UC San Diego
UC San Diego Previously Published Works

Title
On Gaussian comparison inequality and its application to spectral analysis of large random matrices

Permalink
https://escholarship.org/uc/item/8t5003cm

Journal
Bernoulli, 24(3)

ISSN
1350-7265

Authors
Han, Fang
Xu, Sheng
Zhou, Wen-Xin

Publication Date
2018

DOI
10.3150/16-bej912

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Bernoulli **24**(3), 2018, 1787–1833
https://doi.org/10.3150/16-BEJ912

On Gaussian comparison inequality and its application to spectral analysis of large random matrices

FANG HAN[1], SHENG XU[2] and WEN-XIN ZHOU[3,4]

[1]*Department of Statistics, University of Washington, Seattle, WA 98195, USA. E-mail: fanghan@uw.edu*
[2]*Department of Statistics, Yale University, New Haven, CT 06511, USA. E-mail: sheng.xu@yale.edu*
[3]*Department of Mathematics, University of California, San Diego, La Jolla, CA 92093, USA.*
[4]*Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ 08544, USA. E-mail: stevewenxin@hotmail.com*

Recently, Chernozhukov, Chetverikov, and Kato (*Ann. Statist.* **42** (2014) 1564–1597) developed a new Gaussian comparison inequality for approximating the suprema of empirical processes. This paper exploits this technique to devise sharp inference on spectra of large random matrices. In particular, we show that two long-standing problems in random matrix theory can be solved: (i) simple bootstrap inference on sample eigenvalues when true eigenvalues are tied; (ii) conducting two-sample Roy's covariance test in high dimensions. To establish the asymptotic results, a generalized ε-net argument regarding the matrix rescaled spectral norm and several new empirical process bounds are developed and of independent interest.

Keywords: extreme value theory; Gaussian comparison inequality; random matrix theory; Roy's largest root test; spectral analysis

1. Introduction

Spectral analysis of large random matrices plays an important role in multivariate statistical estimation and testing problems. For example, variances of the principal components are functions of covariance eigenvalues (Muirhead [34]), and Roy's largest root test statistic is the spectral distance between the sample covariance and its population counterpart (Roy [37]).

Asymptotic behaviors of sample covariance eigenvalues have been extensively studied in the literature. When the dimension d is small and the population eigenvalues are distinct, Anderson [2] and Waternaux [44] proved the asymptotic normality for sample eigenvalues. Fujikoshi [20] established the Edgeworth expansion and showed that the convergence rate is of order $O(n^{-1/2})$ under various settings when d is fixed. For non-Gaussian data, Waternaux [44] and Fujikoshi [20] illustrated the effects of skewness and kurtosis on the limiting distribution.

When d is large, Johnstone [24] revealed for Gaussian data that the largest sample eigenvalue, after proper standardization, follows the Tracy–Widom law asymptotically (Tracy and Widom [39]). Johnstone [25] further proved that the convergence rate to the Tracy–Widom law is of order $O(d^{-2/3})$, which is astonishingly fast. Despite these elegant properties, existing results rely heavily on some simple Gaussian or sub-Gaussian assumptions (Bao, Pan and Zhou [4], Péché

1350-7265 © 2018 ISI/BS

[35], Pillai and Yin [36]). Their applications to hypothesis testing and constructing confidence intervals under more general settings are largely unknown.

Motivated by the covariance testing problem, the major focus of this paper is to study asymptotic behaviors of a particular type of spectral statistics related to the covariance matrix. Here we are interested in the non-Gaussian setting with the dimension d allowed to grow with the sample size n. Specifically, let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be n independent realizations of a d-dimensional random vector $\boldsymbol{X}$ with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$. Denote the sample covariance matrix by $\widehat{\boldsymbol{\Sigma}} = n^{-1}\sum_{i=1}^n \boldsymbol{X}_i\boldsymbol{X}_i^\mathsf{T}$. We shall derive the limiting distribution and establish bootstrap confidence intervals for the following statistic

$$\widehat{Q}_{\max} := \sup_{\|\boldsymbol{v}\|_2\le 1, \|\boldsymbol{v}\|_0\le s} \left| \frac{\sqrt{n}\boldsymbol{v}^\mathsf{T}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}}{\boldsymbol{v}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{v}} \right|, \tag{1.1}$$

where $1 \le s \le d$ is a prespecified integer-valued parameter representing the "degree of sparsity". The statistic $\widehat{Q}_{\max}$ is of general and strong practical interest. By setting $s = d$, it reduces to the conventional Roy's test statistic $\sqrt{n}\|\boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2} - \mathbf{I}_d\|_2$, where $\|\mathbf{M}\|_2$ denotes the spectral norm of $\mathbf{M}$. If $s \le d-1$, we obtain a generalized version of Roy's test statistic, allowing us to deal with large covariance matrices.[1]

To study the limiting behavior of $\widehat{Q}_{\max}$ in high dimensions, a major insight is to build the connection between the analysis of the maximum eigenvalue and recent developments in extreme value theory. In particular, by viewing the maximum eigenvalue as the extreme value of a specific infinite-state stochastic process, the Gaussian comparison inequality recently developed in Chernozhukov, Chetverikov and Kato [14] can be used. New empirical process bounds are established to ensure the validity of the inference procedure. In the end, bootstrap inference follows.

Two interesting observations are discovered. First, in the low-dimensional regime ($d/n \to 0$), the results in this paper solve a long standing question on bootstrap inference of eigenvalues when multiple roots exist (Beran and Srivastava [5], Eaton and Tyler [17]). The m-out-of-n bootstrap (Hall, Härdle and Simar [21]) is known to be rather sensitive to the choice of m. In comparison, the multiplier-bootstrap-based inference procedure used in this paper does not involve any tuning parameter, and is fairly accurate in approximating the distribution of the test statistic. Secondly, it is well known that Roy's largest root test is optimal against rank-one alternatives (Kritchman and Nadler [29]). Previously it was unclear whether such a result could be extended to high dimensional settings. This paper demonstrates that such a generalization can be made.

1.1. Notation

Throughout the paper, let $\mathbb{R}$ and $\mathbb{Z}$ denote the sets of real numbers and integers. Let $\mathbb{1}(\cdot)$ be the indicator function. Let $\boldsymbol{v} = (v_1, \ldots, v_d)^\mathsf{T}$ and $\mathbf{M} = (\mathbf{M}_{jk}) \in \mathbb{R}^{d\times d}$ be a d dimensional real vector and a $d \times d$ real matrix. For sets $I, J \subset \{1, \ldots, d\}$, let $\boldsymbol{v}_I$ be the subvector of $\boldsymbol{v}$ with entries

[1]The techniques built in this paper can also be exploited to study the non-normalized version of $\widehat{Q}_{\max}$, i.e., $\sup_{\|\boldsymbol{v}\|_2\le 1, \|\boldsymbol{v}\|_0\le s} |\boldsymbol{v}^\mathsf{T}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}|$. We defer to Section 2 for more details.

indexed by I, and $\mathbf{M}_{I,J}$ be the submatrix of $\mathbf{M}$ with entries indexed by I and J. We define the vector ℓ_0 and ℓ_2 (pseudo-)norms of $\boldsymbol{v}$ to be $\|\boldsymbol{v}\|_0 = \sum_j \mathbb{1}(v_j \neq 0)$ and $\|\boldsymbol{v}\|_2 = (\sum_{j=1}^d |v_j|^2)^{1/2}$. We define the matrix spectral (ℓ_2) norm as $\|\mathbf{M}\|_2 = \max_{\boldsymbol{v}} \|\mathbf{M}\boldsymbol{v}\|_2/\|\boldsymbol{v}\|_2$. For every real symmetric matrix $\mathbf{M}$, we define $\lambda_{\max}(\mathbf{M})$ and $\lambda_{\min}(\mathbf{M})$ to be its largest and smallest eigenvalues. For any integer $1 \leq s \leq d$ and real symmetric matrix $\mathbf{M}$, we define the s-sparse smallest and largest eigenvalues of $\mathbf{M}$ to be

$$\lambda_{\min,s}(\mathbf{M}) = \inf_{\boldsymbol{v}\in\mathbb{V}(s,d)} \boldsymbol{v}^\mathsf{T}\mathbf{M}\boldsymbol{v} \quad \text{and} \quad \lambda_{\max,s}(\mathbf{M}) = \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \boldsymbol{v}^\mathsf{T}\mathbf{M}\boldsymbol{v},$$

where

$$\mathbb{V}(s,d) := \left\{\boldsymbol{v} \in \mathbb{R}^d : \|\boldsymbol{v}\|_2 = 1, \|\boldsymbol{v}\|_0 \leq s\right\} \tag{1.2}$$

is the set of all s-sparse vectors on the d-dimensional sphere $\mathbb{S}^{d-1}$. Moreover, we write $\gamma_s(\mathbf{M}) = \sqrt{\lambda_{\max,s}(\mathbf{M})/\lambda_{\min,s}(\mathbf{M})}$ for any positive definite matrix $\mathbf{M}$. For any $\boldsymbol{v} \in \mathbb{R}^d$ and positive definite real-valued matrix $\mathbf{M}$, we write

$$\|\boldsymbol{v}\|_{\mathbf{M}} = \left(\boldsymbol{v}^\mathsf{T}\mathbf{M}\boldsymbol{v}\right)^{1/2} \quad \text{and} \quad \boldsymbol{v}_{\mathbf{M}} = \boldsymbol{v}/\|\boldsymbol{v}\|_{\mathbf{M}}.$$

For any random vectors $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^d$, we write $\boldsymbol{X} \overset{\mathsf{d}}{=} \boldsymbol{Y}$ if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are identically distributed. We use c, C to denote absolute positive constants, which may take different values at each occurrence. For any two real sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$, $a_n = O(b_n)$, or equivalently $b_n \gtrsim a_n$, if there exists an absolute constant C such that $|a_n| \leq C|b_n|$ for any large enough n. We write $a_n \asymp b_n$ if both $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold. We write $a_n = o(b_n)$ if for any absolute constant C, we have $|a_n| \leq C|b_n|$ for any large enough n. We write $a_n = O_{\mathbb{P}}(b_n)$ and $a_n = o_{\mathbb{P}}(b_n)$ if $a_n = O(b_n)$ and $a_n = o(b_n)$ hold stochastically. For arbitrary positive integer n, we write $[n] = \{a \in \mathbb{Z} : 1 \leq a \leq n\}$. For any set $\mathbb{A}$, denote by $|\mathbb{A}|$ its cardinality and $\operatorname{supp}(\mathbb{A})$ its support. For any $a, b \in \mathbb{R}$, we write $a \vee b = \max(a, b)$.

1.2. Structure of the paper

The rest of this paper is organized as follows. Section 2 provides the main results and technical tools involved. Sections 3 and 4 give two applications of the main results. In particular, Section 3 discusses the application to bootstrap inference of largest and smallest eigenvalues for spherical distributions. Section 4 extends the main results to conduct the two-sample Roy's largest root test. In Section 5, we conclude the paper with a short discussion. The technical proofs are relegated to Section 6.

2. Main results

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent and identically distributed (i.i.d.) realizations of $\boldsymbol{X} \in \mathbb{R}^d$ with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, and let $\widehat{\boldsymbol{\Sigma}}$ be the sample covariance matrix. Define $\widehat{Q}_{\boldsymbol{v}}$ and

$\widehat{Q}_{\max}$ to be the normalized rank-one projection and normalized s-sparse largest singular value of $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$, given respectively, by

$$\widehat{Q}_{\boldsymbol{v}} = \frac{\sqrt{n}\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v}} \quad \text{and} \quad \widehat{Q}_{\max} = \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} |\widehat{Q}_{\boldsymbol{v}}|.$$

We aim to derive the limiting distribution of $\widehat{Q}_{\max}$. Of note, when setting $s = d$ and assuming the positive definiteness of $\boldsymbol{\Sigma}$, we have

$$\widehat{Q}_{\max} = \sqrt{n}\left\|\boldsymbol{\Sigma}^{-1/2}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\Sigma}^{-1/2} - \mathbf{I}_d\right\|_2,$$

which coincides with Roy's largest root test statistic (Johnstone and Nadler [27], Roy [37]). The statistic $\widehat{Q}_{\max}$ is of strong practical interest. We shall discuss in Sections 3 and 4 two applications based on its limiting properties stated below.

To derive the limiting distribution of $\widehat{Q}_{\max}$, we impose the following two assumptions.

Assumption 2.1. There exists a random vector $\boldsymbol{U} \in \mathbb{R}^d$ satisfying $\mathbb{E}(\boldsymbol{U}) = \mathbf{0}$ and $\mathbb{E}(\boldsymbol{U}\boldsymbol{U}^{\mathsf{T}}) = \mathbf{I}_d$, such that

$$\boldsymbol{X} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{U} \quad \text{and} \quad K_1 := \sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}} \left\|\boldsymbol{v}^{\mathsf{T}}\boldsymbol{U}\right\|_{\psi_2} < \infty.$$

Here $\|\cdot\|_{\psi_2}$ stands for the standard Orlicz norm with respect to the function $\psi_2(x) := \exp(x^2) - 1$ (van der Vaart and Wellner [42]).

Assumption 2.2. $\{\boldsymbol{X}_i\}_{i=1}^n$ are independent realizations of $\boldsymbol{X}$.

Assumptions 2.1 and 2.2 are mild and are regularly imposed in the literature. Note that the sub-Gaussian condition in Assumption 2.1 can be easily relaxed at the cost of a more stringent scaling constraint on (n, d, s) (Cai, Liu and Xia [8]). Assumption 2.2 can also be slightly relaxed. Such relaxations are beyond the scope of this paper, and we will not pursue the details here.

With Assumptions 2.1 and 2.2 satisfied, the following theorem gives a Gaussian comparison result regarding the limiting distribution of $\widehat{Q}_{\max}$. Below, for an arbitrary set $\mathbb{A}$ equipped with a metric ρ, we call $\mathbb{N}_\varepsilon$ an ε-net of $\mathbb{A}$ if for every $a \in \mathbb{A}$ there exists some $a' \in \mathbb{N}_\varepsilon$ such that $\rho(a, a') \le \varepsilon$. For each $s \in [d]$, with slight abuse of notation, we write $\gamma_s = \gamma_s(\boldsymbol{\Sigma})$ for simplicity. By Lemma 6.1 in Section 6, for any $\varepsilon \in (0, 1)$, there exists an ε-net $\mathbb{N}_\varepsilon^0 = \{\boldsymbol{v}_j : j = 1, \ldots, p_\varepsilon^0\}$ of $\mathbb{V}(s, d)$ equipped with the Euclidean metric, with its cardinality $p_\varepsilon^0 = |\mathbb{N}_\varepsilon^0|$ satisfying $\log p_\varepsilon^0 \lesssim s\log(ed/\varepsilon s)$.

Theorem 2.1. *Let Assumptions 2.1 and 2.2 be satisfied and put $\varepsilon_1 = (n\gamma_s)^{-1}$. Then, for any ε_1-net $\mathbb{N}_{\varepsilon_1}$ of $\mathbb{V}(s, d)$ with cardinality $p_{\varepsilon_1} = |\mathbb{N}_{\varepsilon_1}|$, there exists a p_{ε_1}-dimensional centered Gaussian random vector $(G_1, \ldots, G_{p_{\varepsilon_1}})^{\mathsf{T}}$ satisfying $\mathbb{E}(G_jG_k) = \mathbb{E}(R_jR_k)$ for $1 \le j, k \le p_{\varepsilon_1}$ with $R_j = \boldsymbol{v}_j^{\mathsf{T}}(\boldsymbol{X}\boldsymbol{X}^{\mathsf{T}} - \boldsymbol{\Sigma})\boldsymbol{v}_j/\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v}_j$, such that*

$$\sup_{t\ge 0}\left|\mathbb{P}(\widehat{Q}_{\max} \le t) - \mathbb{P}\Big(\max_{1\le j\le p_{\varepsilon_1}} |G_j| \le t\Big)\right| \le CK_2^2\frac{\{\gamma_n(s, d) \vee \log p_{\varepsilon_1}\}^{9/8}}{n^{1/8}}, \tag{2.1}$$

where $C > 0$ *is an absolute constant,* $K_2 := K_1^2 + 1$, *and* $\gamma_n(s,d) := s\log(\gamma_s \cdot ed/s) \vee s\log n$.

There are several interesting observations drawn from Theorem 2.1. First, as long as $s\log(\gamma_s \cdot ed/s) \vee s\log n \vee \log p_\varepsilon = o(n^{1/9})$ for a properly chosen ε, the distribution of $\widehat{Q}_{\max}$ can be well approximated by that of the maximum of a Gaussian sequence. It is worth noting that no parametric assumption is imposed on the data generating scheme. Secondly, the result in Theorem 2.1, though not reflecting the exact limiting distribution of $\widehat{Q}_{\max}$, sheds light on its asymptotic behavior. Following the standard extreme value theory, when $s = 1$ and the covariance matrix $\mathbf{\Sigma}$ is sparse, $\widehat{Q}_{\max}$ follows a Gumbel distribution asymptotically as $d \to \infty$ (Cai, Liu and Xia [8]). Thirdly, we note that when $\mathbf{\Sigma} = \mathbf{I}_d$, the techniques used to prove Theorem 2.1 can be adapted to derive the limiting distributions of extreme sample eigenvalues. See Section 3 for details.

The detailed proof of Theorem 2.1 is involved. Hence, a heuristic sketch is useful. A major ingredient stems from a Gaussian comparison inequality recently developed by Chernozhukov, Chetverikov and Kato [14].

Lemma 2.1 (Gaussian comparison inequality). *Let* $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ *be independent random vectors in* $\mathbb{R}^d$ *with mean zero and finite absolute third moments, that is,* $\mathbb{E}(X_{ij}) = 0$ *and* $\mathbb{E}(|X_{ij}|^3) < \infty$ *for all* $1 \le i \le n$ *and* $1 \le j \le d$. *Consider the statistic* $Z := \max_{1\le j\le d}\sum_{i=1}^n X_{ij}$. *Let* $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_n$ *be independent random vectors in* $\mathbb{R}^d$ *with* $\boldsymbol{Y}_i \sim N_d(\mathbf{0}, \mathbb{E}(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}))$, $1 \le i \le n$. *Then, for every* $\delta > 0$, *there exists a random variable* $\widetilde{Z} \overset{\mathrm{d}}{=} \max_{1\le j\le d}\sum_{i=1}^n Y_{ij}$ *such that*

$$\mathbb{P}\big(|Z - \widetilde{Z}| \ge 16\delta\big) \lesssim \delta^{-2}\log(dn)\big\{D_1 + \delta^{-1}\log(dn)(D_2 + D_3)\big\} + n^{-1}\log n,$$

where we write

$$D_1 = \mathbb{E}\left[\max_{1\le j,l\le d}\left|\sum_{i=1}^n\big\{X_{ij}X_{il} - \mathbb{E}(X_{ij}X_{il})\big\}\right|\right], \qquad D_2 = \mathbb{E}\left(\max_{1\le j\le d}\sum_{i=1}^n |X_{ij}|^3\right),$$

$$D_3 = \sum_{i=1}^n \mathbb{E}\left[\max_{1\le j\le d}|X_{ij}|^3 \cdot \mathbb{1}\left\{\max_{1\le j\le d}|X_{ij}| > \frac{\delta}{\log(dn)}\right\}\right].$$

In view of Lemma 2.1 and the fact that $\widehat{Q}_{\max}$ is the supremum of an infinite-state process, the proof can be divided into three steps. In the first step, we prove that the difference between $\widehat{Q}_{\max}$ and its "discretized version" is negligible asymptotically. This is implied by the following generalized ε-net argument for the rescaled spectral norm. It extends the standard ε-net argument (Vershynin [43]).

Lemma 2.2. *For any* $\boldsymbol{v}, \widetilde{\boldsymbol{v}} \in \mathbb{V}(s,d)$ *with the same support, positive definite matrix* $\mathbf{\Sigma}$, *and any real symmetric matrix* $\mathbf{M}$, *we have*

$$\big|\big|\boldsymbol{v}_{\mathbf{\Sigma}}^{\mathsf{T}}\mathbf{M}\boldsymbol{v}_{\mathbf{\Sigma}}\big| - \big|\widetilde{\boldsymbol{v}}_{\mathbf{\Sigma}}^{\mathsf{T}}\mathbf{M}\widetilde{\boldsymbol{v}}_{\mathbf{\Sigma}}\big|\big| \le 2\gamma_s\|\boldsymbol{v} - \widetilde{\boldsymbol{v}}\|_2 \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\big|\boldsymbol{v}_{\mathbf{\Sigma}}^{\mathsf{T}}\mathbf{M}\boldsymbol{v}_{\mathbf{\Sigma}}\big|.$$

In the second step, we show that this discretized version of $\widehat{Q}_{\max}$ converges in distribution to the maximum of a finite Gaussian sequence. This can be achieved by exploiting Lemma 2.1. Lastly, anti-concentration bounds (Chernozhukov, Chetverikov and Kato [15]) are established to bridge the gap between the distributions of $\widehat{Q}_{\max}$ and its discretized version. The complete proof is provided in Section 6.

The asymptotic result in Theorem 2.1 is difficult to use in practice. To estimate the limiting distribution of $\widehat{Q}_{\max}$ empirically, bootstrap approximation is preferred. For any $\boldsymbol{v} \in \mathbb{V}(s,d)$, define

$$\widehat{B}_{\boldsymbol{v}} = \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n \xi_i (\boldsymbol{v}^\top \boldsymbol{X}_i \boldsymbol{X}_i^\top \boldsymbol{v} - \boldsymbol{v}^\top \boldsymbol{\Sigma} \boldsymbol{v})}{\boldsymbol{v}^\top \boldsymbol{\Sigma} \boldsymbol{v}} \quad \text{and} \quad \widehat{B}_{\max} = \sup_{\boldsymbol{v} \in \mathbb{V}(s,d)} |\widehat{B}_{\boldsymbol{v}}|, \tag{2.2}$$

where $\xi_1, \dots, \xi_n$ are i.i.d. standard normal that are independent of $\{\boldsymbol{X}_i\}_{i=1}^n$. We use the conditional distribution of $\widehat{B}_{\max}$ given the data to approximate the distribution of $\widehat{Q}_{\max}$. The next theorem characterizes the validity of bootstrap approximation.

Theorem 2.2. *Let Assumptions 2.1 and 2.2 be satisfied, and assume that $\gamma_n(s,d) = s\log(\gamma_s \cdot ed/s) \vee s\log n = o(n^{1/9})$ as $n \to \infty$. Then, there exists a sufficiently large absolute constant $C > 0$ such that*

$$\mathbb{P}\left\{ \sup_{t \geq 0} \left| \mathbb{P}(\widehat{Q}_{\max} \leq t) - \mathbb{P}(\widehat{B}_{\max} \leq t | \boldsymbol{X}_1, \dots, \boldsymbol{X}_n) \right| \geq C \frac{\gamma_n^{9/8}(s,d)}{n^{1/8}} \right\} = o(1).$$

In other words, we have

$$\sup_{t \geq 0} \left| \mathbb{P}(\widehat{Q}_{\max} \leq t) - \mathbb{P}(\widehat{B}_{\max} \leq t | \boldsymbol{X}_1, \dots, \boldsymbol{X}_n) \right| = o_{\mathbb{P}}(1).$$

The proof of Theorem 2.2 heavily relies on characterizing the convergence rates of sub-Gaussian fourth-order terms. We defer this result and the detailed proof of Theorem 2.2 to Section 6.

The rest of this section gives asymptotic results for the non-normalized version of $\widehat{Q}_{\max}$. To this end, let $\widetilde{Q}_{\boldsymbol{v}}$ and $\widetilde{Q}_{\max}$ be the rank-one projection and s-sparse largest singular value of $\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}$, given respectively, by

$$\widetilde{Q}_{\boldsymbol{v}} = \sqrt{n} \boldsymbol{v}^\top (\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}) \boldsymbol{v} \quad \text{and} \quad \widetilde{Q}_{\max} = \sup_{\boldsymbol{v} \in \mathbb{V}(s,d)} |\widetilde{Q}_{\boldsymbol{v}}|.$$

Technically speaking, $\widetilde{Q}_{\max}$ is a simpler version of $\widehat{Q}_{\max}$. We show that, under an additional eigenvalue assumption, $\widetilde{Q}_{\max}$ converges weakly to the extreme of a Gaussian sequence. In particular, the following condition assumes that the s-sparse (restricted) largest eigenvalue of $\boldsymbol{\Sigma}$ is upper bounded by an absolute constant.

Assumption 2.3. There exists an absolute constant $L > 0$ such that $\lambda_{\max,s}(\boldsymbol{\Sigma}) \leq L$.

We define, for any $\boldsymbol{v} \in \mathbb{V}(s,d)$,

$$\widetilde{B}_{\boldsymbol{v}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \big(\boldsymbol{v}^{\mathsf{T}} \boldsymbol{X}_i \boldsymbol{X}_i^{\mathsf{T}} \boldsymbol{v} - \boldsymbol{v}^{\mathsf{T}} \boldsymbol{\Sigma} \boldsymbol{v} \big) \quad \text{and} \quad \widetilde{B}_{\max} = \sup_{\boldsymbol{v} \in \mathbb{V}(s,d)} |\widetilde{B}_{\boldsymbol{v}}|, \tag{2.3}$$

where $\xi_1, \ldots, \xi_n$ are i.i.d. standard normal random variables independent of $\{\boldsymbol{X}_i\}_{i=1}^n$. The following theorem gives the Gaussian approximation result for $\widetilde{Q}_{\max}$.

Theorem 2.3. *Let Assumptions 2.1–2.3 be satisfied and set $\varepsilon_2 = n^{-1}$. Then, for any ε_2-net $\mathbb{N}_{\varepsilon_2}$ of $\mathbb{V}(s,d)$ with cardinality $p_{\varepsilon_2} = |\mathbb{N}_{\varepsilon_2}|$, there exists a p_{ε_2}-dimensional centered Gaussian random vector $(\widetilde{G}_1, \ldots, \widetilde{G}_{p_{\varepsilon_2}})^{\mathsf{T}}$ satisfying $\mathbb{E}(\widetilde{G}_j \widetilde{G}_k) = \mathbb{E}(\widetilde{R}_j \widetilde{R}_k)$ for $1 \le j,k \le p_{\varepsilon_2}$ with $\widetilde{R}_j := \boldsymbol{w}_j^{\mathsf{T}}(\boldsymbol{X}\boldsymbol{X}^{\mathsf{T}} - \boldsymbol{\Sigma})\boldsymbol{w}_j$, such that*

$$\sup_{t \ge 0} \Big| \mathbb{P}(\widetilde{Q}_{\max} \le t) - \mathbb{P}\Big(\max_{1 \le j \le p_{\varepsilon_2}} |\widetilde{G}_j| \le t \Big) \Big| \le C_L K_2^2 \frac{\delta_n^{9/8}(s,d)}{n^{1/8}}, \tag{2.4}$$

where $C_L > 0$ is a constant depending only on L, $K_2 = K_1^2 + 1$, and $\delta_n(s,d) := s\log(ed/s) \vee s\log n$. In addition, if (s,d,n) satisfies $s\log(ed/s) \vee s\log n = o(n^{1/9})$ as $n \to \infty$, then there exists an absolute constant $C > 0$ large enough such that

$$\mathbb{P}\Big\{ \sup_{t \ge 0} \big| \mathbb{P}(\widetilde{Q}_{\max} \le t) - \mathbb{P}(\widetilde{B}_{\max} \le t | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \big| \ge C \frac{\delta_n^{9/8}(s,d)}{n^{1/8}} \Big\} = o(1). \tag{2.5}$$

In other words, we have

$$\sup_{t \ge 0} \big| \mathbb{P}(\widetilde{Q}_{\max} \le t) - \mathbb{P}(\widetilde{B}_{\max} \le t | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \big| = o_{\mathbb{P}}(1).$$

Remark 2.1. By comparing Theorems 2.1 and 2.3, we immediately observe some difference between the properties of $\widehat{Q}_{\max}$ and $\widetilde{Q}_{\max}$. To ensure the validity of the multiplier bootstrap approximation for $\widehat{Q}_{\max}$, we only require $s\log(\gamma_s ed/s) \vee s\log n = o(n^{1/9})$, and thus allow $\lambda_{\max,s}(\boldsymbol{\Sigma})$ to grow quickly. In contrast, the bootstrap approximation consistency for $\widetilde{Q}_{\max}$ relies on C_L, a constant of the same order as $\lambda_{\max,s}^2(\boldsymbol{\Sigma})$.

3. Application I: Bootstrap inference on largest and smallest eigenvalues for spherical distributions

A direct application of Theorem 2.1 is on inferring extreme sample eigenvalues of spherical distributions. A random vector is said to be spherically distributed if its covariance matrix is proportional to the identity. Note that this definition is slightly different from its counterpart in robust statistics, where a more stringent rotation-invariant property is required (Fang, Kotz and Ng [19]).

It is known that when multiple roots exist (i.e., the population eigenvalues are not distinct), the sample eigenvalues are not asymptotically normal even under the Gaussian assumption (Anderson [2]). Waternaux [44] and Tyler [41] showed that inference is even more challenging for non-Gaussian data as the limiting distributions of the sample eigenvalues rely on the skewness and kurtosis of the underlying distribution. Estimation of these parameters is statistically costly. Bootstrap methods are hence recommended for conducting inference.

In the presence of multiple roots, Beran and Srivastava (Beran and Srivastava [5]) pointed out that the nonparametric bootstrap for eigenvalue inference is inconsistent. The m-out-of-n bootstrap (Hall, Härdle and Simar [21]) and its modification (Hall *et al.* [22]) are hence proposed to correct this. The implementation, however, is complicated since tuning parameters are involved.

Based on Theorem 2.2, we show that a simple multiplier bootstrap method leads to asymptotically valid inference for extreme eigenvalues, as stated in the next theorem.

Theorem 3.1. *Suppose that Assumptions 2.1 and 2.2 hold. In addition, assume that* $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_d$ *with* $\sigma^2 > 0$ *an absolute constant. Then, as long as* $d = o(n^{1/9})$,

$$\sup_t \left| \mathbb{P}\big\{\lambda_{\max}(\widehat{\mathbf{\Sigma}}) \le \sigma^2 + t\big\} - \mathbb{P}\left[\lambda_{\max}\left\{\frac{1}{n}\sum_{i=1}^n \xi_i\big(\boldsymbol{X}_i\boldsymbol{X}_i^\top - \sigma^2\mathbf{I}_d\big)\right\} \le t \,\middle|\, \boldsymbol{X}_1, \dots, \boldsymbol{X}_n\right]\right| = o_{\mathbb{P}}(1),$$

and

$$\sup_t \left| \mathbb{P}\big\{\lambda_{\min}(\widehat{\mathbf{\Sigma}}) \ge \sigma^2 + t\big\} - \mathbb{P}\left[\lambda_{\min}\left\{\frac{1}{n}\sum_{i=1}^n \xi_i\big(\boldsymbol{X}_i\boldsymbol{X}_i^\top - \sigma^2\mathbf{I}_d\big)\right\} \le t \,\middle|\, \boldsymbol{X}_1, \dots, \boldsymbol{X}_n\right]\right| = o_{\mathbb{P}}(1).$$

Here $\{\xi_i\}_{i=1}^n$ *forms an independent standard Gaussian sequence independent of the data.*

Theorem 3.1 shows that the distributions of $\lambda_{\min}(\widehat{\mathbf{\Sigma}})$ and $\lambda_{\max}(\widehat{\mathbf{\Sigma}})$ can be consistently estimated by those of their bootstrapped counterparts. Such an approximation is data adaptive, and does not require any parametric or semiparametric (such as elliptical distribution) assumption on the data generating scheme. Thus, such an inference procedure enjoys the distribution-free property (Kendall and Stuart [28]). In addition, the implementation is simple, and does not involve any tuning parameter.

Note that in Theorem 3.1, we allow the dimension to slightly increase with the sample size. This is a more relaxed setting than that for the conventional bootstrap inference (Beran and Srivastava [5]). Assumption 2.1 is required for the case that d increases. However, when d is fixed, this assumption can be easily relaxed.

Finally, a comment on the scaling condition $d = o(n^{1/9})$ could be instructive. In detail, we aim to explore how sharp this condition is. For this, we generate $n = 200$, 500, and 1000 data points from the multivariate standard Gaussian $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \sim N_d(\mathbf{0}, \mathbf{I}_d)$. We increase d from 2, 5, 10, 20, 50, to 100. Figure 1 illustrates the exact distribution of $\lambda_{\max}(\widehat{\mathbf{\Sigma}})$ (denoted as "exact") and its multiplier-bootstrap and nonparametric bootstrap based counterparts (denoted as "m-boots" and "n-boots").

Figure 1 shows that, for $n = 200$, the multiplier-bootstrap-based approach well approximates the exact distribution for d not greater than 10. For $n = 500$ and 1000, the dimension d can be as

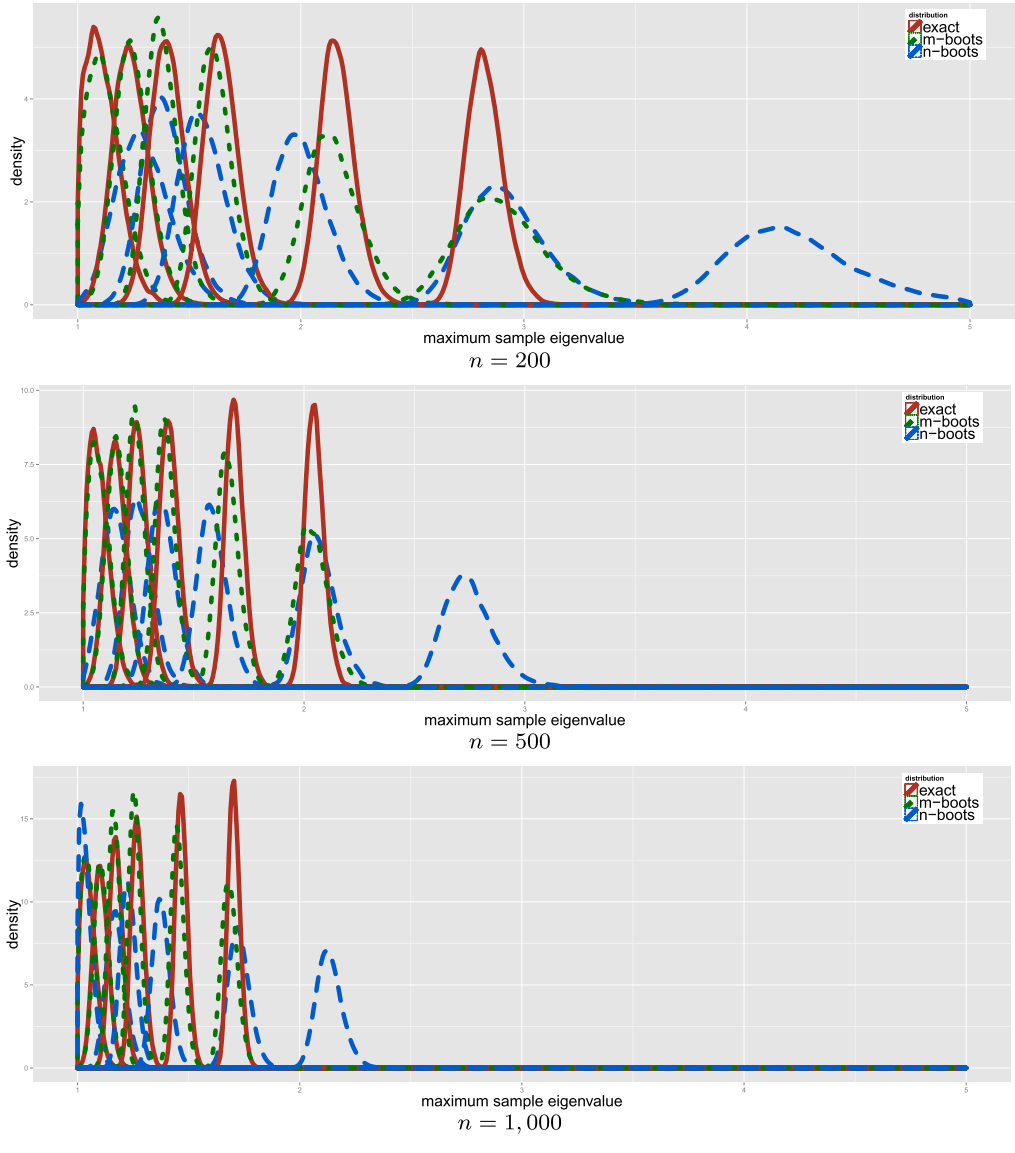

Figure 1. Density plots of three approximation distributions based on $n = 200, 500$, and 1000 data points (from top to bottom) randomly drawn from the standard Gaussian distribution. These three distributions are: (i) the exact distribution of $\lambda_{\max}(\widehat{\mathbf{\Sigma}})$ (solid, denoted as "exact"); (ii) the distribution from the multiplier-bootstrap method (dotted, denoted as "m-boots"); (iii) the distribution from the nonparametric-bootstrap method (broken, denoted as "n-boots"). Within each graph, for each distribution, the curves correspond to the settings $d = 2, 4, 10, 20, 50, 100$ from left to right. The calculation is based on 40 000 replications.

large as 20 and 50 to ensure reasonable approximation results. This indicates that the dependence between n and d to guarantee efficient bootstrap approximation is almost linear, while we do need d to be reasonably small compared to n. Also, Figure 1 shows that the nonparametric-bootstrap method leads to an extremely biased estimate of the distribution of $\lambda_{\max}(\widehat{\boldsymbol{\Sigma}})$.

4. Application II: Two-sample Roy's largest root test

In multivariate analysis, tests for the equality of covariance matrices are of central interest (Anderson [3]). High dimensionality brings new challenges, and many existing methods cannot be used.

Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ and $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ be n and m independent realizations of centered random vectors $\boldsymbol{X} \in \mathbb{R}^d$ and $\boldsymbol{Y} \in \mathbb{R}^d$ with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. In this section, we aim to test the hypothesis

$$\mathbf{H_0}: \quad \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \quad \text{versus} \quad \mathbf{H_1}: \quad \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$$

under the scenario where d is allowed to grow with n and m.

We first briefly review the literature on testing $\mathbf{H_0}$ in high dimensions. Johnstone and Nadler [27] pointed out that the most common tests fall into two categories: the first is based on "linear statistics" of the eigenvalues, and the second is based on extreme value statistics. In the high dimensional setting, Chen, Zhang and Zhong [13], Li and Chen [30], Cai and Ma [10], among many others, have proposed tests based on linear statistics. Asymptotic normality is established even when d/n tends to infinity. Initiated by Jiang [23], another track of tests is developed based on extreme values of the entries of the sample covariance matrix. Cai, Liu and Xia [8] studied the problem of testing the equality of two unknown covariance matrices for possibly non-Gaussian data in the "sparsity" scenario. Recently, Chang *et al.* [12] proposed a bootstrap procedure to conduct inference for the test statistic considered in Cai, Liu and Xia [8] and relaxed the sparsity assumption.

Though significant progress has been made in this area, there has not been much research on Roy's largest root type tests (Roy [37]), an important method in covariance testing. Such tests are built on extreme eigenvalues and are optimal against low-rank alternatives (Kritchman and Nadler [29]). Absence of the corresponding results in high dimensions is largely due to the uncommon behavior of extreme eigenvalues. Built on the results derived in Section 2, we are able to fill this gap. This is done via exploiting a sparse-PCA-type thinking which is advocated by Iain Johnstone and many others (Cai, Ma and Wu [11], Johnstone and Lu [26], Ma [31]) in dealing with large random matrices. The techniques we developed here generalize those in Cai, Liu and Xia [8] and Chang *et al.* [12], and are of independent interest.

4.1. Method

The proposed test extends Roy's largest root test to high dimensions. In detail, let $\widehat{\boldsymbol{\Sigma}}_1$ and $\widehat{\boldsymbol{\Sigma}}_2$ be the sample covariance matrices given by

$$\widehat{\boldsymbol{\Sigma}}_1 = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^{\mathsf{T}} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_2 = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{Y}_i \boldsymbol{Y}_i^{\mathsf{T}}.$$

Let s be a prespecified parameter characterizing the sparsity level we wish to balance. To guarantee valid inference, we recommend s to be chosen less than $\min(n, m)/10$. Recall in (1.2) that $\mathbb{V}(s, d)$ represents the set of all s-sparse vectors in the unit sphere $\mathbb{S}^{d-1}$. Let $\widehat{\mathcal{Q}}_{\boldsymbol{v}}$ and $\widehat{\mathcal{Q}}_{\max}$ be the normalized rank-one projection and normalized s-sparse largest singular value of $\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}}_2$:

$$\widehat{\mathcal{Q}}_{\boldsymbol{v}} = \sqrt{\frac{n+m}{nm}} \frac{\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}}_2)\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1/n + \widehat{\boldsymbol{\Sigma}}_2/m)\boldsymbol{v}} \quad \text{and} \quad \widehat{\mathcal{Q}}_{\max} = \sup_{\boldsymbol{v} \in \mathbb{V}(s,d)} |\mathcal{Q}_{\boldsymbol{v}}|.$$

The proposed test is multiplier-bootstrap-based. In detail, we define

$$\widehat{\mathcal{B}}_{\boldsymbol{v}} = \sqrt{\frac{n+m}{nm}} \frac{\boldsymbol{v}^{\mathsf{T}}\{\sum_{i=1}^n \xi_i(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}} - \widehat{\boldsymbol{\Sigma}}_1)/n - \sum_{i=1}^m \eta_i(\boldsymbol{Y}_i\boldsymbol{Y}_i^{\mathsf{T}} - \widehat{\boldsymbol{\Sigma}}_2)/m\}\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1/n + \widehat{\boldsymbol{\Sigma}}_2/m)\boldsymbol{v}}$$

and

$$\widehat{\mathcal{B}}_{\max} = \sup_{\boldsymbol{v} \in \mathbb{V}(s,d)} |\widehat{\mathcal{B}}_{\boldsymbol{v}}|,$$

where $\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_m$ are independent standard Gaussian random variables that are independent of $\{\boldsymbol{X}_i\}_{i=1}^n$ and $\{\boldsymbol{Y}_i\}_{i=1}^m$. Let $\mathcal{B}_1^*, \ldots, \mathcal{B}_N^*$ be N realizations of $\widehat{\mathcal{B}}_{\max}$ (via fixing the data and changing $\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_m$) for some large enough N.[2] Let q_α be the corresponding $(1-\alpha)$th quantile. The proposed test is

$$T_\alpha = \mathbb{1}(\widehat{\mathcal{Q}}_{\max} \geq q_\alpha), \tag{4.1}$$

and we reject $\mathbf{H_0}$ whenever $T_\alpha = 1$.

Remark 4.1. Computing the extreme eigenvalues of large covariance matrices under sparsity constraint involves a combinatorial optimization and is NP-complete in general. Several computationally feasible methods based on the recent developments in the sparse PCA literature can be used to compute $\widehat{\mathcal{Q}}_{\max}$ approximately. The theoretical guarantees, however, remain unclear. For example, a greedy search to shrink the candidate set (Moghaddam, Weiss and Avidan [33]), followed by a second-step brutal search, may work well in practice. Recently, Berthet and Rigollet [6] proposed a computationally efficient method using convex relaxations to compute their sparse eigenvalue statistic for sparse principal component testing. It is interesting to investigate whether their method can be adapted to deal with the current problem. We leave this to future work.

Remark 4.2. A "non-normalized" version of the test T_α, based on the restricted spectral gap $\sup_{\boldsymbol{v} \in \mathbb{V}(s)} |\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}}_2)\boldsymbol{v}|$, can be similarly defined and calculated using the truncated power method (Yuan and Zhang [45]). However, boundeness on the restricted eigenvalue $\lambda_{\max,s}(\boldsymbol{\Sigma}_1)$ is required for the validity of the non-normalized test.

[2]In the sequel, for ease of presentation, we focus on the ideal case that we know the exact conditional distribution of $\widehat{\mathcal{B}}_{\max}$ given the data. This is equivalent to setting N infinitely large. In practice, the accuracy of bootstrap by setting a finitely large N is guaranteed by the Dvoretzky-Kiefer-Wolfowitz inequality (Dvoretzky, Kiefer and Wolfowitz [16]).

4.2. Theory

This section provides the theoretical properties of T_α in (4.1). First, we show that the size of the test is well controlled. Secondly, we study the power and prove the minimax optimality for the proposed test against "low-rank" alternatives.

To ensure the size consistency of T_α, we require the following two assumptions on the data generating scheme. They are analogous to those in Section 2. Of note, we do not require $\boldsymbol{X} \overset{\mathsf{d}}{=} \boldsymbol{Y}$.

Assumption 4.1. There exist random vectors $\boldsymbol{U}_1, \boldsymbol{U}_2 \in \mathbb{R}^d$ satisfying $\mathbb{E}(\boldsymbol{U}_i) = \mathbf{0}$ and $\mathbb{E}(\boldsymbol{U}_i \boldsymbol{U}_i^\mathsf{T}) = \mathbf{I}_d$ for $i = 1, 2$, such that

$$\boldsymbol{X} = \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{U}_1, \qquad \boldsymbol{Y} = \boldsymbol{\Sigma}_2^{1/2} \boldsymbol{U}_2, \qquad L_1 := \max\Big(\sup_{\boldsymbol{v} \in \mathbb{S}^{d-1}} \big\| \boldsymbol{v}^\mathsf{T} \boldsymbol{U}_1 \big\|_{\psi_2}, \sup_{\boldsymbol{v} \in \mathbb{S}^{d-1}} \big\| \boldsymbol{v}^\mathsf{T} \boldsymbol{U}_2 \big\|_{\psi_2} \Big) < \infty.$$

Assumption 4.2. $\{\boldsymbol{X}_i\}_{i=1}^n$ and $\{\boldsymbol{Y}_i\}_{i=1}^m$ are independent realizations of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Moreover, the sample sizes are comparable, that is, $n \asymp m$.

The following result provides the theoretical guarantee for the validity of the multiplier bootstrap test T_α.

Theorem 4.1. *Let Assumptions 4.1 and 4.2 be satisfied. Under the null hypothesis* $\mathbf{H_0}$, *we have*

$$\sup_{t \geq 0} \big| \mathbb{P}(\widehat{\mathcal{Q}}_{\max} \leq t) - \mathbb{P}(\widehat{\mathcal{B}}_{\max} \leq t | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m) \big| = o_{\mathbb{P}}(1) \tag{4.2}$$

whenever $s \log\{\gamma_s(\boldsymbol{\Sigma}_1) ed/s\} \vee s \log n = o(n^{1/9})$. *Moreover, as* $n, m \to \infty$,

$$\mathbb{P}_{\mathbf{H_0}}(T_\alpha = 1) = \alpha + o(1)$$

uniformly in $0 < \alpha < 1$.

Next, we analyze the power of the test T_α and show that it is minimax rate-optimal. For this, we consider the alternative class of matrices

$$\mathbb{M}(\lambda) := \Bigg\{ (\mathbf{M}_1, \mathbf{M}_2) \in \mathbb{R}^{d \times d} : \lambda_{\max,s}(\mathbf{M}_\ell)/\lambda_{\min,s}(\mathbf{M}_\ell) \leq C, \ell = 1, 2, \\ \sqrt{\frac{n+m}{nm}} \sup_{\boldsymbol{v} \in \mathbb{V}(s,d)} \left| \frac{\boldsymbol{v}^\mathsf{T}(\mathbf{M}_1 - \mathbf{M}_2)\boldsymbol{v}}{\boldsymbol{v}^\mathsf{T}(\mathbf{M}_1/n + \mathbf{M}_2/m)\boldsymbol{v}} \right| \geq \lambda \sqrt{s \log(ed/s)} \Bigg\}$$

for some constant $C > 0$ independent of (n, m, d). The following two theorems illustrate the power and minimax lower bound in differentiating two covariance matrices within some matrix set $\mathbb{M}(\lambda)$.

Theorem 4.2 (Power analysis). *Suppose that Assumptions 4.1 and 4.2 hold. Further assume that (n, m, d, s) satisfies $s\log(ed/s)\log n = o(n)$. Then, for all sufficiently large $\lambda > 0$, we have*

$$\inf_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathbb{M}(\lambda)} \mathbb{P}_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)}(T_\alpha = 1) \to 1, \qquad \text{as } n \to \infty,$$

where $\mathbb{P}_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)}$ represents the joint distribution of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ with covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$.

Theorem 4.3 (Minimax lower bound). *Assume that the conditions in Theorem 4.2 hold. Then, for all sufficiently small $\lambda > 0$, we have*

$$\inf_{\Phi_\alpha\in\mathcal{T}_\alpha} \sup_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\in\mathbb{M}(\lambda)} \mathbb{P}_{(\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)}(\Phi_\alpha = 0) \geq 1 - \alpha - o(1), \qquad \text{as } n \to \infty,$$

where $\mathcal{T}_\alpha = \{\Phi_\alpha : \mathbb{P}_{\mathbf{H_0}}(\Phi_\alpha = 1) \leq \alpha\}$ denotes the class of all α-level tests.

Remark 4.3. The scaling condition, $s\log(ed/s)\log n = o(n)$, in Theorems 4.2 and 4.3, is weaker than the corresponding one in Theorem 4.1. Accordingly, combining Theorems 4.1, 4.2, and 4.3, under Assumptions 4.1 and 4.2, the scaling condition in Theorem 4.1, and the boundedness assumption on $\lambda_{\max,s}(\boldsymbol{\Sigma}_\ell)/\lambda_{\min,s}(\boldsymbol{\Sigma}_\ell)$ for $\ell = 1, 2$, the proposed test is minimax rate-optimal.

Remark 4.4. The $\log n$ term in the scaling condition, $s\log(ed/s)\log n = o(n)$, is required for handling the multiplier-bootstrap-based statistic $\widehat{\mathcal{B}}_{\max}$, which involves the cubes of sub-Gaussian random variables. We tackle it via a truncation argument, which is also exploited in Cai and Liu [7] and Cai, Liu and Xia [8]. Similar scaling conditions are also posed therein.

4.3. Empirical results

In this section, we compare the numerical performance of the proposed approach with two existing ones. Specifically, we consider the following tests.

- HXZ3: the proposed covariance test with $s = 3$;
- HXZ5: the proposed covariance test with $s = 5$;
- HXZ10: the proposed covariance test with $s = 10$;
- CZZW: the L_∞-type bootstrap test proposed in Chang *et al.* [12];
- LC: the L_2-type covariance test proposed in Li and Chen [30].

These five tests represent the three notable tracks in covariance testing problems. Specifically, our proposed tests are more sensitive to low-rank alternatives, while CZZW and LC are more sensitive to elementwise changes of the covariance matrix. To implement the proposed tests and that of CZZW, we take the bootstrap sample size to be 1000.

We consider three settings for the structure of $\boldsymbol{\Sigma}_1$. Let $\mathbf{O}$ be a diagonal matrix with its diagonals generated from a uniform distribution Unif(0.5, 1.5). Set $\boldsymbol{\Sigma}_1 = \mathbf{O}\boldsymbol{\Sigma}^*\mathbf{O}$, where $\boldsymbol{\Sigma}^* = [\boldsymbol{\Sigma}^*_{jk}] \in \mathbb{R}^{d\times d}$ is specified as follows:

- *long range*: $\mathbf{\Sigma}^*_{jk} = \mathbb{1}(j = k) + 0.5\mathbb{1}(j \neq k)$;
- *short range*: $\mathbf{\Sigma}^*_{jk} = 0.1^{|j-k|}$;
- *isotropic*: $\mathbf{\Sigma}^* = \mathbf{I}_d$.

To compare the power, we consider three types of alternatives:

- *alternative* 1: $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1 + c_1 \cdot \boldsymbol{v}_S \boldsymbol{v}_S^\mathsf{T}$, where $\boldsymbol{v}_S$ is sparse with the support size 5 and non-zero entries all equal to $1/\sqrt{5}$;
- *alternative* 2: $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_1 + c_2\mathbf{D}^{(1)}$, where $\mathbf{D}^{(1)}$ satisfies $\mathbf{D}^{(1)}_{12} = \mathbf{D}^{(1)}_{21} = 1$ and has 0 elsewhere;
- *alternative* 3: $\mathbf{\Sigma}_2 = (\mathbf{I}_d + c_3\mathbf{D}^{(2)})^\mathsf{T}\mathbf{\Sigma}_1(\mathbf{I}_d + c_3\mathbf{D}^{(2)})$, where $\mathbf{D}^{(2)}$ satisfies $\mathbf{D}^{(2)}_{j+1,j} = 1$ for $j = 1, \ldots, d-1$ and has 0 elsewhere.

We also generate $n = m$ independent samples from $N_d(\mathbf{0}, \mathbf{\Sigma}_1)$ and $N_d(\mathbf{0}, \mathbf{\Sigma}_2)$ separately. Here we set $n = 20, 500, 1000$ and $d = 40, 100$. The values of c_1, c_2, c_3 are specified as follows:

- *long range*: Set $(c_1, c_2, c_3) = (0.9, 0.35, 0.1)$ for $d = 40$, and set $(c_1, c_2, c_3) = (1.1, 0.4, 0.1)$ for $d = 100$;
- *short range*: Set $(c_1, c_2, c_3) = (0.7, 0.4, 0.1)$ for $d = 40$, and set $(c_1, c_2, c_3) = (0.85, 0.4, 0.1)$ for $d = 100$;
- *isotropic*: Set $(c_1, c_2, c_3) = (0.8, 0.3, 0.12)$ for $d = 40$, and set $(c_1, c_2, c_3) = (0.85, 0.3, 0.1)$ for $d = 100$.

We repeat the simulation 5000 times. Tables 1–3 illustrate the size and powers (corresponding to the three alternatives) for each combination of n and d. There are several noteworthy obser-

Table 1. (Long range) Comparison of the five competing tests under the null and alternatives 1 to 3 when $n = 200, 500, 1000$ and $d = 40, 100$. The results are computed based on 5000 replications

		Empirical size					Empirical power (alternative 1)				
d	n	HXZ3	HXZ5	HXZ10	CZZW	LC	HXZ3	HXZ5	HXZ10	CZZW	LC
40	200	0.031	0.022	0.000	0.073	0.109	0.230	**0.342**	0.284	0.108	0.102
	500	0.046	0.036	0.013	0.055	0.095	0.802	0.911	**0.931**	0.358	0.146
	1000	0.045	0.042	0.027	0.040	0.085	0.990	0.997	**1.000**	0.756	0.150
100	200	0.018	0.007	0.000	0.088	0.111	0.120	**0.184**	0.064	0.089	0.105
	500	0.044	0.016	0.003	0.035	0.093	0.824	0.878	**0.894**	0.340	0.106
	1000	0.043	0.029	0.018	0.048	0.108	0.993	**0.996**	0.995	0.770	0.122
		Empirical power (alternative 2)					Empirical power (alternative 3)				
40	200	0.121	0.202	**0.347**	0.220	0.106	0.068	0.036	0.007	0.269	**0.398**
	500	0.141	0.210	0.386	**0.788**	0.096	0.211	0.123	0.078	0.503	**0.654**
	1000	0.120	0.203	0.431	**0.999**	0.089	0.537	0.505	0.353	0.854	**0.923**
100	200	0.042	0.071	0.147	**0.252**	0.123	0.044	0.003	0.000	0.200	**0.369**
	500	0.079	0.097	0.158	**0.847**	0.118	0.170	0.100	0.028	0.550	**0.671**
	1000	0.052	0.095	0.185	**1.000**	0.101	0.525	0.418	0.255	0.840	**0.912**

Table 2. (Short range) Comparison of the five competing tests under the null and alternatives 1 to 3 when $n = 200, 500, 1000$ and $d = 40, 100$. The results are computed based on 5000 replications

		Empirical size					Empirical power (alternative 1)				
d	n	HXZ3	HXZ5	HXZ10	CZZW	LC	HXZ3	HXZ5	HXZ10	CZZW	LC
40	200	0.023	0.014	0.001	0.039	0.030	0.075	0.053	0.002	0.083	**0.160**
	500	0.036	0.028	0.024	0.066	0.024	0.566	**0.570**	0.465	0.390	0.213
	1000	0.048	0.028	0.032	0.050	0.020	0.936	0.932	**0.943**	0.848	0.426
100	200	0.007	0.001	0.000	0.039	0.127	0.021	0.035	0.001	0.033	**0.209**
	500	0.030	0.020	0.005	0.044	0.034	**0.362**	0.333	0.186	0.223	0.206
	1000	0.036	0.030	0.010	0.047	0.026	0.894	**0.901**	0.898	0.756	0.196
		Empirical power (alternative 2)					Empirical power (alternative 3)				
40	200	0.041	0.020	0.004	**0.179**	0.070	0.027	0.009	0.002	0.154	**0.364**
	500	0.104	0.132	0.081	**0.701**	0.021	0.183	0.192	0.106	0.470	**0.730**
	1000	0.123	0.181	0.294	**0.994**	0.046	0.486	0.462	0.513	0.975	**0.993**
100	200	0.028	0.020	0.002	0.088	**0.159**	0.037	0.025	0.001	0.075	**0.467**
	500	0.052	0.037	0.013	**0.580**	0.161	0.097	0.087	0.025	0.258	**0.756**
	1000	0.080	0.098	0.107	**0.981**	0.169	0.199	0.215	0.168	0.759	**0.990**

Table 3. (Isotropic) Comparison of the five competing tests under the null and alternatives 1 to 3 when $n = 200, 500, 1000$ and $d = 40, 100$. The results are computed based on 5000 replications

		Empirical size					Empirical power (alternative 1)				
d	n	HXZ3	HXZ5	HXZ10	CZZW	LC	HXZ3	HXZ5	HXZ10	CZZW	LC
40	200	0.029	0.013	0.000	0.061	0.042	0.053	0.039	0.009	0.076	**0.159**
	500	0.044	0.015	0.015	0.052	0.053	0.435	**0.500**	0.443	0.253	0.410
	1000	0.042	0.044	0.029	0.050	0.061	0.872	0.920	**0.939**	0.763	0.854
100	200	0.015	0.001	0.000	0.036	0.058	0.035	0.005	0.000	0.062	**0.099**
	500	0.039	0.016	0.006	0.060	0.040	0.430	**0.455**	0.322	0.250	0.238
	1000	0.039	0.032	0.017	0.033	0.065	0.900	0.904	**0.908**	0.806	0.554
		Empirical power (alternative 2)					Empirical power (alternative 3)				
40	200	0.038	0.017	0.006	**0.461**	0.097	0.025	0.007	0.001	0.091	**0.269**
	500	0.120	0.169	0.198	**0.990**	0.288	0.075	0.066	0.070	0.289	**0.729**
	1000	0.141	0.228	0.421	**1.000**	0.604	0.213	0.223	0.223	0.787	**0.994**
100	200	0.030	0.010	0.000	**0.312**	0.073	0.009	0.005	0.000	0.079	**0.253**
	500	0.065	0.063	0.026	**0.982**	0.106	0.043	0.044	0.017	0.223	**0.769**
	1000	0.079	0.135	0.189	**1.000**	0.207	0.160	0.146	0.079	0.820	**0.999**

vations. First, empirical sizes of the proposed tests and CZZW are well controlled, while the size of LC is inflated in the "long range" setting. This is as expected since the "long range" dependence violates the assumptions in Li and Chen [30]. Secondly, regarding the empirical powers, our proposed tests outperform those of CZZW and LC under alternative 1, which is in line with Theorems 4.1 and 4.3; CZZW outperforms the others under alternative 2, and LC outperforms the others under alternative 3, as expected. Finally, we see that HXZ5 performs the best on average in alternative 1, while HXZ3 and HXZ10 perform similarly to HXZ5, and all outperform CZZW and LC. Hence, for testing against low-rank alternatives, our proposed test is more favorable and the choice of s is rather flexible as long as it remains at a moderate size.

5. Discussion

Spectral analysis for large random matrices has a long history and maintains one of the most active research areas in statistics. Recent advances include the discovery of the Tracy–Widom law, an important family of distributions that quantifies the fluctuation of sample eigenvalues. A vast literature follows. However, more questions are raised than answered. In particular, no result has been promised for extensions to non-Gaussian distributions with a nontrivial covariance structure. This paper fills this long-standing gap from a new perspective grown in the literature of extreme value theory. The obtained results prove to work in many cases which for a long time are known to be challenging to deal with.

Very recently, Fan and Wang [18] studied asymptotic behaviors of sample covariance eigenvalues under a pervasive assumption, that is, the largest eigenvalue grows quickly with the dimension. Under this assumption, they proved the asymptotic normality for the sample eigenvalues. In comparison, our results are built on the normalized covariance matrix and are obtained in the settings where the signals are not too strong. A natural question arises that whether a phase transition phenomenon occurs when signals change from weak to strong. In particular, how do the asymptotic distributions of sample eigenvalues change with the growing magnitudes of extreme eigenvalues? We conjecture that this problem may be related to the normal mean problem in extreme value theory, and leave that question for future research.

6. Proofs

This section contains the proofs of the results in this paper.

6.1. Proof of the main results

6.1.1. *Proof of Theorem* 2.1

We first give an outline of the proof, which consists of three main steps. (i) In the first step, we approximate $\widehat{Q}_{\max}$, the supremum over a continuous function space induced by $\mathbb{V}(s, d)$, by the maximum over a discrete function space induced by $\mathbb{N}_{\varepsilon_1}$, for ε_1 as in Theorem 2.1. (ii) In the second step, we show that the above discretized version of $\widehat{Q}_{\max}$ over $\mathbb{N}_{\varepsilon_1}$ converges weakly to

the maximum of a Gaussian sequence. (iii) Lastly, we employ the anti-concentration inequality (Lemma 6.11) to complete the proof.

Step I. Let $\varepsilon \in (0, 1)$ be an arbitrary number. We first employ the following lemma to connect the supremum over a continuous function space induced by $\mathbb{V}(s, d)$ to the maximum over a discrete function space induced by $\mathbb{N}_\varepsilon$.

Lemma 6.1. *There exists an ε-net $\mathbb{N}_\varepsilon^0$ of $\mathbb{V}(s, d)$ equipped with the Euclidean metric satisfying that $\log p_\varepsilon^0 = \log |\mathbb{N}_\varepsilon^0| \lesssim s \log \frac{ed}{\varepsilon s}$. Further, for any ε-net $\mathbb{N}_\varepsilon$ of $\mathbb{V}(s, d)$, we have*

$$\widehat{Q}_{\max} \le 2\gamma_s \varepsilon \cdot \widehat{Q}_{\max} + M_{\max,\varepsilon},$$

where

$$M_{\max,\varepsilon} := \max_{\boldsymbol{v} \in \mathbb{N}_\varepsilon} |Q_{\boldsymbol{v}}| = \max_{\boldsymbol{v} \in \mathbb{N}_\varepsilon} \left| \frac{\sqrt{n}\boldsymbol{v}^\top(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}}{\boldsymbol{v}^\top \boldsymbol{\Sigma} \boldsymbol{v}} \right|. \tag{6.1}$$

Lemma 6.1 and the fact $\widehat{Q}_{\max} \ge M_{\max,\varepsilon}$ yield that, for any $\varepsilon \in (0, 1)$,

$$|\widehat{Q}_{\max} - M_{\max,\varepsilon}| \le 2\gamma_s \varepsilon \cdot \widehat{Q}_{\max}. \tag{6.2}$$

Next we use Lemma 6.2 below to bound $\widehat{Q}_{\max}$. Note that, by Lemma 6.9, we have for any $\boldsymbol{v} \in \mathbb{S}^{d-1}$ that $\|\boldsymbol{v}^\top \boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{v} - 1\|_{\psi_1} \le \|(\boldsymbol{v}^\top \boldsymbol{U})^2\|_{\psi_1} + 1 = \|\boldsymbol{v}^\top \boldsymbol{U}\|_{\psi_2}^2 + 1$. Taking maximum over $\boldsymbol{v} \in \mathbb{S}^{d-1}$ on both sides yields

$$\sup_{\boldsymbol{v} \in \mathbb{S}^{d-1}} \left\| \boldsymbol{v}^\top \boldsymbol{U}\boldsymbol{U}^\top \boldsymbol{v} - 1 \right\|_{\psi_1} \le K_1^2 + 1 = K_2.$$

Lemma 6.2. *For any $t > 0$, we have*

$$\mathbb{P}\left[\widehat{Q}_{\max} \le C_{12} K_2 \gamma_n^{1/2}(s, d) + K_2 \max\left\{ 2\sqrt{2t}, C_{11}\gamma_n(s, d)\frac{t}{\sqrt{n}} \right\} \right] \ge 1 - 4e^{-t},$$

where $C_{11}, C_{12} > 0$ are absolute constants, $K_2 = K_1^2 + 1$ and $\gamma_n(s, d) = s \log(\gamma_s \cdot ed/s) \vee s \log n$ are as in Theorem 2.1.

Using Lemma 6.2, it follows from (6.2) that for any $t > 0$ and $\varepsilon \in (0, 1)$,

$$\mathbb{P}\left(|\widehat{Q}_{\max} - M_{\max,\varepsilon}| \le C K_2 \gamma_s \varepsilon \left[\gamma_n^{1/2}(s, d) + \max\left\{ \sqrt{t}, \gamma_n(s, d)\frac{t}{\sqrt{n}} \right\} \right] \right) \ge 1 - 4e^{-t}.$$

Taking $\varepsilon = \varepsilon_1 = (n\gamma_s)^{-1}$ and $t = \log n$, we deduce that

$$\begin{aligned} &\mathbb{P}\left(\left| \widehat{Q}_{\max} - \max_{\boldsymbol{v} \in \mathbb{N}_{\varepsilon_1}} |Q_{\boldsymbol{v}}| \right| \le C K_2 \left[\frac{\gamma_n^{1/2}(s, d)}{n} + \max\left\{ \frac{\sqrt{\log n}}{n}, \gamma_n(s, d)\frac{\log n}{n^{3/2}} \right\} \right] \right) \\ &\quad \ge 1 - \frac{4}{n}. \end{aligned} \tag{6.3}$$

Step II. For any $\varepsilon \in (0,1)$, write $\mathbb{N}_\varepsilon = \{\boldsymbol{v}_j : j = 1, \dots, p_\varepsilon\}$ for the ε-net $\mathbb{N}_\varepsilon$ constructed in *Step* I, and recall that for $i = 1, \dots, n$ and $j = 1, \dots, p_\varepsilon$,

$$R_{ij} = \frac{\boldsymbol{v}_j^\mathsf{T}(\boldsymbol{X}_i \boldsymbol{X}_i^\mathsf{T} - \boldsymbol{\Sigma})\boldsymbol{v}_j}{\boldsymbol{v}_j^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{v}_j}.$$

It follows from (6.1) that

$$M_{\max,\varepsilon} = \max_{\boldsymbol{v}\in\mathbb{N}_\varepsilon} |Q_{\boldsymbol{v}}| = \max_{1\le j\le p_\varepsilon} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n R_{ij} \right|.$$

The following lemma gives a Gaussian coupling inequality for $M_{\max,\varepsilon}$.

Lemma 6.3. *For every $\delta > 0$, we have*

$$\begin{aligned}
&\mathbb{P}\Big(\Big|M_{\max,\varepsilon} - \max_{1\le j\le p_\varepsilon} |G_j|\Big| \ge 16\delta\Big) \\
&\quad\lesssim K_2^2 \frac{\sqrt{\log(2p_\varepsilon)}\log(2p_\varepsilon \vee n)}{\delta^2 n^{1/2}} + K_0^2 K_2^2 \frac{\log(2p_\varepsilon)\{\log(np_\varepsilon+1)\}^2 \log(2p_\varepsilon \vee n)}{\delta^2 n} \\
&\qquad + K_2^3 \frac{\{\log(2p_\varepsilon \vee n)\}^2}{\delta^3 n^{1/2}} + K_0^3 K_2^3 \frac{(\log p_\varepsilon)\{\log(np_\varepsilon+1)\}^3\{\log(2p_\varepsilon \vee n)\}^2}{\delta^3 n^{3/2}} \\
&\qquad + K_0^4 K_2^4 \frac{\{\log(p_\varepsilon+1)\}^4\{\log(2p_\varepsilon \vee n)\}^3}{\delta^4 n} + \frac{\log n}{n},
\end{aligned}$$

where $M_{\max,\varepsilon}$ is as in (6.1), K_0 is the constant in Lemma 6.13 with $\alpha = 1$ and $K_2 = K_1^2 + 1$.

In view of Lemma 6.3, by taking $\varepsilon = \varepsilon_1$ we have

$$\begin{aligned}
&P\Big(\Big|\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}} |Q_{\boldsymbol{v}}| - \max_{1\le j\le p_{\varepsilon_1}} |G_j|\Big| \ge 16\delta\Big) \\
&\quad\lesssim K_2^3 \frac{\{\log(2p_{\varepsilon_1} \vee n)\}^2}{\delta^3 n^{1/2}} + K_0^4 K_2^4 \frac{\{\log(p_{\varepsilon_1}+1)\}^4\{\log(2p_{\varepsilon_1} \vee n)\}^3}{\delta^4 n},
\end{aligned}$$

where $p_{\varepsilon_1} = |\mathbb{N}_{\varepsilon_1}|$. Putting $\{G_{\boldsymbol{v}}\}_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}} = \{G_j\}_{j=1}^{p_{\varepsilon_1}}$, we have

$$\begin{aligned}
&P\Big(\Big|\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}} |Q_{\boldsymbol{v}}| - \max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}} |G_{\boldsymbol{v}}|\Big| \ge 16\delta\Big) \\
&\quad\lesssim K_2^3 \frac{(\log p_{\varepsilon_1})^2}{\delta^3 n^{1/2}} + K_0^4 K_2^4 \frac{(\log p_{\varepsilon_1})^7}{\delta^4 n}.
\end{aligned} \tag{6.4}$$

Without loss of generality, assume that $\log p_{\varepsilon_1} \ge \gamma_n(s,d)$ (the case when $\log p_{\varepsilon_1} < \gamma_n(s,d)$ can be similarly dealt with by replacing all $\log p_{\varepsilon_1}$ below by $\gamma_n(s,d)$). Combining (6.3) and (6.4),

we have

$$\begin{aligned}\mathbb{P}\bigg\{\Big|\widehat{Q}_{\max}-\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}}|G_{\boldsymbol{v}}|\Big|\geq 16\delta+CK_2\bigg(\frac{\sqrt{\log p_{\varepsilon_1}}}{n}+\log p_{\varepsilon_1}\frac{\log n}{n^{3/2}}\bigg)\bigg\}\\ \lesssim K_2^3\frac{(\log p_{\varepsilon_1})^2}{\delta^3 n^{1/2}}+K_0^4K_2^4\frac{(\log p_{\varepsilon_1})^7}{\delta^4 n},\end{aligned}\tag{6.5}$$

where $C>0$ is an absolute constant.

Step III. Taking $\delta=\frac{K_0^4(\log p_{\varepsilon_1})^{5/8}}{K_2 n^{1/8}}$, we deduce from (6.5) that there exists an absolute positive constant C such that

$$\mathbb{P}\bigg\{\Big|\widehat{Q}_{\max}-\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}}|G_{\boldsymbol{v}}|\Big|\geq CK_0^4\frac{(\log p_{\varepsilon_1})^{5/8}}{K_2 n^{1/8}}\bigg\}\lesssim K_2^2\frac{(\log p_{\varepsilon_1})^{1/8}}{n^{1/8}}+K_2^2\frac{(\log p_{\varepsilon_1})^{9/2}}{n^{1/2}}.\tag{6.6}$$

By Lemma 6.11, we have

$$\sup_{x\in\mathbb{R}}\mathbb{P}\bigg\{\Big|\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}}|G_{\boldsymbol{v}}|-x\Big|\leq C_{13}K_0^4\frac{(\log p_{\varepsilon_1})^{5/8}}{K_2 n^{1/8}}\bigg\}\lesssim K_2^2\frac{(\log p_{\varepsilon_1})^{9/8}}{n^{1/8}}\tag{6.7}$$

for some absolute constant $C_{13}>0$. Note that, for every $t>0$ and $\eta>0$,

$$\begin{aligned}&\Big|\mathbb{P}(\widehat{Q}_{\max}\leq t)-\mathbb{P}\Big(\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}}|G_{\boldsymbol{v}}|\leq t\Big)\Big|\\ &\quad\leq\mathbb{P}\Big(\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}}|G_{\boldsymbol{v}}|\in[t-\eta,t+\eta]\Big)+\mathbb{P}\Big(\Big|\widehat{Q}_{\max}-\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}}|G_{\boldsymbol{v}}|\Big|>\eta\Big).\end{aligned}$$

Taking $\eta=C_{13}\frac{K_0^4(\log p_{\varepsilon_1})^{5/8}}{K_2 n^{1/8}}$ in the last display, we deduce from (6.6) and (6.7) that

$$\sup_{t\geq 0}\Big|\mathbb{P}(\widehat{Q}_{\max}\leq t)-\mathbb{P}\Big(\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}}|G_{\boldsymbol{v}}|\leq t\Big)\Big|\lesssim K_2^2\frac{(\log p_{\varepsilon_1})^{9/8}}{n^{1/8}}+K_2^2\frac{(\log p_{\varepsilon_1})^{9/2}}{n^{1/2}}.$$

This completes the proof.

6.1.2. *Proof of Lemma 2.2*

Proof. Noting that

$$\begin{aligned}\|\boldsymbol{v}_{\boldsymbol{\Sigma}}-\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\|_{\boldsymbol{\Sigma}}^2&=2-2\frac{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{v}}}{(\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v})^{1/2}(\widetilde{\boldsymbol{v}}^{\mathsf{T}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{v}})^{1/2}}\\ &=\frac{(\boldsymbol{v}-\widetilde{\boldsymbol{v}})^{\mathsf{T}}\boldsymbol{\Sigma}(\boldsymbol{v}-\widetilde{\boldsymbol{v}})-\{(\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v})^{1/2}-(\widetilde{\boldsymbol{v}}^{\mathsf{T}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{v}})^{1/2}\}^2}{(\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v})^{1/2}(\widetilde{\boldsymbol{v}}^{\mathsf{T}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{v}})^{1/2}},\end{aligned}$$

we have

$$\|\boldsymbol{v}_{\boldsymbol{\Sigma}}-\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\|_{\boldsymbol{\Sigma}}^2\leq\frac{(\boldsymbol{v}-\widetilde{\boldsymbol{v}})^{\mathsf{T}}\boldsymbol{\Sigma}(\boldsymbol{v}-\widetilde{\boldsymbol{v}})}{(\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v})^{1/2}(\widetilde{\boldsymbol{v}}^{\mathsf{T}}\boldsymbol{\Sigma}\widetilde{\boldsymbol{v}})^{1/2}}\leq\gamma_s^2(\boldsymbol{\Sigma})\cdot\|\boldsymbol{v}-\widetilde{\boldsymbol{v}}\|_2^2.\tag{6.8}$$

By the triangle inequality,

$$\begin{aligned}\big|\big|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\boldsymbol{v}_{\boldsymbol{\Sigma}}\big| - \big|\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\big|\big| &\le \big|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\big| \\ &= \big|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}(\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}) + (\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}})^{\mathsf{T}}\mathbf{M}\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\big|.\end{aligned}$$

It follows that

$$\big|\big|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\boldsymbol{v}_{\boldsymbol{\Sigma}}\big| - \big|\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\big|\big| \le \big|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}(\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}})\big| + \big|(\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}})^{\mathsf{T}}\mathbf{M}\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\big|.$$

Using Lemma 6.8, we deduce that

$$\begin{aligned}\big|\big|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\boldsymbol{v}_{\boldsymbol{\Sigma}}\big| - \big|\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\big|\big| &\le \left|\left(\frac{\boldsymbol{v}_{\boldsymbol{\Sigma}}}{\|\boldsymbol{v}_{\boldsymbol{\Sigma}}\|_2}\right)_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\left(\frac{\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}}{\|\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\|_2}\right)_{\boldsymbol{\Sigma}}\right| \cdot \|\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\|_{\boldsymbol{\Sigma}} \\ &\quad + \left|\left(\frac{\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}}{\|\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\|_2}\right)_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\left(\frac{\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}}{\|\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\|_2}\right)_{\boldsymbol{\Sigma}}\right| \cdot \|\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\|_{\boldsymbol{\Sigma}} \\ &\le 2\|\boldsymbol{v}_{\boldsymbol{\Sigma}} - \widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\|_{\boldsymbol{\Sigma}} \cdot \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \big|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\boldsymbol{v}_{\boldsymbol{\Sigma}}\big|.\end{aligned} \tag{6.9}$$

Combining (6.8) and (6.9) gives

$$\big|\big|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\boldsymbol{v}_{\boldsymbol{\Sigma}}\big| - \big|\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\big|\big| \le 2\gamma_s(\boldsymbol{\Sigma})\|\boldsymbol{v} - \widetilde{\boldsymbol{v}}\|_2 \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \big|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\mathbf{M}\boldsymbol{v}_{\boldsymbol{\Sigma}}\big|,$$

as desired. □

6.1.3. *Proof of Theorem 2.2*

Proof. Based on the ε_1-net $\mathbb{N}_{\varepsilon_1}^0$ described in Lemma 6.1 and the corresponding p_1-dimensional Gaussian random vector $(G_1, \ldots, G_{p_1})^{\mathsf{T}}$ introduced in the proof of Theorem 2.1 with $p_1 := p_{\varepsilon_1}^0 = |\mathbb{N}_{\varepsilon_1}^0|$, we aim to show that

$$\sup_{t\ge 0}\left|\mathbb{P}\Big(\max_{1\le j\le p_1}|G_j| \le t\Big) - \mathbb{P}(\widehat{B}_{\max} \le t|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)\right| = o_{\mathbb{P}}(1).$$

In view of Theorem 2.1, it suffices to prove that

$$\sup_{t\ge 0}\left|\mathbb{P}\Big(\max_{1\le j\le p_1}|G_j| \le t\Big) - \mathbb{P}\Big(\max_{1\le j\le p_1}|\widehat{B}_j| \le t\Big|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\Big)\right| = o_{\mathbb{P}}(1), \tag{6.10}$$

where $\{\widehat{B}_j\}_{j=1}^{p_1} = \{\widehat{B}_{\boldsymbol{v}}\}_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_1}^0}$. In particular, we note that

$$\sup_{t\ge 0}\left|\mathbb{P}(\widehat{B}_{\max} \le t|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) - \mathbb{P}\Big(\max_{1\le j\le p_1}|\widehat{B}_j| \le t\Big|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\Big)\right| = O_{\mathbb{P}}\left\{\frac{\gamma_n^{9/8}(s,d)}{n^{1/8}}\right\},$$

via the proof of Theorem 2.1.

By Lemma 6.19, we have,

$$\begin{aligned}
&\sup_{t\geq 0}\Big|\mathbb{P}\Big(\max_{1\leq j\leq p_1}|G_j|\leq t\Big)-\mathbb{P}\Big(\max_{1\leq j\leq p_1}|\widehat{B}_j|\leq t\,\big|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n\Big)\Big| \\
&\quad\lesssim \Delta_G^{1/3}\big\{\log(2p_1)\big\}^{1/3}\big\{1\vee 2\log(2p_1)\vee\log(1/\Delta_G)\big\}^{1/3} \\
&\quad\lesssim \Delta_G^{1/3}\big\{\log(2p_1)\big\}^{2/3}\vee\Delta_G^{1/3}\big\{\log(1/\Delta_G)\big\}^{1/3}\big\{\log(2p_1)\big\}^{1/3},
\end{aligned}\tag{6.11}$$

where $p_1=|\mathbb{N}^0_{\varepsilon_1}|$ satisfies $\log p_1\lesssim\gamma_n(s,d)=s\log(\gamma_s\cdot ed/s)\vee s\log n$ and

$$\Delta_G:=\max_{1\leq j\leq k\leq p_1}\big|\mathbb{E}(G_jG_k)-\mathbb{E}(\widehat{B}_j\widehat{B}_k|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n)\big|.$$

Next we bound Δ_G. For $1\leq j\leq k\leq p_1$, we have

$$\mathbb{E}(G_jG_k)=\mathbb{E}\bigg\{\frac{(\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{X}_i)^2(\boldsymbol{v}_k^{\mathsf{T}}\boldsymbol{X}_i)^2}{\|\boldsymbol{v}_j\|_{\boldsymbol{\Sigma}}^2\|\boldsymbol{v}_k\|_{\boldsymbol{\Sigma}}^2}\bigg\}-1.$$

By definition (2.2), we have, for $j=1,\ldots,p_1$,

$$\widehat{B}_j=\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i\bigg\{\frac{(\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{X}_i)^2}{\|\boldsymbol{v}_j\|_{\boldsymbol{\Sigma}}^2}-1\bigg\}.$$

It follows that, for $1\leq j\leq k\leq p_1$,

$$\begin{aligned}
&\mathbb{E}(\widehat{B}_j\widehat{B}_k|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n) \\
&\quad=\frac{1}{n}\sum_{i=1}^n\bigg\{\frac{(\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{X}_i)^2(\boldsymbol{v}_k^{\mathsf{T}}\boldsymbol{X}_i)^2}{\|\boldsymbol{v}_j\|_{\boldsymbol{\Sigma}}^2\|\boldsymbol{v}_k\|_{\boldsymbol{\Sigma}}^2}\bigg\}-\frac{1}{n}\sum_{i=1}^n\bigg\{\frac{(\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{X}_i)^2}{\|\boldsymbol{v}_j\|_{\boldsymbol{\Sigma}}^2}-1\bigg\}-\frac{1}{n}\sum_{i=1}^n\bigg\{\frac{(\boldsymbol{v}_k^{\mathsf{T}}\boldsymbol{X}_i)^2}{\|\boldsymbol{v}_k\|_{\boldsymbol{\Sigma}}^2}-1\bigg\}-1.
\end{aligned}$$

For simplicity, we define

$$W_{ij}=\frac{\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{X}_i}{\|\boldsymbol{v}_j\|_{\boldsymbol{\Sigma}}}\quad\text{and}\quad W_{ik}=\frac{\boldsymbol{v}_k^{\mathsf{T}}\boldsymbol{X}_i}{\|\boldsymbol{v}_k\|_{\boldsymbol{\Sigma}}},\qquad i=1,\ldots,n,\,1\leq j\leq k\leq p_1.\tag{6.12}$$

In this notation, we have

$$\begin{aligned}
\Delta_G&=\max_{1\leq j\leq k\leq p_1}\bigg|\frac{1}{n}\sum_{i=1}^n\big\{(W_{ij}W_{ik})^2-\mathbb{E}(W_{ij}W_{ik})^2\big\}+\frac{1}{n}\sum_{i=1}^n\big(W_{ij}^2-1\big)+\frac{1}{n}\sum_{i=1}^n\big(W_{ik}^2-1\big)\bigg| \\
&\leq\max_{1\leq j\leq k\leq p_1}\bigg|\frac{1}{n}\sum_{i=1}^n\big\{(W_{ij}W_{ik})^2-\mathbb{E}(W_{ij}W_{ik})^2\big\}\bigg|+2\max_{1\leq j\leq p_1}\bigg|\frac{\boldsymbol{v}_j^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{v}_j}{\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v}_j}\bigg|.
\end{aligned}$$

Further, define

$$\Delta_{G,1} = \max_{1\le j\le k\le p_1}\left|\frac{1}{n}\sum_{i=1}^{n}\{(W_{ij}W_{ik})^2 - \mathbb{E}(W_{ij}W_{ik})^2\}\right|,$$

$$\Delta_{G,2} = 2\max_{1\le j\le p_1}\left|\frac{\boldsymbol{v}_j^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}_j}{\boldsymbol{v}_j^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v}_j}\right|.$$

The following lemma gives an upper bound for $\Delta_{G,1}$.

Lemma 6.4. *For any $M > 0$, there exists an absolute positive constant C_{21} only depending on M such that*

$$\mathbb{P}\left[\max_{1\le j,k\le p_1}\left|\frac{1}{n}\sum_{i=1}^{n}\{(W_{ij}W_{ik})^2 - \mathbb{E}(W_{ij}W_{ik})^2\}\right| \ge C_{21}\sqrt{\frac{\log p_1}{n}}\right] = O\big(p_1^{-M}\big),$$

where $p_1 = |\mathbb{N}_{\varepsilon_1}^0|$, W_{ij} and W_{ik} for $1 \le i \le n$ and $1 \le j, k \le p_1$ are defined in (6.12).

By Lemma 6.4, there exists an absolute positive constant C_{21} depending only on M such that

$$\mathbb{P}\left(\Delta_{G,1} \ge C_{21}\sqrt{\frac{\log p_1}{n}}\right) = O\big(p_1^{-M}\big). \tag{6.13}$$

Turning to $\Delta_{G,2}$, by Lemma 6.2, there exists a constant $C > 0$ depending only on K_2 such that

$$\mathbb{P}\left[\Delta_{G,2} \ge C\left\{\frac{\gamma_n^{1/2}(s,d)}{n^{1/2}} + \gamma_n(s,d)\frac{\log n}{n}\right\}\right] \le \frac{4}{n}. \tag{6.14}$$

Combining (6.13) and (6.14), we have with probability greater than $1 - O(p_1^{-M})$,

$$\Delta_G^{1/3}\big\{\log(2p_1)\big\}^{2/3} \le C\frac{\{\log(2p_1)\}^{2/3}\gamma_n^{1/6}(s,d)}{n^{1/6}}. \tag{6.15}$$

Since $x \mapsto x\log(1/x)$ is non-decreasing for $0 < x \le e^{-1}$, we have with probability greater than $1 - O(p_1^{-M})$,

$$\Delta_G^{1/3}\big\{\log(2p_1)\big\}^{1/3}\big\{\log(1/\Delta_G)\big\}^{1/3} \le C\frac{\{\log(2p_1)\}^{2/3}\gamma_n^{1/6}(s,d)}{n^{1/6}}. \tag{6.16}$$

Putting (6.11), (6.15), and (6.16) together, we conclude that

$$\mathbb{P}\left\{\sup_{t\ge 0}\left|\mathbb{P}\Big(\max_{1\le j\le p_1}|G_j| \le t\Big) - \mathbb{P}\Big(\max_{1\le j\le p_1}|\widehat{B}_j| \le t\,\big|\boldsymbol{X}_1,\dots,\boldsymbol{X}_n\Big)\right| \ge C\frac{\gamma_n^{5/6}(s,d)}{n^{1/6}}\right\}$$
$$= O\big(p_1^{-M}\big).$$

This proves (6.10).

Finally, using Theorem 2.1, we deduce that for any $M > 0$, there exists a constant $C_M > 0$ depending only on M and K_1 such that

$$\mathbb{P}\left[\sup_{t\geq 0}\left|\mathbb{P}(\widehat{Q}_{\max} \leq t) - \mathbb{P}(\widehat{B}_{\max} \leq t|\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)\right| \geq C_M\left\{\frac{\gamma_n^{9/8}(s,d)}{n^{1/8}} + \frac{\gamma_n^{5/6}(s,d)}{n^{1/6}}\right\}\right] \lesssim p_1^{-M}.$$

This completes the proof. □

6.1.4. *Proof of Theorems 2.3 and 3.1*

Theorems 2.3 and 3.1 can be proved based on similar arguments used in the proofs of Theorems 2.1 and Theorem 2.2. The details are hence omitted.

6.1.5. *Proof of Theorem 4.1*

Proof. To begin with, we introduce the following notations. Define

$$\mathcal{Q}_{\boldsymbol{v}} = \sqrt{\frac{nm}{n+m}}\frac{\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}}_2)\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}}, \qquad \mathcal{Q}_{\max} = \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}|\mathcal{Q}_{\boldsymbol{v}}|, \quad \text{and} \quad \mathcal{M}_{\max} = \max_{\boldsymbol{v}\in\mathbb{N}^0_{\varepsilon_3}}|\mathcal{Q}_{\boldsymbol{v}}|,$$

where $\varepsilon_3 := \{m\gamma_s(\boldsymbol{\Sigma}_2)\}^{-1}$.

We divide the proof into three main steps. (i) First, using the discretized version $\mathcal{M}_{\max}$ as a bridge, we show that $\mathcal{Q}_{\max}$ converges weakly to the maximum of a Gaussian sequence. (ii) Next we show that the difference between $\mathcal{Q}_{\max}$ and the test statistic $\widehat{\mathcal{Q}}_{\max}$ is negligible asymptotically. (iii) Finally, we show that the Gaussian maximum can be approximated by its multiplier bootstrap counterpart. The technical details are stated as lemmas with their proofs deferred to Section 6.2.

Lemma 6.5. *Let Assumptions 4.1 and 4.2 be satisfied. Under the null hypothesis* $\mathbf{H_0} : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, *we have the following two assertions hold.*

(i) *We have*

$$\mathbb{P}\left[\mathcal{Q}_{\max} \leq C_{31}L_2\left\{\gamma_m^{1/2}(s,d) + (\log m)^{1/2} + \gamma_m(s,d)\frac{\log m}{\sqrt{m}}\right\}\right] \geq 1 - \frac{4}{n} - \frac{4}{m}, \tag{6.17}$$

where $C_{31} > 0$ *is an absolute constant,* $\gamma_m(s,d) := s\log\{\gamma_s(\boldsymbol{\Sigma}_2)\frac{ed}{s}\} \vee s\log m$, *and* $L_2 := L_1^2 + 1$.

(ii) *Let* $\mathbb{N}^0_{\varepsilon_3} = \{\boldsymbol{u}_j\}_{j=1}^{p_3}$ *be an* ε_3*-net with* $\varepsilon_3 = \{m\gamma_s(\boldsymbol{\Sigma}_2)\}^{-1}$ *and* $p_3 = |\mathbb{N}^0_{\varepsilon_3}|$. *Then, there exists a* p_3*-dimensional Gaussian random vector* $(\mathcal{G}_1, \ldots, \mathcal{G}_{p_3})^{\mathsf{T}}$ *satisfying*

$$\mathbb{E}(\mathcal{G}_j\mathcal{G}_k) = \frac{n}{m(n+m)}\sum_{i=1}^{m}\mathbb{E}(\mathcal{R}_{ij}\mathcal{R}_{ik}), \qquad 1 \leq j \leq k \leq p_3, \tag{6.18}$$

with

$$\mathcal{R}_{ij} := \begin{cases} \dfrac{m}{n}\dfrac{\boldsymbol{u}_j^\top(\boldsymbol{X}_i\boldsymbol{X}_i^\top - \boldsymbol{\Sigma}_2)\boldsymbol{u}_j}{\boldsymbol{u}_j^\top\boldsymbol{\Sigma}_2\boldsymbol{u}_j} - \dfrac{\boldsymbol{u}_j^\top(\boldsymbol{Y}_i\boldsymbol{Y}_i^\top - \boldsymbol{\Sigma}_2)\boldsymbol{u}_j}{\boldsymbol{u}_j^\top\boldsymbol{\Sigma}_2\boldsymbol{u}_j}, & \text{if } 1 \le i \le n, \\ -\dfrac{\boldsymbol{u}_j^\top(\boldsymbol{Y}_i\boldsymbol{Y}_i^\top - \boldsymbol{\Sigma}_2)\boldsymbol{u}_j}{\boldsymbol{u}_j^\top\boldsymbol{\Sigma}_2\boldsymbol{u}_j}, & \text{if } n+1 \le i \le m \end{cases} \tag{6.19}$$

(*here, without loss of generality, we assume* $n \le m$) *such that*

$$\mathbb{P}\left[\left|\mathcal{Q}_{\max} - \max_{1\le j\le p_3}|\mathcal{G}_j|\right| \ge C_{32}\frac{K_0^4\gamma_m^{5/8}(s,d)}{L_2 m^{1/8}}\right] \lesssim L_2^2\frac{\gamma_m^{1/8}(s,d)}{m^{1/8}} + L_2^2\frac{\gamma_m^{9/2}(s,d)}{m^{1/2}}, \tag{6.20}$$

where $C_{32} > 0$ *is an absolute constant and* K_0 *is the constant in Lemma 6.13 by taking* $\alpha = 1$.

Lemma 6.6. *Let Assumptions 4.1 and 4.2 be satisfied. Under the null hypothesis* $\mathbf{H_0} : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, *we have, as* $n, m \to \infty$,

$$\begin{aligned} &\mathbb{P}\left[|\widehat{\mathcal{Q}}_{\max} - \mathcal{Q}_{\max}| \le C_{33}L_2\sqrt{\frac{nm}{n+m}}\left\{\frac{\gamma_m^{1/2}(s,d)}{\sqrt{m}} + \gamma_m(s,d)\frac{\log m}{m}\right\}^2\right] \\ &\quad \ge 1 - \frac{4}{n} - \frac{4}{m}, \end{aligned} \tag{6.21}$$

where $C_{33} > 0$ *is an absolute constant.*

Lemma 6.7. *Let Assumptions 4.1 and 4.2 be satisfied. Under the null hypothesis* $\mathbf{H_0} : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, *we have, as* $n, m \to \infty$,

$$\sup_{t\ge 0}\left|\mathbb{P}\Big(\max_{1\le j\le p_3}|\mathcal{G}_j| \le t\Big) - \mathbb{P}(\widehat{\mathcal{B}}_{\max} \le t|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n,\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m)\right| = o_{\mathbb{P}}(1).$$

Combining (6.20) and (6.21) we deduce that there exists an absolute constant $C > 0$ such that

$$\mathbb{P}\left[\left|\widehat{\mathcal{Q}}_{\max} - \max_{1\le j\le p_3}|\mathcal{G}_j|\right| \ge C\frac{K_0^4\gamma_m^{5/8}(s,d)}{L_2 m^{1/8}}\right] \lesssim L_2^2\frac{\gamma_m^{1/8}(s,d)}{m^{1/8}} + L_2^2\frac{\gamma_m^{9/2}(s,d)}{m^{1/2}}.$$

Using arguments similar to those used in the proof of Theorem 2.1, we deduce that

$$\sup_{t\ge 0}\left|\mathbb{P}(\widehat{\mathcal{Q}}_{\max} \le t) - \mathbb{P}\Big(\max_{1\le j\le p_3}|\mathcal{G}_j| \le t\Big)\right| \lesssim L_2^2\frac{\gamma_m^{9/8}(s,d)}{m^{1/8}} + L_2^2\frac{\gamma_m^{9/2}(s,d)}{m^{1/2}}.$$

This, together with Lemma 6.7 yields that

$$\sup_{t\ge 0}\left|\mathbb{P}(\widehat{\mathcal{Q}}_{\max} \le t) - \mathbb{P}(\widehat{\mathcal{B}}_{\max} \le t|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n,\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_m)\right| = o_{\mathbb{P}}(1),$$

which completes the proof. □

6.1.6. *Proof of Theorem* 4.2

Proof. It is equivalent to proving that for $\lambda > 0$ sufficiently large,

$$\inf_{(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \in \mathbb{M}(\lambda)} \mathbb{P}_{(\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)}(\widehat{\mathcal{Q}}_{\max} \geq q_\alpha) = 1 - o(1). \tag{6.22}$$

First, we claim that $q_\alpha = O_{\mathbb{P}}\{\sqrt{s \log(ed/s)}\}$. To see this, it suffices to show that

$$\widehat{\mathcal{B}}_{\max} = O_{\mathbb{P}}\big\{\sqrt{s \log(ed/s)}\big\}.$$

It suffices to show

$$\sup_{\boldsymbol{v} \in \mathbb{V}(s,d)} \left| \sqrt{\frac{n+m}{nm}} \frac{\boldsymbol{v}^\top \{\sum_{i=1}^n \xi_i (\boldsymbol{X}_i \boldsymbol{X}_i^\top - \widehat{\boldsymbol{\Sigma}}_1)/n - \sum_{i=1}^m \eta_i (\boldsymbol{Y}_i \boldsymbol{Y}_i^\top - \widehat{\boldsymbol{\Sigma}}_2)/m\} \boldsymbol{v}}{\boldsymbol{v}^\top (\boldsymbol{\Sigma}_1/n + \boldsymbol{\Sigma}_2/m) \boldsymbol{v}} \right|$$
$$= O_{\mathbb{P}}\big\{\sqrt{s \log(ed/s)}\big\},$$

since, by exactly the same argument as in the proof of Lemma 6.6, the difference between

$$\sup_{\boldsymbol{v} \in \mathbb{V}(s,d)} \left| \sqrt{\frac{n+m}{nm}} \frac{\boldsymbol{v}^\top \{\sum_{i=1}^n \xi_i (\boldsymbol{X}_i \boldsymbol{X}_i^\top - \widehat{\boldsymbol{\Sigma}}_1)/n - \sum_{i=1}^m \eta_i (\boldsymbol{Y}_i \boldsymbol{Y}_i^\top - \widehat{\boldsymbol{\Sigma}}_2)/m\} \boldsymbol{v}}{\boldsymbol{v}^\top (\boldsymbol{\Sigma}_1/n + \boldsymbol{\Sigma}_2/m) \boldsymbol{v}} \right|$$

and

$$\sup_{\boldsymbol{v} \in \mathbb{V}(s,d)} \left| \sqrt{\frac{n+m}{nm}} \frac{\boldsymbol{v}^\top \{\sum_{i=1}^n \xi_i (\boldsymbol{X}_i \boldsymbol{X}_i^\top - \widehat{\boldsymbol{\Sigma}}_1)/n - \sum_{i=1}^m \eta_i (\boldsymbol{Y}_i \boldsymbol{Y}_i^\top - \widehat{\boldsymbol{\Sigma}}_2)/m\} \boldsymbol{v}}{\boldsymbol{v}^\top (\widehat{\boldsymbol{\Sigma}}_1/n + \widehat{\boldsymbol{\Sigma}}_2/m) \boldsymbol{v}} \right|$$

is of order $O_{\mathbb{P}}\{\sqrt{s \log(ed/s)}\}$. It then reduces to show

$$\sup_{\boldsymbol{v} \in \mathbb{V}(s,d)} \left| \boldsymbol{v}^\top \left\{ \frac{1}{n} \sum_{i=1}^n \xi_i \big(\boldsymbol{X}_i \boldsymbol{X}_i^\top - \widehat{\boldsymbol{\Sigma}}_1\big) - \frac{1}{m} \sum_{i=1}^m \eta_i \big(\boldsymbol{Y}_i \boldsymbol{Y}_i^\top - \widehat{\boldsymbol{\Sigma}}_2\big) \right\} \boldsymbol{v} \right|$$
$$= O_{\mathbb{P}}\left\{ \sqrt{\frac{s \log(ed/s)}{m}} \right\},$$

since we have, for any $\boldsymbol{v} \in \mathbb{V}(s,d)$,

$$\left| \frac{\boldsymbol{v}^\top \{\sum_{i=1}^n \xi_i (\boldsymbol{X}_i \boldsymbol{X}_i^\top - \widehat{\boldsymbol{\Sigma}}_1)/n - \sum_{i=1}^m \eta_i (\boldsymbol{Y}_i \boldsymbol{Y}_i^\top - \widehat{\boldsymbol{\Sigma}}_2)/m\} \boldsymbol{v}}{\boldsymbol{v}^\top (\boldsymbol{\Sigma}_1/n + \boldsymbol{\Sigma}_2/m) \boldsymbol{v}} \right|$$
$$\lesssim n \cdot \left| \boldsymbol{v}^\top \left\{ \frac{1}{n} \sum_{i=1}^n \xi_i \big(\boldsymbol{X}_i \boldsymbol{X}_i^\top - \widehat{\boldsymbol{\Sigma}}_1\big) - \frac{1}{m} \sum_{i=1}^m \eta_i \big(\boldsymbol{Y}_i \boldsymbol{Y}_i^\top - \widehat{\boldsymbol{\Sigma}}_2\big) \right\} \boldsymbol{v} \right|.$$

This is due to the fact that $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2 \in \mathbb{M}(\lambda)$ and $m \asymp n$. Then, we can further write

$$\begin{aligned}
&\mathbb{P}\left[\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\boldsymbol{v}^{\mathsf{T}}\left\{\frac{1}{n}\sum_{i=1}^{n}\xi_i\big(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}-\widehat{\mathbf{\Sigma}}_1\big)-\frac{1}{m}\sum_{i=1}^{m}\eta_i\big(\boldsymbol{Y}_i\boldsymbol{Y}_i^{\mathsf{T}}-\widehat{\mathbf{\Sigma}}_2\big)\right\}\boldsymbol{v}\right|\geq C_{41}\sqrt{\frac{s\log(ed/s)}{m}}\right]\\
&\quad\leq\mathbb{P}\left\{H_1\geq\frac{C_{41}}{4}\sqrt{\frac{s\log(ed/s)}{m}}\right\}+\mathbb{P}\left\{H_2\geq\frac{C_{41}}{4}\sqrt{\frac{s\log(ed/s)}{m}}\right\}\\
&\qquad+\mathbb{P}\left\{H_3\geq\frac{C_{41}}{4}\sqrt{\frac{s\log(ed/s)}{m}}\right\}+\mathbb{P}\left\{H_4\geq\frac{C_{41}}{4}\sqrt{\frac{s\log(ed/s)}{m}}\right\},
\end{aligned}$$

where

$$\begin{aligned}
H_1&:=\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i\boldsymbol{v}^{\mathsf{T}}\big(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}-\mathbf{\Sigma}_1\big)\boldsymbol{v}\right|,\\
H_2&:=\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{1}{m}\sum_{i=1}^{m}\eta_i\boldsymbol{v}^{\mathsf{T}}\big(\boldsymbol{Y}_i\boldsymbol{Y}_i^{\mathsf{T}}-\mathbf{\Sigma}_2\big)\boldsymbol{v}\right|,\\
H_3&:=\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i\boldsymbol{v}^{\mathsf{T}}(\widehat{\mathbf{\Sigma}}_1-\mathbf{\Sigma}_1)\boldsymbol{v}\right|,\quad\text{and}\\
H_4&:=\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{1}{m}\sum_{i=1}^{m}\eta_i\boldsymbol{v}^{\mathsf{T}}(\widehat{\mathbf{\Sigma}}_2-\mathbf{\Sigma}_2)\boldsymbol{v}\right|.
\end{aligned}$$

We bound H_1, H_2, H_3, and H_4, respectively. Without loss of generality, we only need to consider H_1 and H_3. For H_1, define $\overline{\xi}_i=\xi_i I(|\xi_i|\leq\tau\sqrt{\log n})$ for some sufficiently large $\tau>0$. Using the standard ε-net argument, it can be shown that (using Lemma 5.4 in Vershynin [43])

$$\begin{aligned}
&\mathbb{P}\left\{H_1\geq\frac{C_{41}}{4}\sqrt{\frac{s\log(ed/s)}{m}}\right\}\\
&\quad\leq\binom{d}{s}9^s\cdot\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i\boldsymbol{v}^{\mathsf{T}}\big(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}-\mathbf{\Sigma}_1\big)\boldsymbol{v}\right|\geq\frac{C_{41}}{8}\sqrt{\frac{s\log(ed/s)}{m}}\right\}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i\boldsymbol{v}^{\mathsf{T}}\big(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}-\mathbf{\Sigma}_1\big)\boldsymbol{v}\right|\geq\frac{C_{41}}{8}\sqrt{\frac{s\log(ed/s)}{m}}\right\}\\
&\quad\leq n\max_{1\leq i\leq n}\mathbb{P}\big(|\xi_i|>\tau\sqrt{\log n}\big)+\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\overline{\xi}_i\boldsymbol{v}^{\mathsf{T}}\big(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}-\mathbf{\Sigma}_1\big)\boldsymbol{v}\right|\geq\frac{C_{41}}{8}\sqrt{\frac{s\log(ed/s)}{m}}\right\}.
\end{aligned}$$

Similar to Lemma 6.4, define $V_i := \overline{\xi}_i \boldsymbol{v}^\top (\boldsymbol{X}_i \boldsymbol{X}_i^\top - \boldsymbol{\Sigma}_1)\boldsymbol{v}$ and, by Markov's inequality, we have for any $t > 0$,

$$\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^n \overline{\xi}_i \boldsymbol{v}^\top\big(\boldsymbol{X}_i \boldsymbol{X}_i^\top - \boldsymbol{\Sigma}_1\big)\boldsymbol{v} \geq \frac{C_{41}}{8}\sqrt{\frac{s\log(ed/s)}{m}}\right\}$$
$$\leq \exp\left\{-\frac{C_{41}}{8} t\sqrt{ms\log(ed/s)}\right\}\prod_{i=1}^n \mathbb{E}\exp(tV_i).$$

Taking $t = \sqrt{s\log(ed/s)/m}$, it follows

$$\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^n \overline{\xi}_i \boldsymbol{v}^\top\big(\boldsymbol{X}_i \boldsymbol{X}_i^\top - \boldsymbol{\Sigma}_1\big)\boldsymbol{v} \geq \frac{C_{41}}{8}\sqrt{\frac{s\log(ed/s)}{m}}\right\}$$
$$\leq \exp\left(-\frac{C_{41}}{8} s\log(ed/s) + \frac{s\log(ed/s)}{m}\sum_{i=1}^n \mathbb{E}\left[V_i^2 \exp\left\{\sqrt{\frac{s\log(ed/s)}{m}}|V_i|\right\}\right]\right).$$

Similar to (6.45), we get $H_1 = O_{\mathbb{P}}\{\sqrt{s\log(ed/s)/m}\}$ as long as $s\log(ed/s)\log n = o(n)$. Furthermore, using the fact

$$H_3 \leq \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \big|\boldsymbol{v}^\top(\widehat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1)\boldsymbol{v}\big| \cdot \frac{1}{n}\sum_{i=1}^n |\xi_i| = O_{\mathbb{P}}\big\{\sqrt{s\log(ed/s)/m}\big\},$$

we deduce that $H_3 = O_{\mathbb{P}}\{\sqrt{s\log(ed/s)/m}\}$. Putting together the pieces, we conclude that $q_\alpha = O_{\mathbb{P}}\{\sqrt{s\log(ed/s)}\}$.

Secondly, we study $\widehat{\mathcal{Q}}_{\max}$. As in Lemma 6.6, we bound $\mathcal{Q}'_{\max}$ instead, where

$$\mathcal{Q}'_{\max} := \sqrt{\frac{n+m}{nm}}\frac{\boldsymbol{v}^\top(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}}_2)\boldsymbol{v}}{\boldsymbol{v}^\top(\boldsymbol{\Sigma}_1/n + \boldsymbol{\Sigma}_2/m)\boldsymbol{v}}.$$

This is, again, because the difference between them is of order $O_{\mathbb{P}}\{\sqrt{s\log(ed/s)}\}$. Note that

$$\mathcal{Q}'_{\max} \geq \sqrt{\frac{n+m}{nm}}\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{\boldsymbol{v}^\top(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)\boldsymbol{v}}{\boldsymbol{v}^\top(\boldsymbol{\Sigma}_1/n + \boldsymbol{\Sigma}_2/m)\boldsymbol{v}}\right| - (H_5 + H_6),$$

where

$$H_5 := \sqrt{\frac{n+m}{nm}}\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{\boldsymbol{v}^\top(\widehat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1)\boldsymbol{v}}{\boldsymbol{v}^\top(\boldsymbol{\Sigma}_1/n + \boldsymbol{\Sigma}_2/m)\boldsymbol{v}}\right|,$$
$$H_6 := \sqrt{\frac{n+m}{nm}}\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{\boldsymbol{v}^\top(\widehat{\boldsymbol{\Sigma}}_2 - \boldsymbol{\Sigma}_2)\boldsymbol{v}}{\boldsymbol{v}^\top(\boldsymbol{\Sigma}_1/n + \boldsymbol{\Sigma}_2/m)\boldsymbol{v}}\right|.$$

Equation (6.22) then follows from the fact that $H_5 + H_6 = O_{\mathbb{P}}\{\sqrt{s\log(ed/s)}\}$. This completes the proof. □

6.1.7. *Proof of Theorem 4.3*

Proof. Define the class of rank one perturbations of the identity matrix as follows:

$$\mathbb{H}(\lambda) := \left\{\mathbf{M} = \mathbf{I}_d + \lambda\sqrt{\frac{s\log(ed/s)}{n}}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}} : \boldsymbol{v} \in \mathbb{V}(s,d)\right\}.$$

Then, it suffices to prove the conclusion with $\mathbb{M}(\lambda)$ replaced by all $\boldsymbol{\Sigma}_1 \in \mathbb{H}(\lambda)$ and $\boldsymbol{\Sigma}_2 = \mathbf{I}_d$. Let λ be sufficiently small. For any two distributions F and G, we write $F \otimes G$ to represent the product measure of F and G. In particular, we use $F^{\otimes n}$ to denote the product distribution of n independent copies of F. Recall that the minimax risk is lower bounded by the Bayesian risk. Define $\mathbb{P}^0_{\mu_\lambda} = \mathbb{E}\{N_d(\mathbf{0}, \mathbf{I}_d + \lambda\sqrt{s\log(ed/s)/n}\boldsymbol{v}\boldsymbol{v}^{\mathsf{T}})^{\otimes n} \otimes N_d(\mathbf{0}, \mathbf{I}_d)^{\otimes m}\}$ to be the mixture alternative distribution with a prior distribution on $\boldsymbol{v}$ with $\boldsymbol{v}$ taking values uniformly in $\mathbb{V}(s,d)$:

$$\mathbb{P}^0_{\mu_\lambda}(A) := \int_A \mathrm{d}N_d\left(\mathbf{0}, \mathbf{I}_d + \lambda\sqrt{\frac{s\log(ed/s)}{n}}\boldsymbol{s}\boldsymbol{s}^{\mathsf{T}}\right)^{\otimes n} \cdot \mathrm{d}N_d(\mathbf{0}, \mathbf{I}_d)^{\otimes m}\,\mathrm{d}\mathcal{S}(\boldsymbol{s}),$$

where $A \in \mathbb{R}^{(n+m)d}$ and $\mathcal{S}$ denotes the uniform measure on $\mathbb{V}(s,d)$ with respect to the Haar measure.

Define $\mathbb{P}^0_{(\boldsymbol{\Sigma}_1, \mathbf{I}_d)}$ to be the probability measure of $N_d(\mathbf{0}, \boldsymbol{\Sigma}_1)^{\otimes n} \otimes N_d(\mathbf{0}, \mathbf{I}_d)^{\otimes m}$. In particular, let $\mathbb{P}^0_{(\mathbf{I}_d, \mathbf{I}_d)}$ be the probability measure of $N_d(\mathbf{0}, \mathbf{I}_d)^{\otimes(n+m)}$. Note that, for any measurable set $A \subset \mathbb{R}^{(n+m)d}$, the measure $\mathbb{P}^0_{\mu_\lambda}$ satisfies

$$\sup_{\boldsymbol{\Sigma}_1 \in \mathbb{H}(\lambda)} \mathbb{P}^0_{(\boldsymbol{\Sigma}_1, \mathbf{I}_d)}\big(A^c\big) \geq \mathbb{P}^0_{\mu_\lambda}\big(A^c\big).$$

Also by the definition of the probability measure, we have

$$1 = \mathbb{P}^0_{\mu_\lambda}(A) + \mathbb{P}^0_{\mu_\lambda}\big(A^c\big).$$

Due to the triangular inequality, we have

$$\mathbb{P}^0_{\mu_\lambda}(A) \leq \mathbb{P}^0_{(\mathbf{I}_d, \mathbf{I}_d)}(A) + \big|\mathbb{P}^0_{\mu_\lambda}(A) - \mathbb{P}^0_{(\mathbf{I}_d, \mathbf{I}_d)}(A)\big|.$$

Putting $A = \{\Phi_\alpha = 1\}$, we deduce that

$$\begin{aligned}\inf_{\Phi_\alpha} \sup_{\boldsymbol{\Sigma}_1 \in \mathbb{H}(\lambda)} \mathbb{P}^0_{(\boldsymbol{\Sigma}_1, \mathbf{I}_d)}(\Phi_\alpha = 0) &\geq 1 - \alpha - \sup_{A: \mathbb{P}^0_{(\mathbf{I}_d, \mathbf{I}_d)}(A) \leq \alpha} \big|\mathbb{P}^0_{\mu_\lambda}(A) - \mathbb{P}^0_{(\mathbf{I}_d, \mathbf{I}_d)}(A)\big| \\ &\geq 1 - \alpha - \frac{1}{2}\big\|\mathbb{P}^0_{\mu_\lambda} - \mathbb{P}^0_{(\mathbf{I}_d, \mathbf{I}_d)}\big\|_{\mathrm{TV}},\end{aligned} \tag{6.23}$$

where $\|\mathbb{P}^0_{\mu_\lambda} - \mathbb{P}^0_{(\mathbf{I}_d, \mathbf{I}_d)}\|_{\mathrm{TV}}$ denotes the total variation distance between the two probability measures $\mathbb{P}^0_{\mu_\lambda}$ and $\mathbb{P}^0_{(\mathbf{I}_d, \mathbf{I}_d)}$.

To finish the proof, we introduce another distance measurement over distributions. Let the χ^2-divergence between two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ be defined as

$$\chi^2(\mathbb{P}_1 \parallel \mathbb{P}_2) = \int \left(\frac{d\mathbb{P}_1}{d\mathbb{P}_2} - 1\right)^2 \mathrm{d}\mathbb{P}_2.$$

In view of the proof of Proposition 2 in Cai, Ma and Wu [9], there exists a function $g : (0, 1/36) \mapsto (1, \infty)$ with $g(0+) = 1$ such that

$$\chi^2\big(\mathbb{P}^0_{\mu_\lambda} \parallel \mathbb{P}^0_{(\mathbf{I}_d,\mathbf{I}_d)}\big) \le g(\beta_0) - 1,$$

where β_0 tends to zero as $\lambda \to 0$. Using the Pinsker's inequality (see, for example, Lemma 2.5 in Tsybakov [40])

$$\chi^2\big(\mathbb{P}^0_{\mu_\lambda} \parallel \mathbb{P}^0_{(\mathbf{I}_d,\mathbf{I}_d)}\big) \ge 2\big\|\mathbb{P}^0_{\mu_\lambda} - \mathbb{P}^0_{(\mathbf{I}_d,\mathbf{I}_d)}\big\|^2_{\mathrm{TV}},$$

we deduce from (6.23) that

$$\inf_{\Phi_\alpha} \sup_{\mathbf{\Sigma}_1 \in \mathbb{H}(\lambda)} \mathbb{P}^0_{(\mathbf{\Sigma}_1,\mathbf{I}_d)}(\Phi_\alpha = 0) \ge 1 - \alpha - o(1).$$

This completes the proof. □

6.2. Proofs of the supporting lemmas

6.2.1. *Proof of Lemma 6.1*

Proof. For any $\varepsilon \in (0, 1)$ fixed and $\mathbb{I} \subseteq [d]$ subject to $|\mathbb{I}| = s$, let $\mathbb{S}^{s-1}_{\mathbb{I}} \subseteq \mathbb{R}^d$ be the unit Euclidean sphere whose support is $\mathbb{I}$. Further, let $\mathbb{N}^0_{\mathbb{I},\varepsilon}$ denote an ε-net of $\mathbb{S}^{s-1}_{\mathbb{I}}$ with respect to the Euclidean metric ρ_E satisfying $|\mathbb{N}^0_{\mathbb{I},\varepsilon}| \le (1 + 2/\varepsilon)^s$. Due to the decomposition

$$\mathbb{V}(s, d) = \big\{\boldsymbol{v} \in \mathbb{S}^{d-1} : |\boldsymbol{v}|_0 = s\big\} = \bigcup_{\mathbb{I}\subseteq[d]:|\mathbb{I}|=s} \big\{\boldsymbol{v} \in \mathbb{S}^{d-1} : \mathrm{supp}(\boldsymbol{v}) = \mathbb{I}\big\} = \bigcup_{\mathbb{I}\subseteq[d]:|\mathbb{I}|=s} \mathbb{S}^{s-1}_{\mathbb{I}},$$

we can construct an ε-net of $(\mathbb{V}(s, d), \rho_E)$ by $\mathbb{N}^0_\varepsilon := \bigcup_{\mathbb{I}\subseteq[d]:|\mathbb{I}|=s} \mathbb{N}^0_{\mathbb{I},\varepsilon}$. Then, it is straightforward to see that

$$p^0_\varepsilon = \big|\mathbb{N}^0_\varepsilon\big| \le \binom{d}{s}\left(1 + \frac{2}{\varepsilon}\right)^s.$$

Using the binomial coefficient bound

$$\binom{d}{s} \le \left(\frac{ed}{s}\right)^s,$$

we get

$$\log p^0_\varepsilon \lesssim s \log \frac{ed}{\varepsilon s}.$$

Next, we prove the second assertion. For every $\boldsymbol{v} \in \mathbb{V}(s,d)$ with support $\mathbb{I}$ and its ε-net $\mathbb{N}_{\mathbb{I},\varepsilon}$, we can find some $\widetilde{\boldsymbol{v}} \in \mathbb{N}_{\mathbb{I},\varepsilon}$ satisfying that $\operatorname{supp}(\boldsymbol{v}) = \operatorname{supp}(\widetilde{\boldsymbol{v}})$ and $\|\boldsymbol{v} - \widetilde{\boldsymbol{v}}\|_2 \le \varepsilon$. By Lemma 2.2, we have

$$\begin{aligned}\left|\left|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}_{\boldsymbol{\Sigma}}\right| - \left|\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\right|\right| &\le 2\gamma_s \|\boldsymbol{v} - \widetilde{\boldsymbol{v}}\|_2 \cdot \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}_{\boldsymbol{\Sigma}}\right| \\ &\le 2\gamma_s \varepsilon \cdot \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}_{\boldsymbol{\Sigma}}\right|.\end{aligned}$$

Therefore, we have

$$\sup_{\boldsymbol{v}\in\mathbb{V}(s,d):\operatorname{supp}(\boldsymbol{v})=\mathbb{I}} \left|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}_{\boldsymbol{\Sigma}}\right| \le 2\gamma_s \varepsilon \cdot \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}_{\boldsymbol{\Sigma}}\right| + \max_{\boldsymbol{v}\in\mathbb{N}_{\mathbb{I},\varepsilon}} \left|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}_{\boldsymbol{\Sigma}}\right|.$$

Taking maximum over $\mathbb{I} \subseteq [d]$ with $|\mathbb{I}| = s$ on both sides yields

$$\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}_{\boldsymbol{\Sigma}}\right| \le 2\gamma_s \varepsilon \cdot \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}_{\boldsymbol{\Sigma}}\right| + \max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon}} \left|\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma})\boldsymbol{v}_{\boldsymbol{\Sigma}}\right|.$$

Together, the last two displays imply $\widehat{Q}_{\max} \le 2\gamma_s \varepsilon \cdot \widehat{Q}_{\max} + M_{\max,\varepsilon}$. This completes the proof. □

6.2.2. *Proof of Lemma 6.2*

Proof. We follow a standard procedure. First, we show concentration of $\widehat{Q}_{\max}$ around its expectation $\mathbb{E}\widehat{Q}_{\max}$. Next we upper bound $\mathbb{E}\widehat{Q}_{\max}$. To prove the concentration, we define for every $\boldsymbol{v} \in \mathbb{V}(s,d)$ that

$$g_{\boldsymbol{v}}(\boldsymbol{X}_i) = \frac{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v}} - 1.$$

By Lemma 6.12, there exists an absolute constant $C_{11} > 0$ such that for every $t > 0$,

$$\begin{aligned}&\mathbb{P}\left[\widehat{Q}_{\max} \le 2\mathbb{E}\widehat{Q}_{\max} + \max\left\{\underbrace{2\sigma_{\boldsymbol{v}}\sqrt{\frac{t}{n}}}_{J_1}, \underbrace{C_{11}\frac{t}{\sqrt{n}}\left\|\max_{1\le i\le n}\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|g_{\boldsymbol{v}}(\boldsymbol{X}_i)\right|\right\|_{\psi_1}}_{J_2}\right\}\right] \\ &\quad \ge 1 - 4e^{-t},\end{aligned} \tag{6.24}$$

where $\sigma_{\boldsymbol{v}}^2 := \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \sum_{i=1}^n \mathbb{E}g_{\boldsymbol{v}}^2(\boldsymbol{X}_i)$. We first bound J_1 and J_2, starting with J_1. Under Assumption 2.1, we have

$$\sigma_{\boldsymbol{v}}^2 \le n \cdot \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \mathbb{E}\left\{\left(\frac{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}}\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v}} - 1\right)^2\right\} \le 2n \sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}} \left\|\boldsymbol{v}^{\mathsf{T}}\boldsymbol{U}_i\boldsymbol{U}_i^{\mathsf{T}}\boldsymbol{v} - 1\right\|_{\psi_1}^2 \le 2K_2^2 n, \tag{6.25}$$

and hence $\sigma_{\boldsymbol{v}} \le K_2\sqrt{2n}$. For J_2, using a similar argument as in the proof of Lemma 6.1, we deduce that for every $0 < \varepsilon < (2\gamma_s)^{-1}$,

$$\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{\boldsymbol{v}^\top \boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{v}}{\boldsymbol{v}^\top\boldsymbol{\Sigma}\boldsymbol{v}}-1\right| \le (1-2\gamma_s\varepsilon)^{-1}\max_{\boldsymbol{v}\in\mathbb{N}_\varepsilon^0}\left|\frac{\boldsymbol{v}^\top \boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{v}}{\boldsymbol{v}^\top\boldsymbol{\Sigma}\boldsymbol{v}}-1\right|.$$

By taking $\varepsilon = \varepsilon_4 := (4\gamma_s)^{-1}$, we have

$$\begin{aligned}\Big\|\max_{1\le i\le n}\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\big|g_{\boldsymbol{v}}(\boldsymbol{X}_i)\big|\Big\|_{\psi_1} &= \Big\|\max_{1\le i\le n}\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{\boldsymbol{v}^\top \boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{v}}{\boldsymbol{v}^\top\boldsymbol{\Sigma}\boldsymbol{v}}-1\right|\Big\|_{\psi_1}\\ &\lesssim \Big\|\max_{1\le i\le n}\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_4}^0}\left|\frac{\boldsymbol{v}^\top \boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{v}}{\boldsymbol{v}^\top\boldsymbol{\Sigma}\boldsymbol{v}}-1\right|\Big\|_{\psi_1},\end{aligned}$$

where $\mathbb{N}_{\varepsilon_4}^0$ is an ε_4-net of $\mathbb{V}(s,d)$ with properties in Lemma 6.1. Using Lemma 6.13, we have

$$\Big\|\max_{1\le i\le n}\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\big|g_{\boldsymbol{v}}(\boldsymbol{X}_i)\big|\Big\|_{\psi_1} \lesssim \left(s\log\frac{ed}{\varepsilon_4 s}+\log n\right)\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left\|\frac{\boldsymbol{v}^\top \boldsymbol{X}_i\boldsymbol{X}_i^\top\boldsymbol{v}}{\boldsymbol{v}^\top\boldsymbol{\Sigma}\boldsymbol{v}}-1\right\|_{\psi_1}.$$

It follows that

$$\Big\|\max_{1\le i\le n}\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\big|g_{\boldsymbol{v}}(\boldsymbol{X}_i)\big|\Big\|_{\psi_1} \lesssim K_2\big\{s\log(\gamma_s\cdot ed/s)+\log n\big\}. \tag{6.26}$$

Combining (6.24), (6.25), and (6.26) gives

$$\mathbb{P}\left[\widehat{Q}_{\max} \le 2\mathbb{E}\widehat{Q}_{\max} + K_2\max\left\{2\sqrt{2t}, C_{11}\gamma_n(s,d)\frac{t}{\sqrt{n}}\right\}\right] \ge 1-4e^{-t}, \tag{6.27}$$

where we recall that $\gamma_n(s,d) = s\log(\gamma_s\cdot ed/s)\vee s\log n$.

Now we bound the expectation $\mathbb{E}\widehat{Q}_{\max}$. Here we use a result that involves the generic chaining complexity, $\gamma_2(T,\rho)$, of a metric space (T,ρ). See Definition 2.2.19 in Talagrand [38]. We refer the readers to Talagrand [38] for a systematic introduction. Note that

$$\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\|\boldsymbol{v}_{\boldsymbol{\Sigma}}^\top\boldsymbol{X}_i\|_{\psi_1} = \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\|(\boldsymbol{\Sigma}^{1/2}\boldsymbol{v}_{\boldsymbol{\Sigma}})^\top\boldsymbol{U}_i\|_{\psi_1} \le K_1,$$

and

$$\|(\boldsymbol{v}_{\boldsymbol{\Sigma}}-\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}})^\top\boldsymbol{X}_i\|_{\psi_2} = \|(\boldsymbol{v}_{\boldsymbol{\Sigma}}-\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}})^\top\boldsymbol{\Sigma}^{1/2}\boldsymbol{U}_i\|_{\psi_2} \le K_1\|\boldsymbol{v}_{\boldsymbol{\Sigma}}-\widetilde{\boldsymbol{v}}_{\boldsymbol{\Sigma}}\|_{\boldsymbol{\Sigma}},$$

for any $\boldsymbol{v},\widetilde{\boldsymbol{v}}\in\mathbb{V}(s,d)$. It follows from Lemma 6.14 and Lemma 6.15 that

$$\begin{aligned}\mathbb{E}\widehat{Q}_{\max} &= \sqrt{n}\mathbb{E}\left\{\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{1}{n}\sum_{i=1}^n(\boldsymbol{v}_{\boldsymbol{\Sigma}}^\top\boldsymbol{X}_i)^2-1\right|\right\}\\ &\lesssim K_1^2\left\{\gamma_2\big(\mathbb{V}(s,d),\|\cdot\|_{\boldsymbol{\Sigma}}\big)+\frac{\gamma_2(\mathbb{V}(s,d),\|\cdot\|_{\boldsymbol{\Sigma}})^2}{\sqrt{n}}\right\}.\end{aligned} \tag{6.28}$$

By Lemma 6.16, we have

$$\gamma_2\big(\mathbb{V}(s,d), \|\cdot\|_{\boldsymbol{\Sigma}}\big) \lesssim \mathbb{E}\Big\{\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \big(\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\boldsymbol{Z}\big)\Big\}, \tag{6.29}$$

where $\boldsymbol{Z} \overset{\mathsf{d}}{=} N_d(\boldsymbol{0}, \boldsymbol{\Sigma})$. Similar to the proof of Lemma 6.1, we have

$$\begin{aligned}\mathbb{E}\Big\{\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \big(\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\boldsymbol{Z}\big)\Big\} &\le \frac{4}{3}\mathbb{E}\Big\{\max_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_4}^0} \big(\boldsymbol{v}_{\boldsymbol{\Sigma}}^{\mathsf{T}}\boldsymbol{Z}\big)\Big\} \\ &\le 2\big(\log\big|\mathbb{N}_{\varepsilon_4}^0\big|\big)^{1/2} \lesssim \gamma_n^{1/2}(s,d),\end{aligned} \tag{6.30}$$

where $\varepsilon_4 = (4\gamma_s)^{-1}$. Together, (6.28), (6.29), and (6.30) imply that

$$\mathbb{E}\widehat{Q}_{\max} \lesssim K_1^2\gamma_n^{1/2}(s,d). \tag{6.31}$$

Combining (6.27) and (6.31), we deduce that

$$\mathbb{P}\bigg[\widehat{Q}_{\max} \le C_{12}K_2\gamma_n^{1/2}(s,d) + K_2\max\bigg\{2\sqrt{2t}, C_{11}\gamma_n(s,d)\frac{t}{\sqrt{n}}\bigg\}\bigg] \ge 1 - 4e^{-t}.$$

This completes the proof. □

6.2.3. *Proof of Lemma 6.3*

Proof. Recall that

$$M_{\max,\varepsilon} = \max_{1\le j\le p_\varepsilon}\bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^n R_{ij}\bigg|,$$

and $\mathbb{E}R_{ij} = 0$ for $i = 1,\dots,n$ and $j = 1,\dots,p_\varepsilon$. Moreover, define $R_{ij} = -R_{i,j-p_\varepsilon}$ for $j = p_\varepsilon + 1,\dots,2p_\varepsilon$ and put $\boldsymbol{R}_i = (R_{i1},\dots,R_{i,2p_\varepsilon})^{\mathsf{T}}$ for $i = 1,\dots,n$. Let $\boldsymbol{G} = (G_1,\dots,G_{p_\varepsilon},-G_1,\dots,-G_{p_\varepsilon})^{\mathsf{T}}$ be a $(2p_\varepsilon)$-dimensional Gaussian random vector satisfying

$$\mathbb{E}(G_jG_k) = \mathbb{E}(R_{ij}R_{ik}), \qquad 1 \le j \le k \le p_\varepsilon.$$

Applying Lemma 2.1 to $\{\boldsymbol{R}_i\}_{i=1}^n$ and $\boldsymbol{G}$, we have, for any $\delta > 0$,

$$\begin{aligned}&\mathbb{P}\bigg(\bigg|\max_{1\le j\le p_\varepsilon}\bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^n R_{ij}\bigg| - \max_{1\le j\le p_\varepsilon}|G_j|\bigg| \ge 16\delta\bigg) \\ &\quad\lesssim D_1\frac{\log(2p_\varepsilon\vee n)}{\delta^2 n} + (D_2+D_3)\frac{\{\log(2p_\varepsilon\vee n)\}^2}{\delta^3 n^{3/2}} + \frac{\log n}{n},\end{aligned} \tag{6.32}$$

where we put

$$D_1 = \mathbb{E}\left[\max_{1\le j,k\le 2p_\varepsilon}\left|\sum_{i=1}^n \{R_{ij}R_{ik} - \mathbb{E}(R_{ij}R_{ik})\}\right|\right],$$

$$D_2 = \mathbb{E}\left(\max_{1\le j\le 2p_\varepsilon}\sum_{i=1}^n |R_{ij}|^3\right),$$

$$D_3 = \sum_{i=1}^n \mathbb{E}\left[\max_{1\le j\le 2p_\varepsilon}|R_{ij}|^3 \mathbb{1}\left\{\max_{1\le j\le 2p_\varepsilon}|R_{ij}| > \frac{\delta n^{1/2}}{\log(2p_\varepsilon\vee n)}\right\}\right].$$

Note that, for $i = 1,\dots,n$,

$$\mathbb{E}\Big(\max_{1\le j\le 2p_\varepsilon} R_{ij}^4\Big) \ge \frac{\delta n^{1/2}}{\log(2p_\varepsilon\vee n)}\mathbb{E}\left[\max_{1\le j\le 2p_\varepsilon}|R_{ij}|^3\mathbb{1}\left\{\max_{1\le j\le 2p_\varepsilon}|R_{ij}| > \frac{\delta n^{1/2}}{\log(2p_\varepsilon\vee n)}\right\}\right],$$

we have

$$D_3 \le \frac{\log(2p_\varepsilon\vee n)}{\delta n^{1/2}}\sum_{i=1}^n \mathbb{E}\Big(\max_{1\le j\le 2p_\varepsilon} R_{ij}^4\Big).$$

Hence, we deduce from (6.32) that

$$\begin{aligned}\mathbb{P}&\left(\left|\max_{1\le j\le p_\varepsilon}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n R_{ij}\right| - \max_{1\le j\le p_\varepsilon}|G_j|\right| \ge 16\delta\right)\\ &\lesssim D_1\frac{\log(2p_\varepsilon\vee n)}{\delta^2 n} + D_2\frac{\{\log(2p_\varepsilon\vee n)\}^2}{\delta^3 n^{3/2}}\\ &\quad + D_4\frac{\{\log(2p_\varepsilon\vee n)\}^3}{\delta^4 n^2} + \frac{\log n}{n},\end{aligned} \tag{6.33}$$

where

$$D_4 := \sum_{i=1}^n \mathbb{E}\Big(\max_{1\le j\le 2p_\varepsilon} R_{ij}^4\Big).$$

Next we bound D_1, D_2, and D_4, starting with D_1. By Lemma 6.17,

$$\begin{aligned}D_1 &\lesssim \sqrt{\log(2p_\varepsilon)}\max_{1\le j\le 2p_\varepsilon}\left(\sum_{i=1}^n \mathbb{E}R_{ij}^4\right)^{1/2} + \log(2p_\varepsilon)\left\{\mathbb{E}\Big(\max_{1\le i\le n}\max_{1\le j\le 2p_\varepsilon} R_{ij}^4\Big)\right\}^{1/2}\\ &= \sqrt{\log(2p_\varepsilon)}\underbrace{\max_{1\le j\le p_\varepsilon}\left(\sum_{i=1}^n \mathbb{E}R_{ij}^4\right)^{1/2}}_{D_{11}} + \log(2p_\varepsilon)\underbrace{\left\{\mathbb{E}\Big(\max_{1\le i\le n}\max_{1\le j\le p_\varepsilon} R_{ij}^4\Big)\right\}^{1/2}}_{D_{12}}.\end{aligned} \tag{6.34}$$

For D_{11}, using Lemma 6.9, we deduce that

$$\mathbb{E}R_{ij}^4 \le 4!\|R_{ij}\|_{\psi_1}^4 = 4!\left\| \frac{\boldsymbol{v}_j^\top \boldsymbol{\Sigma}^{1/2}\boldsymbol{U}_i\boldsymbol{U}_i^\top\boldsymbol{\Sigma}^{1/2}\boldsymbol{v}_j}{\boldsymbol{v}_j^\top\boldsymbol{\Sigma}\boldsymbol{v}_j} - 1\right\|_{\psi_1}^4.$$

This gives $\mathbb{E}R_{ij}^4 \le 4!\sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\|\boldsymbol{v}^\top\boldsymbol{U}_i\boldsymbol{U}_i^\top\boldsymbol{v}-1\|_{\psi_1}^4 = 4!K_2^4$ and hence

$$D_{11} \le \left(4!K_2^4 n\right)^{1/2} = 24^{1/2}K_2^2\sqrt{n}. \tag{6.35}$$

To bound D_{12}, by Lemmas 6.9 and 6.13, we have

$$\begin{aligned}
\mathbb{E}\Big(\max_{1\le i\le n}\max_{1\le j\le p_\varepsilon} R_{ij}^4\Big) &\le 4!\Big\|\max_{1\le i\le n}\max_{1\le j\le p_\varepsilon}|R_{ij}|\Big\|_{\psi_1}^4 \\
&\le 4!K_0^4\big\{\log(np_\varepsilon+1)\big\}^4 \max_{1\le i\le n}\max_{1\le j\le p_\varepsilon}\|R_{ij}\|_{\psi_1}^4 \\
&\le 4!K_0^4\big\{\log(np_\varepsilon+1)\big\}^4 \sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\big\|\boldsymbol{v}^\top\boldsymbol{U}\boldsymbol{U}^\top\boldsymbol{v}-1\big\|_{\psi_1}^4 \\
&= 4!K_0^4K_2^4\big\{\log(np_\varepsilon+1)\big\}^4,
\end{aligned}$$

which further implies

$$D_{12} \le 24^{1/2}K_0^2K_2^2\big\{\log(np_\varepsilon+1)\big\}^2. \tag{6.36}$$

Combining (6.34), (6.35), and (6.36) yields

$$D_1 \lesssim K_2^2(\log 2p_\varepsilon)^{1/2}\sqrt{n} + K_0^2K_2^2\big\{\log(np_\varepsilon+1)\big\}^2\log(2p_\varepsilon). \tag{6.37}$$

For D_2, it follows from Lemma 6.18 that

$$D_2 = \mathbb{E}\left(\max_{1\le j\le p_\varepsilon}\sum_{i=1}^n|R_{ij}|^3\right) \lesssim \max_{1\le j\le p_\varepsilon}\underbrace{\sum_{i=1}^n\mathbb{E}|R_{ij}|^3}_{D_{21}} + (\log p_\varepsilon)\underbrace{\mathbb{E}\Big(\max_{1\le i\le n}\max_{1\le j\le p_\varepsilon}|R_{ij}|^3\Big)}_{D_{22}}. \tag{6.38}$$

By Lemma 6.9, we have $\mathbb{E}|R_{ij}|^3 \le 3!K_2^3$ and hence

$$D_{21} \le 3!K_2^3 n. \tag{6.39}$$

Further, in view of Lemma 6.13, we have

$$D_{22} \le 3!K_0^3\big\{\log(np_\varepsilon+1)\big\}^3 \max_{1\le i\le n}\max_{1\le j\le p_\varepsilon}\|R_{ij}\|_{\psi_1}^3 \le 3!K_0^3K_2^3\big\{\log(np_\varepsilon+1)\big\}^3. \tag{6.40}$$

Together, (6.38), (6.39), and (6.40) yield that

$$D_2 \lesssim K_2^3 n + K_0^3K_2^3\big\{\log(np_\varepsilon+1)\big\}^3\log p_\varepsilon. \tag{6.41}$$

For D_4, using Lemmas 6.9 and 6.13, we deduce that

$$\begin{aligned}\mathbb{E}\Big(\max_{1\le j\le 2p_\varepsilon} R_{ij}^4\Big) &= \mathbb{E}\Big(\max_{1\le j\le p_\varepsilon} R_{ij}^4\Big) \le 4!\Big\|\max_{1\le j\le p_\varepsilon}|R_{ij}|\Big\|_{\psi_1}^4 \\ &\le 4!K_0^4\big\{\log(p_\varepsilon+1)\big\}^4 \sup_{\boldsymbol{v}\in\mathbb{S}^{d-1}}\big\|\boldsymbol{v}^\mathsf{T}\boldsymbol{U}\boldsymbol{U}^\mathsf{T}\boldsymbol{v}-1\big\|_{\psi_1}^4 \\ &= 4!K_0^4K_2^4\big\{\log(p_\varepsilon+1)\big\}^4.\end{aligned}$$

Consequently, we have

$$D_4 \lesssim K_0^4K_2^4\big\{\log(p_\varepsilon+1)\big\}^4 n. \tag{6.42}$$

Finally, putting (6.33), (6.37), (6.41), and (6.42) together, we obtain

$$\begin{aligned}&\mathbb{P}\bigg(\bigg|\max_{1\le j\le p_\varepsilon}\bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^n R_{ij}\bigg| - \max_{1\le j\le p_\varepsilon}|G_j|\bigg| \ge 16\delta\bigg) \\ &\lesssim K_2^2\frac{\sqrt{\log(2p_\varepsilon)}\log(2p_\varepsilon\vee n)}{\delta^2\sqrt{n}} + K_0^2K_2^2\frac{\log(2p_\varepsilon)\{\log(np_\varepsilon+1)\}^2\log(2p_\varepsilon\vee n)}{\delta^2 n} \\ &\quad + K_2^3\frac{\{\log(2p_\varepsilon\vee n)\}^2}{\delta^3\sqrt{n}} + K_0^3K_2^3\frac{(\log p_\varepsilon)\{\log(np_\varepsilon+1)\}^3\{\log(2p_\varepsilon\vee n)\}^2}{\delta^3 n^{3/2}} \\ &\quad + K_0^4K_2^4\frac{\{\log(p_\varepsilon+1)\}^4\{\log(2p_\varepsilon\vee n)\}^3}{\delta^4 n} + \frac{\log n}{n},\end{aligned}$$

as desired. □

6.2.4. *Proof of Lemma 6.4*

Proof. Define $\overline{W}_{ik} = W_{ik}\mathbb{1}(|W_{ik}| \le \tau\{\log(p_1+n)\}^{1/2})$ for some sufficiently large τ. Then, for some constant $C_{22} > 0$,

$$\begin{aligned}&\mathbb{P}\bigg\{\max_{1\le j,k\le p_1}\bigg|\frac{1}{n}\sum_{i=1}^n(W_{ij}W_{ik})^2 - \mathbb{E}(W_{ij}W_{ik})^2\bigg| \ge C_{22}\sqrt{\frac{\log p_1}{n}}\bigg\} \\ &\le np_1\max_{1\le i\le n,1\le k\le p_1}\mathbb{P}\big[|W_{ik}| > \tau\big\{\log(p_1+n)\big\}^{1/2}\big] \\ &\quad + \mathbb{P}\bigg\{\max_{1\le j,k\le p_1}\bigg|\frac{1}{n}\sum_{i=1}^n(W_{ij}\overline{W}_{ik})^2 - \mathbb{E}(W_{ij}W_{ik})^2\bigg| \ge C_{22}\sqrt{\frac{\log p_1}{n}}\bigg\}.\end{aligned}$$

Using Cauchy-Schwarz inequality, we deduce that, for any $\eta > 0$,

$$\begin{aligned}\big|\mathbb{E}(W_{ij}W_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2\big| &\le \big(\mathbb{E}W_{ij}^4\big)^{1/2}\cdot\big[\mathbb{E}\big\{W_{ik}^4\cdot\mathbb{1}\big(|W_{ik}| > \tau\big\{\log(p_1+n)\big\}^{1/2}\big)\big\}\big]^{1/2} \\ &\le \big(\mathbb{E}W_{ij}^4\big)^{1/2}(n+p_1)^{-\tau^2\eta/4}\cdot\big[\mathbb{E}\big\{W_{ik}^4\exp\big(\eta W_{ik}^2/2\big)\big\}\big]^{1/2}.\end{aligned}$$

By the elementary inequality $x^2e^x \le e^{2x}$, $x > 0$, we have, for any $\eta > 0$,

$$\left|\mathbb{E}(W_{ij}W_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2\right| \le \left(\mathbb{E}W_{ij}^4\right)^{1/2}(n+p_1)^{-\tau^2\eta/4}\cdot 2\eta^{-1}\left\{\mathbb{E}\exp\left(\eta W_{ik}^2\right)\right\}^{1/2}.$$

Under Assumptions 2.1 and 2.2, for any $\eta \in (0, K_1^{-2})$, there exists a constant $C_{23} > 0$ such that

$$\left|\mathbb{E}(W_{ij}W_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2\right| \le 2C_{23}\eta^{-1}(n+p_1)^{-\tau^2\eta/4}.$$

Hence, for all sufficiently large τ, n, and p_1, we have

$$2C_{23}\eta^{-1}(n+p_1)^{-\tau^2\eta/4} \le \frac{C_{22}}{2}\sqrt{\frac{\log p_1}{n}}.$$

It follows that

$$\begin{aligned}
&\mathbb{P}\left\{\max_{1\le j,k\le p_1}\left|\frac{1}{n}\sum_{i=1}^n (W_{ij}W_{ik})^2 - \mathbb{E}(W_{ij}W_{ik})^2\right| \ge C_{22}\sqrt{\frac{\log p_1}{n}}\right\} \\
&\quad\le \underbrace{np_1 \max_{1\le i\le n, 1\le k\le p_1}\mathbb{P}\left[|W_{ik}| > \tau\left\{\log(p_1+n)\right\}^{1/2}\right]}_{F_1} \\
&\qquad + \underbrace{\mathbb{P}\left\{\max_{1\le j,k\le p_1}\left|\frac{1}{n}\sum_{i=1}^n (W_{ij}\overline{W}_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2\right| \ge \frac{C_{22}}{2}\sqrt{\frac{\log p_1}{n}}\right\}}_{F_2}.
\end{aligned}$$

For F_1, we have, for any $\eta \in (0, K_1^{-2})$, $M > 0$, and sufficiently large τ,

$$F_1 \le np_1(n+p_1)^{-\tau^2\eta}\max_{1\le i\le n, 1\le k\le p_1}\mathbb{E}\left\{\exp\left(\eta W_{ik}^2\right)\right\} = O\left(p_1^{-M}\right).$$

To bound F_2, it suffices to show that, for any $M > 0$, there exists an absolute constant $C_{24} > 0$ depending only on M such that

$$\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^n (W_{ij}\overline{W}_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2 \ge C_{24}\sqrt{\frac{\log p_1}{n}}\right\} = O\left(p_1^{-M-2}\right). \tag{6.43}$$

Define $W_{ijk} = (W_{ij}\overline{W}_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2$ for $1 \le i \le n$ and $1 \le j, k \le p_1$. By Markov's inequality, we have for any $t > 0$,

$$\begin{aligned}
&\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n\left\{(W_{ij}\overline{W}_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2\right\} \ge C_{25}\sqrt{\frac{\log p_1}{n}}\right] \\
&\quad\le \exp(-C_{25}t\sqrt{n\log p_1})\prod_{i=1}^n \mathbb{E}\exp(tW_{ijk}).
\end{aligned}$$

Using inequalities $e^x \le 1 + x + x^2 e^{|x|}$ and $1 + x \le e^x$ for $x > 0$, we deduce that

$$\begin{aligned}
&\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n}(W_{ij}\overline{W}_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2 \ge C_{25}\sqrt{\frac{\log p_1}{n}}\right\} \\
&\quad \le \exp(-C_{25}t\sqrt{n\log p_1})\prod_{i=1}^{n}\left[1 + \mathbb{E}\left\{t^2 W_{ijk}^2 \exp\left(t|W_{ijk}|\right)\right\}\right] \\
&\quad \le \exp\left[-C_{25}t\sqrt{n\log p_1} + \sum_{i=1}^{n}\mathbb{E}\left\{t^2 W_{ijk}^2 \exp\left(t|W_{ijk}|\right)\right\}\right].
\end{aligned}$$

Taking $t = \eta\tau^{-2}\sqrt{(\log p_1)/n}$ gives

$$\begin{aligned}
&P\left\{\frac{1}{n}\sum_{i=1}^{n}(W_{ij}\overline{W}_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2 \ge C_{25}\sqrt{\frac{\log p_1}{n}}\right\} \\
&\quad \le \exp\left[-C_{25}\frac{\eta}{\tau^2}\log p_1 + \frac{\eta^2\log p_1}{\tau^4 n}\sum_{i=1}^{n}\underbrace{\mathbb{E}\left\{W_{ijk}^2\exp\left(\frac{\eta}{\tau^2}\sqrt{\frac{\log p_1}{n}}|W_{ijk}|\right)\right\}}_{F_3}\right].
\end{aligned} \tag{6.44}$$

Using Cauchy–Schwarz inequality, we have

$$F_3 \le \left\{\mathbb{E}\left(W_{ijk}^4\right)\right\}^{1/2}\left[\mathbb{E}\left\{\exp\left(\frac{2\eta}{\tau^2}\sqrt{\frac{\log p_1}{n}}|W_{ijk}|\right)\right\}\right]^{1/2}.$$

According to Assumption 2.1, for any $\eta \in (0, K_1^{-2})$ and sufficiently large n and p_1 satisfying that $\{\log(p_1 + n)\}^2 \log p_1 = o(n)$, there exists a constant $C_{26} > 0$ depending on K_1, η, and τ such that

$$\mathbb{E}\left\{\exp\left(\frac{2\eta}{\tau^2}\sqrt{\frac{\log p_1}{n}}|W_{ijk}|\right)\right\} \le C_{26}. \tag{6.45}$$

Consequently, there exists a positive constant C_{27} depending on K_1, η, and τ such that

$$F_3 \le C_{27}. \tag{6.46}$$

Combining (6.44) and (6.46), we obtain that for $\eta \in (0, K_1^{-2})$ and sufficiently large τ, n, and p_1,

$$\begin{aligned}
&P\left\{\frac{1}{n}\sum_{i=1}^{n}(W_{ij}\overline{W}_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2 \ge C_{25}\sqrt{\frac{\log p_1}{n}}\right\} \\
&\quad \le \exp\left(-C_{25}\frac{\eta}{\tau^2}\log p_1 + C_{27}\frac{\eta^2}{\tau^4}\log p_1\right)
\end{aligned}$$

for any $C_{25} > 0$. Therefore, for any $M > 0$, there exists a constant $C_{24} > 0$ depending only on M such that (6.43) holds. Similarly, it can be shown that

$$\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n}(W_{ij}\overline{W}_{ik})^2 - \mathbb{E}(W_{ij}\overline{W}_{ik})^2 \le -C_{24}\sqrt{\frac{\log p_1}{n}}\right\} = O\big(p_1^{-M-2}\big).$$

By taking $C_{22} = 2C_{24}$, we get $F_2 = O(p_1^{-M})$, which completes the proof. □

6.2.5. *Proof of Lemma 6.5*

Proof. Similar to Lemma 6.1, we have, for any $\varepsilon \in (0, 1)$,

$$\mathcal{Q}_{\max} \le 2\gamma_s(\boldsymbol{\Sigma}_2)\varepsilon \cdot \mathcal{Q}_{\max} + \mathcal{M}_{\max}. \tag{6.47}$$

Of note, we have

$$\mathcal{Q}_{\max} \le \sqrt{\frac{m}{n+m}}\underbrace{\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{\sqrt{n}\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_2)\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}}\right|}_{\mathcal{Q}_{\max,1}} + \sqrt{\frac{n}{n+m}}\underbrace{\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)}\left|\frac{\sqrt{m}\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_2 - \boldsymbol{\Sigma}_2)\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}}\right|}_{\mathcal{Q}_{\max,2}}.$$

Using Lemma 6.2, we deduce that, for any $t > 0$,

$$\mathbb{P}\left(\mathcal{Q}_{\max,1} \le CL_2\left[\gamma_n^{1/2}(s,d) + \max\left\{\sqrt{t}, \frac{t}{\sqrt{n}}\gamma_n(s,d)\right\}\right]\right) \ge 1 - 4e^{-t}, \tag{6.48}$$

and

$$\mathbb{P}\left(\mathcal{Q}_{\max,2} \le CL_2\left[\gamma_m^{1/2}(s,d) + \max\left\{\sqrt{t}, \frac{t}{\sqrt{m}}\gamma_m(s,d)\right\}\right]\right) \ge 1 - 4e^{-t}, \tag{6.49}$$

where $C > 0$ is an absolute constant, $L_2 = L_1^2 + 1$, $\gamma_n(s,d) = s\log\{\gamma_s(\boldsymbol{\Sigma}_2)\frac{ed}{s}\} \vee s\log n$, and $\gamma_m(s,d) = s\log\{\gamma_s(\boldsymbol{\Sigma}_2)\frac{ed}{s}\} \vee s\log m$. It follows that, for any $t_1, t_2 > 0$,

$$\begin{aligned}\mathbb{P}\Bigg(\mathcal{Q}_{\max} \le CL_2\Bigg[&\gamma_n^{1/2}(s,d) + \gamma_m^{1/2}(s,d) + \max\left\{\sqrt{t_1}, \frac{t_1}{\sqrt{n}}\gamma_n(s,d)\right\}\\ &+ \max\left\{\sqrt{t_2}, \frac{t_2}{\sqrt{m}}\gamma_m(s,d)\right\}\Bigg]\Bigg) \ge 1 - 4e^{-t_1} - 4e^{-t_2}.\end{aligned}$$

Taking $t_1 = \log n$ and $t_2 = \log m$ gives

$$\mathbb{P}\left[\mathcal{Q}_{\max} \le C_{31}L_2\left\{\gamma_m^{1/2}(s,d) + \sqrt{\log m} + \frac{\log m}{\sqrt{m}}\gamma_m(s,d)\right\}\right] \ge 1 - \frac{4}{n} - \frac{4}{m},$$

which proves (6.17). Combining (6.17) and (6.47), and taking $\varepsilon = \varepsilon_3 = \{m\gamma_s(\mathbf{\Sigma}_2)\}^{-1}$, we obtain

$$\begin{aligned}\mathbb{P}\bigg(|\mathcal{Q}_{\max} - \mathcal{M}_{\max}| \le CL_2\bigg[\frac{\gamma_m^{1/2}(s,d)}{m} + \max\bigg\{\frac{\sqrt{\log m}}{m}, \frac{\log m}{m^{3/2}}\gamma_m(s,d)\bigg\}\bigg]\bigg) \\ \ge 1 - \frac{4}{n} - \frac{4}{m}.\end{aligned} \tag{6.50}$$

Recalling the definition of $\mathcal{R}_{ij}$ in (6.19), we have

$$\mathcal{M}_{\max} = \sqrt{\frac{nm}{n+m}} \max_{1\le j\le p_3}\bigg|\frac{\boldsymbol{u}_j^{\mathsf{T}}(\widehat{\mathbf{\Sigma}}_1 - \widehat{\mathbf{\Sigma}}_2)\boldsymbol{u}_j}{\boldsymbol{u}_j^{\mathsf{T}}\mathbf{\Sigma}_2\boldsymbol{u}_j}\bigg| = \sqrt{\frac{n}{m(n+m)}} \max_{1\le j\le p_3}\bigg|\sum_{i=1}^m \mathcal{R}_{ij}\bigg|,$$

where $\mathcal{M}_{\max}$ is as in the proof of Theorem 4.1. Moreover, there exists a p_3-dimensional Gaussian random vector $(\mathcal{G}_1, \dots, \mathcal{G}_{p_3})^{\mathsf{T}}$ satisfying

$$\mathbb{E}(\mathcal{G}_j\mathcal{G}_k) = \frac{n}{m(n+m)}\mathbb{E}\bigg(\sum_{i=1}^m \mathcal{R}_{ij}\mathcal{R}_{ik}\bigg), \qquad 1 \le j \le k \le p_3,$$

such that for every $\delta > 0$,

$$\begin{aligned}&\mathbb{P}\bigg(\Big|\mathcal{M}_{\max} - \max_{1\le j\le p_3}|\mathcal{G}_j|\Big| \ge 16\delta\bigg) \\ &\lesssim \bigg(\frac{n+m}{n}\bigg)^3 L_3^3 \frac{\{\log(2p_3 \vee m)\}^2}{\delta^3\{m(n+m)/n\}^{1/2}} \\ &\quad + \bigg(\frac{n+m}{n}\bigg)^4 K_0^4 L_3^4 \frac{\{\log(p_3+1)\}^4\{\log(2p_3\vee m)\}^3}{\delta^4\{m(n+m)/n\}} \\ &\lesssim L_2^3\frac{\{\log(2p_3\vee m)\}^2}{\delta^3 m^{1/2}} + K_0^4 L_2^4 \frac{\{\log(p_3+1)\}^4\{\log(2p_3\vee m)\}^3}{\delta^4 m},\end{aligned}$$

where $L_3 := (m/n+1)L_2$. It follows that

$$\mathbb{P}\bigg(\Big|\mathcal{M}_{\max} - \max_{1\le j\le p_3}|\mathcal{G}_j|\Big| \ge 16\delta\bigg) \lesssim L_2^3\frac{\gamma_m^2(s,d)}{\delta^3 m^{1/2}} + K_0^4 L_2^4 \frac{\gamma_m^7(s,d)}{\delta^4 m}. \tag{6.51}$$

Taking $\delta = \frac{K_0^4\gamma_m^{5/8}(s,d)}{L_2 m^{1/8}}$, it follows from (6.50) and (6.51) that

$$\mathbb{P}\bigg[\Big|\mathcal{Q}_{\max} - \max_{1\le j\le p_3}|\mathcal{G}_j|\Big| \ge C_{32}K_0^4\frac{\gamma_m^{5/8}(s,d)}{L_2 m^{1/8}}\bigg] \lesssim L_2^2\frac{\gamma_m^{1/8}(s,d)}{m^{1/8}} + L_2^2\frac{\gamma_m^{9/2}(s,d)}{m^{1/2}}.$$

This completes the proof. □

6.2.6. *Proof of Lemma 6.6*

Proof. Write $\widehat{\boldsymbol{\Sigma}} = (m\widehat{\boldsymbol{\Sigma}}_1 + n\widehat{\boldsymbol{\Sigma}}_2)/(n+m)$. By definition, we have

$$
\begin{aligned}
&|\widehat{\mathcal{Q}}_{\max} - \mathcal{Q}_{\max}| \\
&= \left| \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left| \frac{\sqrt{\frac{nm}{n+m}}\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}}_2)\boldsymbol{v}}{\frac{nm}{n+m}\boldsymbol{v}^{\mathsf{T}}(\frac{\widehat{\boldsymbol{\Sigma}}_1}{n} + \frac{\widehat{\boldsymbol{\Sigma}}_2}{m})\boldsymbol{v}} \right| - \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left| \frac{\sqrt{\frac{nm}{n+m}}\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}}_2)\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}} \right| \right| \\
&\le \sqrt{\frac{nm}{n+m}} \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left| \frac{\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}}_2)\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}\boldsymbol{v}} - \frac{\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}}_2)\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}} \right| \qquad (6.52) \\
&\le \sqrt{\frac{nm}{n+m}} \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left| \frac{\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1 - \widehat{\boldsymbol{\Sigma}}_2)\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}} \right| \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left| \frac{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}\boldsymbol{v}} \right| \\
&\quad \times \left\{ \frac{m}{n+m} \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left| \frac{\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_2)\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}} \right| + \frac{n}{n+m} \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left| \frac{\boldsymbol{v}^{\mathsf{T}}(\widehat{\boldsymbol{\Sigma}}_2 - \boldsymbol{\Sigma}_2)\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}} \right| \right\}.
\end{aligned}
$$

Combining (6.48) and (6.49), and taking $t_1 = \log n$ and $t_2 = \log m$, we have

$$
\mathbb{P}\left[\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left| \frac{\boldsymbol{v}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}} - 1 \right| \le C_{34}L_2 \left\{ \frac{\gamma_m^{1/2}(s,d)}{\sqrt{m}} + \frac{\log m}{m}\gamma_m(s,d) \right\} \right] \ge 1 - \frac{4}{n} - \frac{4}{m},
$$

where $C_{34} > 0$ is an absolute constant, $\gamma_m(s,d) = s\log\{\gamma_s(\boldsymbol{\Sigma}_2)\frac{ed}{s}\} \vee s\log m$, and $L_2 = L_1^2 + 1$. It follows that, for all sufficiently large m,

$$
\mathbb{P}\left(\sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} \left| \frac{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}\boldsymbol{v}} \right| \le \left[1 - C_{34}L_2 \left\{ \frac{\gamma_m^{1/2}(s,d)}{\sqrt{m}} + \frac{\log m}{m}\gamma_m(s,d) \right\} \right]^{-1} \right) \ge 1 - \frac{4}{n} - \frac{4}{m}.
$$

This, together with (6.17), (6.48), (6.49), and (6.52), proves (6.21). □

6.2.7. *Proof of Lemma 6.7*

Proof. Define $\mathcal{B}_{\max} = \sup_{\boldsymbol{v}\in\mathbb{V}(s,d)} |\mathcal{B}_{\boldsymbol{v}}|$, where

$$
\mathcal{B}_{\boldsymbol{v}} = \sqrt{\frac{nm}{n+m}} \frac{\boldsymbol{v}^{\mathsf{T}}\{n^{-1}\sum_{i=1}^n \xi_i(\boldsymbol{X}_i\boldsymbol{X}_i^{\mathsf{T}} - \boldsymbol{\Sigma}_2) - m^{-1}\sum_{i=1}^m \eta_i(\boldsymbol{Y}_i\boldsymbol{Y}_i^{\mathsf{T}} - \boldsymbol{\Sigma}_2)\}\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}_2\boldsymbol{v}}
$$

and $\xi_1, \ldots, \xi_n, \eta_1, \ldots, \eta_m$ are independent standard Gaussian random variables that are independent of $\{\boldsymbol{X}_i\}_{i=1}^n$ and $\{\boldsymbol{Y}_i\}_{i=1}^m$. As in Lemma 6.5, we have for $1 \le j \le k \le p_3$,

$$
\begin{aligned}
\mathbb{E}(\mathcal{G}_j\mathcal{G}_k) &= \frac{n}{m(n+m)}\mathbb{E}\left(\sum_{i=1}^n \mathcal{R}_{ij}\mathcal{R}_{ik} + \sum_{i=n+1}^m \mathcal{R}_{ij}\mathcal{R}_{ik} \right) \\
&= \frac{m}{n+m}\mathbb{E}\left\{ \frac{(\boldsymbol{u}_j^{\mathsf{T}}\boldsymbol{X}_i)^2(\boldsymbol{u}_k^{\mathsf{T}}\boldsymbol{X}_i)^2}{\|\boldsymbol{u}_j\|_{\boldsymbol{\Sigma}_2}^2\|\boldsymbol{u}_k\|_{\boldsymbol{\Sigma}_2}^2} - 1 \right\} + \frac{n^2}{m(n+m)}\mathbb{E}\left\{ \frac{(\boldsymbol{u}_j^{\mathsf{T}}\boldsymbol{Y}_i)^2(\boldsymbol{u}_k^{\mathsf{T}}\boldsymbol{Y}_i)^2}{\|\boldsymbol{u}_j\|_{\boldsymbol{\Sigma}_2}^2\|\boldsymbol{u}_k\|_{\boldsymbol{\Sigma}_2}^2} - 1 \right\}
\end{aligned}
$$

$$+\frac{n(m-n)}{m(n+m)}\mathbb{E}\left\{\frac{(\boldsymbol{u}_j^\top \boldsymbol{Y}_i)^2(\boldsymbol{u}_k^\top \boldsymbol{Y}_i)^2}{\|\boldsymbol{u}_j\|_{\boldsymbol{\Sigma}_2}^2\|\boldsymbol{u}_k\|_{\boldsymbol{\Sigma}_2}^2}-1\right\}$$

$$=\frac{m}{n+m}\mathbb{E}\left\{\frac{(\boldsymbol{u}_j^\top \boldsymbol{X}_i)^2(\boldsymbol{u}_k^\top \boldsymbol{X}_i)^2}{\|\boldsymbol{u}_j\|_{\boldsymbol{\Sigma}_2}^2\|\boldsymbol{u}_k\|_{\boldsymbol{\Sigma}_2}^2}-1\right\}+\frac{n}{n+m}\mathbb{E}\left\{\frac{(\boldsymbol{u}_j^\top \boldsymbol{Y}_i)^2(\boldsymbol{u}_k^\top \boldsymbol{Y}_i)^2}{\|\boldsymbol{u}_j\|_{\boldsymbol{\Sigma}_2}^2\|\boldsymbol{u}_k\|_{\boldsymbol{\Sigma}_2}^2}-1\right\}.$$

Putting $\{\mathcal{B}_j\}_{j=1}^{p_3}=\{\mathcal{B}_{\boldsymbol{v}}\}_{\boldsymbol{v}\in\mathbb{N}_{\varepsilon_3}^0}$, we have, for $j=1,\dots,p_3$,

$$\mathcal{B}_j=\sqrt{\frac{nm}{n+m}}\left[\frac{1}{n}\sum_{i=1}^n \xi_i\left\{\frac{(\boldsymbol{u}_j^\top \boldsymbol{X}_i)^2}{\|\boldsymbol{u}_j\|_{\boldsymbol{\Sigma}_2}^2}-1\right\}-\frac{1}{m}\sum_{i=1}^m \eta_i\left\{\frac{(\boldsymbol{u}_j^\top \boldsymbol{Y}_i)^2}{\|\boldsymbol{u}_j\|_{\boldsymbol{\Sigma}_2}^2}-1\right\}\right].$$

It follows that

$$\begin{aligned}&\mathbb{E}(\mathcal{B}_j\mathcal{B}_k|\boldsymbol{X}_1,\dots,\boldsymbol{X}_n,\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_m)\\&\quad=\frac{m}{n+m}\frac{1}{n}\sum_{i=1}^n\left\{\frac{(\boldsymbol{u}_j^\top \boldsymbol{X}_i)^2}{\|\boldsymbol{u}_j\|_{\boldsymbol{\Sigma}_2}^2}-1\right\}\left\{\frac{(\boldsymbol{u}_k^\top \boldsymbol{X}_i)^2}{\|\boldsymbol{u}_k\|_{\boldsymbol{\Sigma}_2}^2}-1\right\}\\&\qquad+\frac{n}{n+m}\frac{1}{m}\sum_{i=1}^m\left\{\frac{(\boldsymbol{u}_j^\top \boldsymbol{Y}_i)^2}{\|\boldsymbol{u}_j\|_{\boldsymbol{\Sigma}_2}^2}-1\right\}\left\{\frac{(\boldsymbol{u}_k^\top \boldsymbol{Y}_i)^2}{\|\boldsymbol{u}_k\|_{\boldsymbol{\Sigma}_2}^2}-1\right\}.\end{aligned}$$

Define

$$\Delta_{\mathcal{G}}=\max_{1\le j\le k\le p_3}\left|\mathbb{E}(\mathcal{G}_j\mathcal{G}_k)-\mathbb{E}(\mathcal{B}_j\mathcal{B}_k|\boldsymbol{X}_1,\dots,\boldsymbol{X}_n,\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_m)\right|.$$

Similar to the proof of Theorem 2.2, it can be shown that with probability greater than $1-O(p_3^{-M})$,

$$\Delta_{\mathcal{G}}^{1/3}\{\log(2p_3)\}^{1/3}\{\log(1/\Delta_{\mathcal{G}})\}^{1/3}\le\Delta_{p_3,m},$$

where $\Delta_{p_3,m}\to 0$ as $p_3,m\to\infty$. By Lemma 6.19, we have

$$\sup_{t\ge 0}\left|\mathbb{P}\Big(\max_{1\le j\le p_3}|\mathcal{G}_j|\le t\Big)-\mathbb{P}\Big(\max_{1\le j\le p_3}|\mathcal{B}_j|\le t\Big|\boldsymbol{X}_1,\dots,\boldsymbol{X}_n,\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_m\Big)\right|=o_{\mathbb{P}}(1).$$

Then, using Lemma 6.1 and Lemma 6.6, we deduce that

$$\sup_{t\ge 0}\left|\mathbb{P}\Big(\max_{1\le j\le p_3}|\mathcal{G}_j|\le t\Big)-\mathbb{P}(\widehat{\mathcal{B}}_{\max}\le t|\boldsymbol{X}_1,\dots,\boldsymbol{X}_n,\boldsymbol{Y}_1,\dots,\boldsymbol{Y}_m)\right|=o_{\mathbb{P}}(1),$$

as desired. □

6.3. Auxillary lemmas

In the sequel, we define $\mathbb{R}^+$ and $\mathbb{Z}^+$ to be the sets of positive real values and integers. The following two lemmas are elementary, yet very useful, in the proofs of the above results.

Lemma 6.8. *For any $\boldsymbol{v} \in \mathbb{S}^{d-1}$, we have*

$$\|\boldsymbol{v}_{\boldsymbol{\Sigma}}\|_{\boldsymbol{\Sigma}} = 1 \quad \textit{and} \quad \left(\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_2}\right)_{\boldsymbol{\Sigma}} = \boldsymbol{v}_{\boldsymbol{\Sigma}}.$$

Proof. By definition, it is straightforward that $\|\boldsymbol{v}_{\boldsymbol{\Sigma}}\|_{\boldsymbol{\Sigma}} = 1$, and

$$\left(\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_2}\right)_{\boldsymbol{\Sigma}} = \frac{\boldsymbol{v}}{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{v}} \Big/ \left\{\frac{\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v}}{(\boldsymbol{v}^{\mathsf{T}}\boldsymbol{v})^2}\right\}^{1/2} = \frac{\boldsymbol{v}}{(\boldsymbol{v}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{v})^{1/2}} = \boldsymbol{v}_{\boldsymbol{\Sigma}},$$

as desired. □

Lemma 6.9. *For $\alpha \in [1, \infty)$, define the function $\psi_\alpha(x) = \exp(x^\alpha) - 1$, $x > 0$. The Orlicz norm for a random variable X is given by*

$$\|X\|_{\psi_\alpha} := \inf\left\{C > 0 : \mathbb{E}\left\{\psi_\alpha\left(\frac{|X|}{C}\right)\right\} \le 1\right\}.$$

Also, define the L^p $(p \ge 1)$ norm of a random variable X by $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$. Then, for every $p \in \mathbb{Z}^+$ we have

$$\|X\|_p \le (p!)^{1/p} \cdot \|X\|_{\psi_1}.$$

Proof. Note that for every $p \in \mathbb{Z}^+$ and $x \in \mathbb{R}^+$, $x^p \le p! \cdot \psi_1(x)$. Then, we have for any $C > 0$,

$$\mathbb{E}\left(\frac{|X|^p}{C^p}\right) \le p! \cdot \mathbb{E}\left\{\psi_1\left(\frac{|X|}{C}\right)\right\}.$$

The conclusion thus follows immediately. □

The following lemma is from Vershynin [43].

Lemma 6.10. *Let (Ω, ρ) be a metric space. For every $\varepsilon > 0$, a subset $\mathbb{N}_\varepsilon(\Omega)$ of Ω is called an ε-net of Ω if for every $\omega \in \Omega$, there is some $\xi \in \mathbb{N}_\varepsilon(\Omega)$ such that $\rho(\omega, \xi) \le \varepsilon$. The minimal cardinality of an ε-net Ω, if finite, is called the covering number of Ω at scale ε, and is denoted by $N(\Omega, \rho, \varepsilon)$. The unit sphere $\mathbb{S}^{d-1}$ equipped with the Euclidean metric satisfies that for every $0 < \varepsilon \le 1$, $N(\mathbb{S}^{d-1}, \rho, \varepsilon) \le (1 + 2/\varepsilon)^d$.*

The following anti-concentration lemma is Theorem 3 in Chernozhukov, Chetverikov and Kato [15] and is used in the proofs of Theorems 2.1 and 4.1.

Lemma 6.11. *Let $(X_1, \ldots, X_d)^{\mathsf{T}}$ be a centered Gaussian random vector in $\mathbb{R}^d$ with $\sigma_j^2 := \mathbb{E}(X_j^2) > 0$ for all $1 \le j \le d$. Define $\underline{\sigma} = \min_{1 \le j \le d} \sigma_j$, $\overline{\sigma} = \max_{1 \le j \le d} \sigma_j$, and $a_d = \mathbb{E}\{\max_{1 \le j \le d}(X_j/\sigma_j)\}$.*

(i) *If* $\underline{\sigma} = \overline{\sigma} = \sigma$, *then for every* $\varepsilon > 0$,

$$\sup_{x\in\mathbb{R}} \mathbb{P}\Big(\Big|\max_{1\le j\le d} X_j - x\Big| \le \varepsilon\Big) \le \frac{4\varepsilon}{\sigma}(a_d + 1).$$

(ii) *If* $\underline{\sigma} < \overline{\sigma}$, *then for every* $\varepsilon > 0$,

$$\sup_{x\in\mathbb{R}} \mathbb{P}\Big(\Big|\max_{1\le j\le d} X_j - x\Big| \le \varepsilon\Big) \le C\varepsilon\big\{a_d + \sqrt{1 \vee \log(\underline{\sigma}/\varepsilon)}\big\},$$

where $C > 0$ *is a constant depending only on* $\underline{\sigma}$ *and* $\overline{\sigma}$.

The following lemma from Adamczak [1] is used in the proof of Lemma 6.2.

Lemma 6.12. *Let* $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ *be independent random variables taking values in a measurable space* $(\mathbb{S}, \mathcal{B})$, *and let* $\mathcal{F}$ *be a countable class of measurable functions* $f : \mathbb{S} \to \mathbb{R}$. *Assume that for* $i = 1, \dots, n$, $\mathbb{E} f(\boldsymbol{X}_i) = 0$ *for every* $f \in \mathcal{F}$ *and* $\|\sup_{f\in\mathcal{F}} |f(\boldsymbol{X}_i)|\|_{\psi_1} < \infty$. *Define*

$$Z = \sup_{f\in\mathcal{F}} \Big|\sum_{i=1}^n f(\boldsymbol{X}_i)\Big| \quad \text{and} \quad \sigma^2 = \sup_{f\in\mathcal{F}} \sum_{i=1}^n \mathbb{E} f^2(\boldsymbol{X}_i).$$

Then, for every $0 < \eta < 1$ *and* $\delta > 0$, *there exists a constant* $C = C(\eta, \delta)$ *such that for all* $t \ge 0$,

$$\begin{aligned}
&\mathbb{P}\big\{Z \ge (1+\eta)\mathbb{E}Z + t\big\} \\
&\quad \le \exp\Big\{-\frac{t^2}{2(1+\delta)\sigma^2}\Big\} + 3\exp\Big\{-\frac{t}{C\|\max_{1\le i\le n}\sup_{f\in\mathcal{F}}|f(\boldsymbol{X}_i)|\|_{\psi_1}}\Big\},
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}\big\{Z \le (1-\eta)\mathbb{E}Z - t\big\} \\
&\quad \le \exp\Big\{-\frac{t^2}{2(1+\delta)\sigma^2}\Big\} + 3\exp\Big\{-\frac{t}{C\|\max_{1\le i\le n}\sup_{f\in\mathcal{F}}|f(\boldsymbol{X}_i)|\|_{\psi_1}}\Big\}.
\end{aligned}$$

The following lemma is Lemma 2.2.2 in van der Vaart and Wellner [42] and is used in the proofs of Lemma 6.2 and Lemma 6.3.

Lemma 6.13. *For any* $\alpha \in [1, \infty)$, *there exists a constant* $K_0 > 0$ *depending only on* α *such that*

$$\Big\|\max_{1\le i\le n} X_i\Big\|_{\psi_\alpha} \le K_0 \psi_\alpha^{-1}(n) \max_{1\le i\le n} \|X_i\|_{\psi_\alpha}.$$

The following lemma is Theorem A in Mendelson [32] and is used in the proof of Lemma 6.2.

Lemma 6.14. *Let $\mathcal{F}$ be a class of mean-zero functions on a probability space $(\mathbb{R}^d, \mu, \mathbb{P})$, and let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent random variables in $\mathbb{R}^d$ distributed according to $\mathbb{P}$. Then, there exists an absolute constant $C > 0$ such that*

$$\mathbb{E}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n f^2(\boldsymbol{X}_i) - \mathbb{E}f^2\right|\right\} \le C\left\{\sup_{f\in\mathcal{F}}\|f\|_{\psi_1}\frac{\gamma_2(\mathcal{F},\psi_2)}{\sqrt{n}} + \frac{\gamma_2^2(\mathcal{F},\psi_2)}{n}\right\}.$$

The complexity parameter $\gamma_2(\mathcal{F},\psi_2)$ of $\mathcal{F}$ is the γ_2 functional with respect to the ψ_2 norm. See Talagrand [38] for its definition and properties.

The following two lemmas are Theorem 2.7.5 and Theorem 2.4.1 in Talagrand [38] on generic chaining, and are used in the proof of Lemma 6.2.

Lemma 6.15. *If $f : (T, \rho) \mapsto (U, \varrho)$ is surjective and there exists a constant $C > 0$ such that*

$$\varrho\big(f(x), f(y)\big) \le C\rho(x, y),$$

for any $x, y \in T$. Then, we have

$$\gamma_\alpha(U, \varrho) \le CK(\alpha)\gamma_\alpha(T, \rho),$$

where $K(\alpha)$ is an absolute constant depending only on α.

Lemma 6.16. *For any metric space (T, ρ) and centered Gaussian process $\{X_t\}_{t\in T}$, there exist universal constants $C > c > 0$ such that*

$$c\gamma_2(T, \rho) \le \mathbb{E}\Big(\sup_{t\in T} X_t\Big) \le C\gamma_2(T, \rho).$$

The following two lemmas are Lemma 1 and Lemma 9 in Chernozhukov, Chetverikov and Kato [14] and are used in the proof of Lemma 6.3.

Lemma 6.17. *Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent centered random vectors in $\mathbb{R}^d$ with $d \ge 2$. Then, there exists a absolute constant $C > 0$ such that*

$$\begin{aligned}
&\mathbb{E}\left[\max_{1\le j,k\le d}\left|\frac{1}{n}\sum_{i=1}^n\big\{X_{ij}X_{ik} - \mathbb{E}(X_{ij}X_{ik})\big\}\right|\right] \\
&\quad\le C\left[\sqrt{\frac{\log d}{n}}\max_{1\le j\le d}\left\{\frac{1}{n}\sum_{i=1}^n\mathbb{E}\big(X_{ij}^4\big)\right\}^{1/2} + \frac{\log d}{n}\left\{\mathbb{E}\Big(\max_{1\le i\le n}\max_{1\le j\le d}X_{ij}^4\Big)\right\}^{1/2}\right].
\end{aligned}$$

Lemma 6.18. *Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent random vectors in $\mathbb{R}^d$ with $d \ge 2$ such that $X_{ij} \ge 0$ for all $1 \le i \le n$ and $1 \le j \le d$. Then*

$$\mathbb{E}\left(\max_{1\le j\le d}\sum_{i=1}^n X_{ij}\right) \lesssim \max_{1\le j\le d}\sum_{i=1}^n\mathbb{E}(X_{ij}) + (\log d)\cdot\mathbb{E}\Big(\max_{1\le i\le n}\max_{1\le j\le d}X_{ij}\Big).$$

The following lemma is Theorem 2 in Chernozhukov, Chetverikov and Kato [15] and is used in the proofs of Theorem 2.2 and Lemma 6.7.

Lemma 6.19. *Let* $\boldsymbol{X} = (X_1, \dots, X_d)^\mathsf{T}$ *and* $\boldsymbol{Y} = (Y_1, \dots, Y_d)^\mathsf{T}$ *be centered Gaussian random vectors in* $\mathbb{R}^d$ *with covariance matrices* $\boldsymbol{\Sigma}^{\boldsymbol{X}} = (\sigma_{jk}^{\boldsymbol{X}})_{1\le j,k\le d}$ *and* $\boldsymbol{\Sigma}^{\boldsymbol{Y}} = (\sigma_{jk}^{\boldsymbol{Y}})_{1\le j,k\le d}$, *respectively. Suppose that* $d \ge 2$ *and* $\sigma_{jj}^{\boldsymbol{Y}} > 0$ *for all* $1 \le j \le d$. *Define*

$$a_d = \mathbb{E}\Big\{\max_{1\le j\le d}\big(Y_j/\sigma_{jj}^{\boldsymbol{Y}}\big)\Big\} \text{ and } \Delta = \max_{1\le j,k\le d}\big|\sigma_{jk}^{\boldsymbol{X}} - \sigma_{jk}^{\boldsymbol{Y}}\big|.$$

Then

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\max_{1\le j\le d} X_j \le x\Big) - \mathbb{P}\Big(\max_{1\le j\le d} Y_j \le x\Big)\Big| \le C\Delta^{1/3}(\log d)^{1/3}\big\{1 \vee a_d^2 \vee \log(1/\Delta)\big\}^{1/3},$$

where $C > 0$ *is an absolute constant depending only on* $\min_{1\le j\le d}\sigma_{jj}^{\boldsymbol{Y}}$ *and* $\max_{1\le j\le d}\sigma_{jj}^{\boldsymbol{Y}}$. *In particular, we have* $a_d \le (2\log d)^{1/2}$ *and*

$$\sup_{x\in\mathbb{R}}\Big|\mathbb{P}\Big(\max_{1\le j\le d} X_j \le x\Big) - \mathbb{P}\Big(\max_{1\le j\le d} Y_j \le x\Big)\Big| \le C'\Delta^{1/3}\big\{1 \vee \log(d/\Delta)\big\}^{2/3},$$

where $C' > 0$ *is an absolute constant depending only on* $\min_{1\le j\le d}\sigma_{jj}^{\boldsymbol{Y}}$ *and* $\max_{1\le j\le d}\sigma_{jj}^{\boldsymbol{Y}}$.

Acknowledgements

The authors sincerely thank the Editor, Associate Editor, and an anonymous referee for their valuable comments and suggestions. A part of this work was carried out when Fang Han was visiting Department of Biostatistics at Johns Hopkins University.

References

[1] Adamczak, R. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.* **13** 1000–1034. MR2424985

[2] Anderson, T.W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Stat.* **34** 122–148. MR0145620

[3] Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd ed. *Wiley Series in Probability and Statistics*. Hoboken, NJ: Wiley-Interscience. MR1990662

[4] Bao, Z., Pan, G. and Zhou, W. (2015). Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.* **43** 382–421. MR3311864

[5] Beran, R. and Srivastava, M.S. (1985). Bootstrap tests and confidence regions for functions of a covariance matrix. *Ann. Statist.* **13** 95–115. MR0773155

[6] Berthet, Q. and Rigollet, P. (2013). Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** 1780–1815. MR3127849

[7] Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106** 672–684. MR2847949

[8] Cai, T., Liu, W. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618

[9] Cai, T., Ma, Z. and Wu, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probab. Theory Related Fields* **161** 781–815. MR3334281

[10] Cai, T.T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19** 2359–2388. MR3160557

[11] Cai, T.T., Ma, Z. and Wu, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110. MR3161458

[12] Chang, J., Zhou, W., Zhou, W.-X. and Wang, L. (2017). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. *Biometrics* **73** 31–41.

[13] Chen, S.X., Zhang, L.-X. and Zhong, P.-S. (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

[14] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461

[15] Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040

[16] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Stat.* **27** 642–669. MR0083864

[17] Eaton, M.L. and Tyler, D.E. (1991). On Wielandt's inequality and its application to the asymptotic distribution of the eigenvalues of a random symmetric matrix. *Ann. Statist.* **19** 260–271. MR1091849

[18] Fan, J. and Wang, W. (2015) Asymptotics of empirical eigen-structure for ultra-high dimensional spiked covariance model. Available at arXiv:1502.04733.

[19] Fang, K.T., Kotz, S. and Ng, K.W. (1990). *Symmetric Multivariate and Related Distributions. Monographs on Statistics and Applied Probability* **36**. London: Chapman & Hall. MR1071174

[20] Fujikoshi, Y. (1980). Asymptotic expansions for the distributions of the sample roots under nonnormality. *Biometrika* **67** 45–51. MR0570503

[21] Hall, P., Härdle, W. and Simar, L. (1993). On the inconsistency of bootstrap distribution estimators. *Comput. Statist. Data Anal.* **16** 11–18. MR1229605

[22] Hall, P., Lee, Y.K., Park, B.U. and Paul, D. (2009). Tie-respecting bootstrap methods for estimating distributions of sets and functions of eigenvalues. *Bernoulli* **15** 380–401. MR2543867

[23] Jiang, T. (2004). The asymptotic distributions of the largest entries of sample correlation matrices. *Ann. Appl. Probab.* **14** 865–880. MR2052906

[24] Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961

[25] Johnstone, I.M. (2008). Multivariate analysis and Jacobi ensembles: Largest eigenvalue, Tracy–Widom limits and rates of convergence. *Ann. Statist.* **36** 2638–2716. MR2485010

[26] Johnstone, I.M. and Lu, A.Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448

[27] Johnstone, I.M. and Nadler, B. (2017). Roy's largest root test under rank-one alternatives. *Biometrika* **104** 181–193.

[28] Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics* **2**, 4th ed. Oxford: Oxford Univ. Press.

[29] Kritchman, S. and Nadler, B. (2009). Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *IEEE Trans. Signal Process.* **57** 3930–3941. MR2683143

[30] Li, J. and Chen, S.X. (2012). Two sample tests for high-dimensional covariance matrices. *Ann. Statist.* **40** 908–940. MR2985938

[31] Ma, Z. (2013). Sparse principal component analysis and iterative thresholding. *Ann. Statist.* **41** 772–801. MR3099121

[32] Mendelson, S. (2010). Empirical processes with a bounded ψ_1 diameter. *Geom. Funct. Anal.* **20** 988–1027. MR2729283

[33] Moghaddam, B., Weiss, Y. and Avidan, S. (2006). Advances in neural information processing systems. 18. In *Proceedings of the* 19*th Annual Conference* (*NIPS*-19) *Held December* 5–10, 2005. *Advances in Neural Information Processing Systems* **18**. Cambridge, MA: MIT Press. Edited by Yair Weiss, Bernard Schölkopf and John C. Platt, A Bradford Book. MR2441315

[34] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR0652932

[35] Péché, S. (2009). Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields* **143** 481–516. MR2475670

[36] Pillai, N.S. and Yin, J. (2012). Edge universality of correlation matrices. *Ann. Statist.* **40** 1737–1763. MR3015042

[37] Roy, S.N. (1957). *Some Aspects of Multivariate Analysis*. New York: Wiley; Calcutta: Indian Statistical Institute. MR0092296

[38] Talagrand, M. (2014). *Upper and Lower Bounds for Stochastic Processes*: *Modern Methods and Classical Problems*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete*. 3. *Folge*. *A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas*. 3*rd Series*. *A Series of Modern Surveys in Mathematics*] **60**. Heidelberg: Springer. MR3184689

[39] Tracy, C.A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** 727–754. MR1385083

[40] Tsybakov, A.B. (2004). *Introduction à L'estimation Non-paramétrique*. *Mathématiques & Applications* (*Berlin*) [*Mathematics & Applications*] **41**. Berlin: Springer. MR2013911

[41] Tyler, D.E. (1983). The asymptotic distribution of principal component roots under local alternatives to multiple roots. *Ann. Statist.* **11** 1232–1242. MR0720268

[42] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. New York: Springer. With applications to statistics. MR1385671

[43] Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge: Cambridge Univ. Press. MR2963170

[44] Waternaux, C.M. (1976). Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika* **63** 639–645. MR0501582

[45] Yuan, X.-T. and Zhang, T. (2013). Truncated power method for sparse eigenvalue problems. *J. Mach. Learn. Res.* **14** 899–925. MR3063614

Received November 2015 and revised July 2016