|\boldsymbol{Z}_i|_2\ge x_p\right)\right)+o(1).\quad\blacksquare$$

Proof of Lemma 7. Based on the proof of Lemma 4 in supplementary material [7] (available on the web at www.unc.edu/~xiayin/mean-suppmaterial.pdf), it is enough to show that, for $i=1,\ldots,p$, we have

$$P(|\boldsymbol{Y}_i|_2\ge x_p+a_p)=(1+o(1))P(|\boldsymbol{Y}_i|_2\ge x_p)+o(p^{-r}).\tag{47}$$

Without loss of generality, suppose $a_n>0$ and $\delta_i^{(jl)}\ge 0$ for $1\le j<l\le K$ and $i=1,\ldots,p$. Because $\boldsymbol{Y}_i\sim N(\boldsymbol{\delta}_i,\boldsymbol{\Sigma}_0)$, let $\boldsymbol{Z}\sim N(\boldsymbol{0},\boldsymbol{I})$, then similarly as the proof from (21) and (27) in Theorem 1 for $t=1$ and $\widetilde{\boldsymbol{\Sigma}}=\boldsymbol{\Sigma}_1=\boldsymbol{\Sigma}_0$, we have

$$\begin{aligned}
P(|\boldsymbol{Y}_i|_2\ge x_p+a_p)&=P(|\boldsymbol{Y}_i+Cp^{-\gamma/4}\boldsymbol{Z}|_2\ge x_p+a_p+O(p^{-\gamma/8}))+O(p^{-2})\\
&=P((\lambda_1z_1+\delta_{i1})^2+(\lambda_2z_2+\delta_{i2})^2+\cdots+(\lambda_bz_b+\delta_{ib})^2\ge(x_p+O(p^{-\gamma/8})+a_p)^2)+O(p^{-2}),
\end{aligned}$$

where $(\delta_{i1}, \ldots, \delta_{ib})^T := U(\delta_i^{(12)}, \ldots, \delta_i^{(K-1K)})^T$, $\mathbf{\Sigma}_0 = U^T \Lambda U$, $(z_1, \ldots, z_b) =: \boldsymbol{z} \sim N(\mathbf{0}, \boldsymbol{I})$ and $\lambda_1^2 \geq \cdots \geq \lambda_b^2 > 0$ are eigenvalues of $\mathbf{\Sigma}_0 + Cn^{-\gamma/4}\boldsymbol{I}$ for some constant C satisfying $C > 0$. Hence, there exist a δ such that

$$\begin{aligned} P(|\boldsymbol{Y}_i|_2 \geq x_p + a_p) &= P((\lambda_1 z_1 + \delta)^2 \geq (x_p + O(p^{-\gamma/8}) + a_p)^2 - (\lambda_2^2 z_2^2 + \cdots + \lambda_b^2 z_b^2), \\ &\qquad \lambda_2^2 z_2^2 + \cdots + \lambda_b^2 z_b^2 \leq r(x_p + O(p^{-\gamma/8}) + a_p)^2) + o(p^{-r}) \\ &= (1 + o(1))P((\lambda_1 z_1 + \delta)^2 \geq (x_p + O(p^{-\gamma/8}))^2 - (\lambda_2^2 z_2^2 + \cdots + \lambda_b^2 z_b^2)) + o(p^{-r}), \end{aligned}$$

where the last equality comes from the proof of Eq. 12 in Lemma 4 in supplementary material [7]. It follows from the fact that

$$P(|\boldsymbol{Y}_i|_2 \geq x) = P(|\boldsymbol{Y}_i + Cp^{-\gamma/4}\boldsymbol{Z}|_2 \geq x_p + O(p^{-\gamma/8})) + O(p^{-2})$$

(47) is proved. ∎

Proof of Theorem 4. We have $P\left(|\hat{\mathbf{\Sigma}}_{\boldsymbol{X}} - \mathbf{\Sigma}|_\infty \leq C\sqrt{\log p/n}\right) \to 1$ as $n, p \to \infty$; see [5]. On the event $\{|\hat{\mathbf{\Sigma}}_{\boldsymbol{X}} - \mathbf{\Sigma}|_\infty \leq C\sqrt{\log p/n}\}$ with $\widehat{\boldsymbol{A}} = \widehat{\mathbf{\Omega}}$,

$$|\widehat{\boldsymbol{A}}\widehat{\mathbf{\Sigma}}\widehat{\boldsymbol{A}} - \boldsymbol{A}\mathbf{\Sigma}\boldsymbol{A}|_\infty \leq Cs_p M_p^{2-q}\left(\frac{\log p}{n}\right)^{(1-q)/2} = o(1/\log p).$$

Hence, as in the proof of Lemma 6, it is easy to show that $P\left(\Phi_\alpha(\widehat{\boldsymbol{A}}) = 1, \boldsymbol{G}^c\right) = P(\boldsymbol{G}^c) + o(1)$ and $P\left(\Phi_\alpha(\boldsymbol{A}) = 1, \boldsymbol{G}^c\right) = P(\boldsymbol{G}^c) + o(1)$. Note that, for $1 \leq j < l \leq K$, $\sqrt{\frac{n_j n_l}{n_j+n_l}}\widehat{\boldsymbol{A}}(\bar{\boldsymbol{X}}_j - \bar{\boldsymbol{X}}_l) = (\widehat{\boldsymbol{A}} - \boldsymbol{A})(\sqrt{\frac{n_j n_l}{n_j+n_l}}(\bar{\boldsymbol{X}}_j - \bar{\boldsymbol{X}}_l) - \boldsymbol{\delta}^{(jl)}) + (\widehat{\boldsymbol{A}} - \boldsymbol{A})\boldsymbol{\delta}^{(jl)} + \sqrt{\frac{n_j n_l}{n_j+n_l}}\boldsymbol{A}(\bar{\boldsymbol{X}}_j - \bar{\boldsymbol{X}}_l)$. On $\boldsymbol{G}$, we have

$$\left|(\widehat{\boldsymbol{A}} - \boldsymbol{A})\left(\sqrt{\frac{n_j n_l}{n_j + n_l}}(\bar{\boldsymbol{X}}_j - \bar{\boldsymbol{X}}_l) - \boldsymbol{\delta}^{(jl)}\right) + (\widehat{\boldsymbol{A}} - \boldsymbol{A})\boldsymbol{\delta}^{(jl)}\right|_\infty = O_P\left(\frac{\log p}{\sqrt{\min\{n_j, n_l\}}}\right) = o_P\left(\frac{1}{\sqrt{\log p}}\right).$$

To prove Theorem 4, it suffices to show that

$$P\left(\max_{1\leq i\leq p} |\boldsymbol{Z}_i|_2 \geq x_p + a_n, \boldsymbol{G}\right) = P\left(\max_{1\leq i\leq p} |\boldsymbol{Z}_i|_2 \geq x_p, \boldsymbol{G}\right) + o(1), \tag{48}$$

for any $a_n = o((\log p)^{-1/2})$, where $\boldsymbol{Z}_i = \frac{1}{\sqrt{b_{i,i}}}\left(\sqrt{\frac{n_1 n_2}{n_1+n_2}}(\boldsymbol{A}(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2))_i, \ldots, \sqrt{\frac{n_{K-1}n_K}{n_{K-1}+n_K}}(\boldsymbol{A}(\bar{\boldsymbol{X}}_{K-1} - \bar{\boldsymbol{X}}_K))_i\right)$. From the proof of Lemma 6, Let $H = \{1 \leq i \leq p : \delta_i^{(jl)} \neq 0 \text{ for some } 1 \leq j < l \leq K\} = \{l_1, \ldots, l_m\}$, then we can get

$$P\left(\max_{1\leq i\leq p} |\boldsymbol{Z}_i|_2 \geq x_p + a_n, \boldsymbol{G}\right) = \alpha P(\boldsymbol{G}) + (1 - \alpha)P\left(\max_{i\in H} |\boldsymbol{Y}_i|_2 \geq x_p + a_n, \boldsymbol{G}\right) + o(1),$$

$$P\left(\max_{1\leq i\leq p} |\boldsymbol{Z}_i|_2 \geq x_p, \boldsymbol{G}\right) = \alpha P(\boldsymbol{G}) + (1 - \alpha)P\left(\max_{i\in H} |\boldsymbol{Y}_i|_2 \geq x_p, \boldsymbol{G}\right) + o(1),$$

where given $\boldsymbol{\delta}$, $\boldsymbol{Y}_i$, $i \in H$ are independent normal random vectors with covariance matrix $\mathbf{\Sigma}_0$. Thus, (48) can be proved by Lemma 7 and Theorem 4 is proved. ∎

Acknowledgments

The research of Tony Cai was supported in part by NSF Grant DMS-1208982 and NIH Grant R01 CA127334.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.jmva.2014.07.002.

References

[1] T.W. Anderson, An introduction to multivariate statistical analysis, third ed., Wiley-Interscience, 2003.
[2] Z. Bai, H. Saranadasa, Effect of high dimension: by an example of a two sample problem, Statist. Sinica 6 (1996) 311–329.
[3] S.M. Berman, A law of large numbers for the maximum of a stationary Gaussian sequence, Ann. Math. Statist. 33 (1962) 93–97.
[4] A. Birnbaum, B. Nadler, 2012. High Dimensional Sparse Covariance Estimation: Accurate Thresholds for the Maximal Diagonal Entry and for the Largest Correlation Coefficient. Technical Report.
[5] T. Cai, W.D. Liu, Adaptive thresholding for sparse covariance matrix estimation, J. Amer. Statist. Assoc. 106 (2011) 672–684.
[6] T. Cai, W.D. Liu, X. Luo, A constrained l_1 minimization approach to sparse precision matrix estimation, J. Amer. Statist. Assoc. 106 (2011) 594–607.
[7] T. Cai, W.D. Liu, Y. Xia, Two-sample test of high dimensional means under dependency, J. Roy. Statist. Soc. Ser. B 76 (2014) 349–372.
[8] T. Cai, H. Zhou, Minimax estimation of large covariance matrices under l1 norm (with discussion), Statist. Sinica 22 (2012) 1319–1378.

[9] S. Chen, Y. Qin, A two-sample test for high-dimensional data with applications to gene-set testing, Ann. Statist. 38 (2010) 808–835.
[10] Y. Fan, J. Jin, Z. Yao, Optimal classification in sparse Gaussian Graphical Model, Ann. Statist. 41 (2013) 2537–2571.
[11] Y. Fujikoshi, T. Himeno, H. Wakaki, Asymptotic results of a high dimensional MANOVA test and power comparisons when the dimension is large compared to the sample size, J. Japan Statist. Soc. 34 (2004) 19–26.
[12] P. Hall, On convergence rates of suprema, Probab. Theory Related Fields 89 (1991) 447–455.
[13] P. Hall, J. Jin, Innovated higher criticism for detecting sparse signals in correlated noise, Ann. Statist. 38 (2010) 1686–1732.
[14] S. Huckemann, T. Hotz, A. Munk, Intrinsic MANOVA for Riemannian manifolds with an application to Kendall's space of planar shapes, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 593–603.
[15] W. Liu, Z.Y. Lin, Q.M. Shao, The asymptotic distribution and Berry–Esseen bound of a new test for independence in high dimension with an application to stochastic optimization, Ann. Appl. Probab. 18 (2008) 2337–2366.
[16] J.R. Schott, Some high-dimensional tests for a one-way MANOVA, J. Multivariate Anal. 98 (2007) 1825–1839.
[17] Y.F. Shen, Z.Y. Lin, J. Zhu, Shrinkage-based regularization tests for high-dimensional data with application to gene set analysis, Comput. Statist. Data Anal. 55 (2011) 2221–2233.
[18] M. Srivastava, A test for the mean vector with fewer observations than the dimension under non-normality, J. Multivariate Anal. 100 (2009) 518–532.
[19] M. Srivastava, Multivariate theory for analyzing high dimensional data, J. Japan Statist. Soc. 37 (2007) 53–86.
[20] M. Srivastava, M. Du, A test for the mean vector with fewer observations than the dimension, J. Multivariate Anal. 99 (2008) 386–402.
[21] C.A. Tsai, J.J. Chen, Multivariate analysis of variance test for gene set analysis, Bioinformatics 25 (2009) 897–903.
[22] V.M. Zolotarev, Concerning a certain probability problem, Theory Probab. Appl. 6 (1961) 201–204.