# A Bootstrap Test for Testing the Equality of Two Ultra-high Dimensional Covariance Matrices

Nilanjan Chakraborty[a,*], Sayar Karmakar[b], Hira L. Koul[c]

[a]*Department of Statistics and Data Science, Washington University St Louis*
[b]*Department of Statistics, University of Florida*
[c]*Department of Statistics and Probability, Michigan State University*

**Abstract**

This paper proposes a test for testing the equality of two high-dimensional covariance matrices in the two-sample set up. This test is based on the maximum of the absolute differences between the entries of the multiplier bootstrap Jackknifed estimators of the two population covariance matrices. The paper also contains an extension of the central limit theorem for one-sample non-degenerate U statistics to the two sample non-degenerate U statistics. This extension is used to derive the asymptotic distributions of the sequence of the proposed test statistics under the null and some local alternative hypotheses. These results are obtained under some weak conditions on the moments of the random vectors and the tails of the marginal distributions. The correlation structures of the random vectors can be arbitrary, the two sample sizes need not be equal and the multivariate dimension is allowed to grow exponentially with the two sample sizes. The test is shown to be consistent against a class of shrinking nonparametric alternatives. A finite sample simulation study reveals some superiority of this test compared to some of the existing tests.

*Keywords:* Jackknife estimator, High dimensional U-statistics, Multiplier bootstrap.
*2020 MSC:* Primary 62H12, Secondary 62F12

## 1. Introduction

The problem of testing the equality of the two covariance matrices in the two sample multivariate set up is a classical problem in statistical inference. It has been well studied in the low-dimensional setting where the multivariate dimension $p$ is fixed and smaller than the sample sizes, see, e.g., Chapter 10, Anderson [1] and the references therein.

In the context of high dimensional data where the number of components $p$ either grows polynomially or even exponentially with increasing sample sizes, this problem has been addressed only in the last decade or so. The tests proposed by Schott [12] and Srivastava and Yanighara [13] are valid for multivariate normal distributions only. A U-statistic test based on an unbiased estimator of the Frobenius norm of the difference of the two population covariance matrices was proposed by Li and Chen [11]. Chen and He [10] proposed another test based on a U-statistic that focuses on the super diagonal elements of the covariance matrices. Cai, Liu and Xia [2] proposed a test based on the maximum of the standardized differences between the entries of the two estimates of the two population covariance matrices. Both of these tests do not require the assumption of Gaussianity of the underlying populations. Further investigation by Cai et al. [2] revealed that the Li and Chen [11] test fails to distinguish between the null and the alternative hypotheses when the difference between the two population covariance matrices is sparse, i.e., when the number of non-zero elements in this difference matrix is small. On the other hand, the Cai et al. [2] test works well only when the difference matrix is sparse.

[*]Corresponding author: Nilanjan Chakraborty, Email address: `chakra46@msu.edu`

Although Cai et al. [2] showed that under certain regularity conditions their test enjoys some optimality in terms of the asymptotic power, it has been pointed out by Fan, Liao and Yao [9] that the convergence of the null distribution of Cai et al. [2] test statistic to Gumbel requires large sample sizes. They also suggested some power-enhancement techniques to achieve the desired power for the test statistic of Cai et al. [2]. Chang, Zhou, Zhou and Wang [3] investigated the finite sample performance of a bootstrap version of the Cai et al. [2] test. The technique involves using multiplier bootstrap approximation result for random vectors after vectorizing the covariance matrices. Their bootstrap method is inapplicable when the two populations means are unknown and unequal because then the sample covariance matrices can no longer be expressed as sums of independent vectors. Moreover, they established consistency of their test under some restrictive conditions like sparsity and other correlational structures.

The need for U statistics based testing approach for covariance matrices can be motivated by noting the fact that the high dimensional central limit theorem fails because the sample covariance matrix can no longer be written as a vectorized sum of independent high dimensional vectors.

In this paper we propose a test for testing the equality of the two population covariance matrices in the high dimensional set up when the two populations means are unknown, under some mild assumptions on the moments and tails of the underlying distributions. The proposed test is based on the maximum of the absolute differences between the entries of the Jackknifed estimators of the two population covariance matrices. We actually use a multiplier bootstrap version of this test statistic. The proposed multiplier bootstrap procedure makes the size and power computation a lot faster. Moreover, the absence of distributional and correlational assumptions makes it applicable more broadly, compared to the above mentioned tests. The proposed test is shown to be consistent against a large class of shrinking alternatives and is argued to be constant rate-optimal against such alternatives. These results are obtained using the seminal works of Chernozhukov, Chetverikov and Kato [5], [6], [7] and Chen [4].

The rest of the article is organized as follows. Section 3 describes the testing problem, the proposed test statistic and the large sample Gaussian approximation of a class of two sample U statistics in the high dimensional set up along with the needed assumptions. This approximation in turn is used to derive the limiting null distribution of the proposed test statistic in Section 4. It is also used to prove the consistency of the test against a sequence of general nonparametric alternatives in Section 5. Findings of a finite sample study reported in Section 6 exhibits some superiority of the proposed test compared to some of the currently popular tests in terms of the empirical level and power. Section 8 contains the proofs of theorems and lemmas.

## 2. Notations

We shall use the following notation and conventions in this paper. The symbol := stands for 'definition'. For a positive integer $p$, $\mathbb{R}^p$ denotes the $p$ dimensional Euclidean space. $\mathbb{R} := \mathbb{R}^1$. For $x \in \mathbb{R}^p$, $x^T$ denotes its transpose and $\|x\|$ its Euclidean norm and $\|x\|_\infty$ denotes its maximum norm. For any two vectors $x = (x_1, \cdots, x_p)^T \in \mathbb{R}^p$ and $y = (y_1, \cdots, y_p)^T \in \mathbb{R}^p$, write $x \le y$ if $x_j \le y_j$ for all $j = 1, \cdots, p$. For any $x = (x_1, \cdots, x_p)^T \in \mathbb{R}^p$ and $a \in \mathbb{R}$, $x + a := (x_1 + a, \cdots, x_p + a)^T$. For any $a, b \in \mathbb{R}$, $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. For any two sequences $a_{d,m,n}$, $b_{d,m,n}, d \wedge m \wedge n \ge 1$, of positive numbers, write $a_{m,n,d} \lesssim b_{m,n,d}$ if for some universal constant $C > 0$, not depending on $m, n, d$, $a_{d,m,n} \le C b_{d,m,n}$, for all sufficiently large $d \wedge m \wedge n$. We write $a_{d,m,n} \sim b_{d,m,n}$ if $a_{d,m,n} \lesssim b_{d,m,n}$ and $b_{d,m,n} \lesssim a_{d,m,n}$.

For a positive integer $q$ and a random vector (r.v.) $Z = (Z_1, \cdots, Z_q)^T$ with finite expectation, $\|Z\|_1 := \sum_{j=1}^p E(|Z_j|)$ and $Z \sim_D G$ means that the distribution function (d.f.) of $Z$ is $G$. For any matrix $A = ((a_{ij}))$, $\|A\|_\infty := \max_{i,j} |a_{ij}|$. For any $p \times p$ symmetric matrix $A$, vec$(A)$ denotes the $d := p(p+1)/2$-dimensional vector consisting of all of the upper-diagonal entries of $A$. For any function $f : \mathbb{R} \to \mathbb{R}$, $\|f\|_\infty := \sup_{z \in \mathbb{R}} |f(z)|$. For a smooth function $g : \mathbb{R}^p \to \mathbb{R}$, we adopt indices to represent the partial derivatives for brevity. For example, $\delta_j \delta_k \delta_l g = g_{jkl}$, where $\delta_j$ denotes the partial derivative with respect to the $j$th coordinate. Let $\psi_\alpha(x) := \exp(x^\alpha) - 1$, $x > 0, \alpha > 0$. For any random variable

$\mathcal{X}$, define

$$\|\mathcal{X}\|_{\psi_\alpha} := \inf\Big\{\lambda > 0 : E\{\psi_\alpha(|\mathcal{X}|/\lambda)\} \le 1\Big\}. \tag{2.1}$$

The entity $\|\mathcal{X}\|_{\psi_\alpha}$ with $\alpha \in [1, \infty)$, is called the Orlicz norm while for $0 < \alpha < 1$, it is called Orlicz quasi-norm. Also let, $A^{Re}$ denote the class of hyper-rectangles in $\mathbb{R}^p$, i.e.,

$$A^{Re} := \Big\{ \prod_{j=1}^{d} [a_j, b_j] : -\infty \le a_j \le b_j \le \infty, \ j = 1, 2, \ldots, d\Big\}.$$

## 3. Gaussian approximation result for U statistics

This section contains the description of the testing problem, the proposed test and the Gaussian approximation results along with the needed assumptions for a large class of the two sample U statistics.

Let $F_j, j = 1, 2$ be possibly two different d.f.'s on $\mathbb{R}^p$. Let $\mu_j$ and $\Sigma_j, j = 1, 2$ denote their mean vectors and covariance matrices, respectively. Let $X^m$ represent the random sample $X_1, \cdots, X_m$ from $F_1$ and $Y^n$ denote the random sample $Y_1, \cdots, Y_n$ from $F_2$. We wish to test $H_0 : \Sigma_1 = \Sigma_2$ versus the alternatives $H_{alt} : \Sigma_1 \ne \Sigma_2$.

To describe the proposed test, let

$$V_m^X := \frac{1}{m(m-1)} \sum_{1\le i\ne j\le m} \frac{\mathrm{vec}((X_i - X_j)(X_i - X_j)^T)}{2} \quad \text{and} \quad V_n^Y := \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \frac{\mathrm{vec}((Y_i - Y_j)(Y_i - Y_j)^T)}{2},$$

be the sample covariance matrices for the $X^m$ and $Y^n$ samples, respectively. Let

$$T_{m,n} := \frac{\sqrt{m}(V_m^X - V_n^Y)}{2}. \tag{3.1}$$

Both $V_m^X, V_n^Y$ are $d := p(p+1)/2$ dimensional U statistics. The proposed test rejects $H_0$ whenever $\|T_{m,n}\|_\infty$ is large. To implement this test in the large samples, we need to know its asymptotic null distribution. Towards that goal, we shall first analyze some asymptotic properties of a general class of two sample U statistics in the high dimensional set up.

Let $\tilde{h}$ be a kernel function from $\mathbb{R}^p \times \mathbb{R}^p \mapsto \mathbb{R}^{p\times p}$ that is symmetric under permutations, i.e., $\tilde{h}(x_1, x_2) = \tilde{h}(x_2, x_1)$, for all $x_1, x_2 \in \mathbb{R}^p$. Thus $\tilde{h}$ is a $p\times p$ symmetric matrix. Let $\mathrm{vec}(h)$ denote the vector representation of $\tilde{h}$, i.e., $\mathrm{vec}(\tilde{h})$ is the $d$-dimensional vector consisting of all the upper-diagonal entries of $\tilde{h}$. Assume $\big\|\mathrm{vec}(\tilde{h}(X_1, X_2))\big\|_1 + \big\|\mathrm{vec}(\tilde{h}(Y_1, Y_2))\big\|_1 < \infty$.

Let $\delta_{m,n}$ be a sequence of real numbers and define

$$U_m^X := \frac{1}{m(m-1)} \sum_{1\le i\ne j\le m} \mathrm{vec}(\tilde{h}(X_i, X_j)), \qquad U_n^Y := \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \mathrm{vec}(\tilde{h}(Y_i, Y_j)),$$

$$\Omega^X := \mathbb{E}\Big(\mathrm{vec}(\tilde{h}(X_1, X_2))\Big), \qquad \Omega^Y := \mathbb{E}\Big(\mathrm{vec}(\tilde{h}(Y_1, Y_2))\Big),$$

$$W_m^X := \frac{\sqrt{m}(U_m^X - \Omega^X)}{2}, \qquad W_n^Y := \frac{\sqrt{n}(U_n^Y - \Omega^Y)}{2}, \qquad W_{m,n} := W_m^X + \delta_{m,n} W_n^Y.$$

Note that if $\tilde{h}(x_1, x_2) \equiv (x_1 - x_2)(x_1 - x_2)^T/2$ and $\delta_{m,n} = -(m/n)^{1/2}$, then $U_m^X \equiv V_m^X$, $U_n^Y \equiv V_n^Y$, $\Omega^X \equiv \Sigma_1$ and $\Omega^Y \equiv \Sigma_2$. Moreover, under $H_0$, $\Omega^X = \Omega^Y$ and $W_{m,n} = T_{m,n}$. We further define the linear projection terms of $U_m^X$ and $U_n^Y$, respectively, to be

$$g(x) := \mathbb{E}\Big(\mathrm{vec}(\tilde{h}(X_1, X_2))\Big| X_1 = x\Big) - \Omega^X, \qquad x \in \mathbb{R}^p,$$

$$\ell(y) := \mathbb{E}\Big(\mathrm{vec}(\tilde{h}(Y_1, Y_2))\Big| Y_1 = y\Big) - \Omega^Y, \qquad y \in \mathbb{R}^p.$$

The $d \times d$ covariance matrices of $g(X)$ and $\ell(Y)$ are, respectively, defined as

$$\Gamma^X := \mathbb{E}\Big(g(X)g(X)^T\Big), \qquad \Gamma^Y := \mathbb{E}\Big(\ell(Y)\ell(Y)^T\Big). \tag{3.2}$$

Let $\text{vec}(h(x_1,x_2)) := \text{vec}(\tilde{h}(x_1,x_2)) - \Omega^X$ and $\text{vec}(h(y_1,y_2)) := \text{vec}(\tilde{h}(y_1,y_2)) - \Omega^Y$ denote the centered version of the kernels $\tilde{h}(x_1,x_2)$ and $\tilde{h}(y_1,y_2))$, respectively. Further, for $x_j, y_j \in \mathbb{R}^p$, $j = 1,2$, let

$$f(x_1,x_2) := \text{vec}(h(x_1,x_2)) - g(x_1) - g(x_2), \qquad q(y_1,y_2) := \text{vec}(h(y_1,y_2)) - \ell(y_1) - \ell(y_2).$$

A kernel $\text{vec}(h) : \mathbb{R}^p \times \mathbb{R}^p \mapsto \mathbb{R}^d$ is said to be non-degenerate if $\text{Var}(g_a(X)) > 0$, for all $a = 1,2,\cdots,d$. It is said to be completely degenerate if $\mathbb{P}(g(X) = 0) = 1$ or equivalently,

$$\mathbb{E}[\text{vec}(h(x_1,X_2))] = \mathbb{E}[\text{vec}(h(X_1,x_2))] = \mathbb{E}[\text{vec}(h(X_1,X_2))] = 0, \quad \forall x_1,x_2 \in \mathbb{R}^p.$$

To state the result about the Gaussian approximation of U statistics we need the following additional notation and assumptions. Let

$$\begin{aligned} L_{m,n} &:= \frac{1}{\sqrt{m}} \sum_{i=1}^{n} g(X_i) + \frac{\delta_{m,n}}{\sqrt{n}} \sum_{j=1}^{n} \ell(Y_j), \\ R_{m,n} &:= \frac{1}{2\sqrt{m}(m-1)} \sum_{1\le i\ne j\le m} f(X_i,X_j) + \frac{\delta_{m,n}}{2\sqrt{n}(n-1)} \sum_{1\le i\ne j\le n} q(Y_i,Y_j). \end{aligned} \tag{3.3}$$

Then,

$$W_{m,n} = L_{m,n} + R_{m,n}. \tag{3.4}$$

Note that, $R_{m,n}$ is a degenerate U statistic while $L_{m,n}$ is non-degenerate. It is reasonable to expect that the distribution of $W_{m,n}$ would be well approximated by that of $L_{m,n}$.

Let $T_m^{G_1}$ and $T_n^{G_2}$ be two independent r.v.'s having $\mathcal{N}_d(0,\Gamma^X)$, $\mathcal{N}_d(0,\Gamma^Y)$ distribution, respectively, and define

$$\begin{aligned} \rho^{**}_{m,n} &:= \sup_{A\in\mathcal{A}^{Re}} \left|\mathbb{P}\Big(\frac{\sqrt{m}(U_m^X - \Omega^X)}{2} + \delta_{m,n}\frac{\sqrt{n}(U_n^Y - \Omega^Y)}{2} \in A\Big) - \mathbb{P}\Big(T_m^{G_1} + \delta_{m,n}T_n^{G_2} \in A\Big)\right| \\ &= \sup_{A\in\mathcal{A}^{Re}} \left|\mathbb{P}\Big(W_m^X + \delta_{m,n}W_n^Y \in A\Big) - \mathbb{P}\Big(T_m^{G_1} + \delta_{m,n}T_n^{G_2} \in A\Big)\right|. \end{aligned}$$

To proceed further we state the needed assumptions, where, for any measurable function $f$ from $\mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^d$, $f_a$ denotes its $a^{\text{th}}$ coordinate, $a = 1(1)d$ and $\|\cdot\|_{\psi_1}$ is as in (2.1).

(a) There exists constants $0 < \underline{b} < \infty$ and $\delta_2 > \delta_1 > 0$ such that $\delta_1 < |\delta_{m,n}| < \delta_2$ and $\inf_{1\le a\le d} \mathbb{E}\Big[g_a^2(X) + \delta_{m,n}^2 \ell_a^2(Y)\Big] > \underline{b}$, $\forall\, m \wedge n \ge 2$.

(b) There exists a sequence of positive constants $B_{m,n}^l, l = 1,2$ such that the following holds with with $\xi_j = X_j,\ j = 1,2$ and $\xi_j = Y_j,\ j = 1,2$.

$$\max_{1\le a\le d} \mathbb{E}\Big[\big|\{\text{vec}(h(\xi_1,\xi_2))\}_a\big|^{2+l}\Big] \le B_{m,n}^l, \quad l = 1,2, \forall\, m \wedge n \ge 2.$$

(c) There exists a sequence of positive constants $B_{m,n}$ such that the following holds with $\xi_j = X_j,\ j = 1,2$ and $\xi_j = Y_j,\ j = 1,2$,

$$\max_{1\le a\le d} \|\{\text{vec}(h(\xi_1,\xi_2))\}_a\|_{\psi_1} \le B_{m,n}, \quad \forall\, m \wedge n \ge 1.$$

(d) $B_{m,n}^2 \log(md)^7 \le Km$ and $B_{m,n}^2 \log(nd)^7 \le Kn$, for some constant $K > 0$ and for all $m \wedge n \ge 2$.

We are now ready to state the following theorem which provides an approximation of the error bound estimate between the probability of interest and its Gaussian counterpart. Its proof is deferred to Section 8.

**Theorem 3.1.** *Under the above set up and assumptions (a)–(d),*

$$\rho^{**}_{m,n} \lesssim \Big(\frac{B^2_{m,n}\log^7(md)}{m}\Big)^{1/6} + \Big(\frac{B^2_{m,n}\log^7(nd)}{n}\Big)^{1/6}. \tag{3.5}$$

**Remark 1.** The above assumptions (a)-(e) have roots in the works of Chernozhukov et al. [7] and Chen [4], which deal with the one sample set up. These conditions are the two sample adaptations of their conditions.

Assumption (a) specifies the restriction on the sequence of constants $\delta_{m,n}$ to be bounded away from zero and infinity. It also ensures the non-degeneracy of the kernel $\mathrm{vec}(h(\xi_1,\xi_2))$ with $\xi_j = X_j, j = 1,2$ and $\xi_j = Y_j, j = 1,2$. Assumption (b) imposes a condition on the third and fourth order moments of the kernel $\mathrm{vec}(h(\xi_1,\xi_2))$ with $\xi_j = X_j, j = 1,2$ and $\xi_j = Y_j,\ j = 1,2$. Li and Chen [11], Cai et al. [2] and Chang et al. [3] assumed the third order moments to be bounded whereas we allow them to diverge to infinity at a rate specified in condition (d). The assumption (c) requires the tail of the kernel to decay exponentially. This assumption is quite common and has been used in all the literatures cited above.

Li and Chen [11] considered some structural assumptions on the traces of the two covariance matrices and Cai et al. [2] imposed some structural assumptions like correlation and sparsity among the components of $X^m$ and $Y^n$. Schott [12], Srivastava and Yanighara [13] assumed the Gaussianity of $X^m$ and $Y^n$.

The above assumptions (a)–(d) are either the same or weaker than those appearing in the above references. They are weaker in the sense that no specific distributional assumption nor additional correlational assumption nor any uniformly bounded moment conditions are needed for the validity of the asymptotic results in this paper about the proposed test.

Although Theorem 3.1 acts as a foundational stone towards the Gaussian approximation of the distribution of $W^X_m + \delta_{m,n}W^Y_n$, but because the limiting distribution is unknown, this theorem is of little use in implementing any test based on $W^X_m + \delta_{m,n}W^Y_n$ for the large sample sizes. To circumvent this problem we are proposing a bootstrap approximation in Theorem 3.2 in the next section, which acts as a crucial step towards bridging this gap.

Instead of applying re-weighted multiplier bootstrap to estimate the unknown covariance matrices we employ the jackknifed version of multiplier bootstrap approximation with jackknifed estimators of the covariance matrices. A reason for choosing this strategy is that the i.i.d re-weighted bootstrap or naive multiplier bootstrap techniques are known to have slower rates of consistency than the jackknifed counterpart, see, e.g., Section 3 in Chen [4].

Let $e_1, e_2, \cdots, e_{m+n}$ be i.i.d. $\mathcal{N}(0,1)$ r.v.'s that are independent of $X^m, Y^n, T^{G_1}_m$ and $T^{G_2}_n$. Define the jackknife versions of $W^X_m$ and $W^Y_n$ , respectively, as

$$W^{eX}_m := \frac{1}{\sqrt{m}}\sum_{i=1}^m\Big[\frac{1}{m-1}\sum_{j\neq i=1}^m \mathrm{vec}(\tilde{h}(X_i,X_j)) - U^X_m\Big]e_i, \quad W^{eY}_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big[\frac{1}{n-1}\sum_{j\neq i=1}^n \mathrm{vec}(\tilde{h}(Y_i,Y_j)) - U^Y_n\Big]e_{i+m}.$$

Define the jackknife estimators of the corresponding covariance matrices of $W^X_m$ and $W^Y_n$ as

$$\hat{\Gamma}^{JK}_m := \frac{1}{(m-1)(m-2)^2}\sum_{i=1}^m\sum_{j\neq i}\sum_{k\neq i}\Big\{\mathrm{vec}(\tilde{h}(X_i,X_j)) - U^X_m\Big\}\Big\{\mathrm{vec}(\tilde{h}(X_i,X_k)) - U^X_m\Big\}^T,$$

$$\hat{\Gamma}^{JK}_n := \frac{1}{(n-1)(n-2)^2}\sum_{i=1}^n\sum_{j\neq i}\sum_{k\neq i}\Big\{\mathrm{vec}(\tilde{h}(Y_i,Y_j)) - U^Y_n\Big\}\Big\{\mathrm{vec}(\tilde{h}(Y_i,Y_k)) - U^Y_n\Big\}^T.$$

Let

$$\tilde{\Gamma}^{JK}_m := \frac{(m-2)^2}{m(m-1)}\hat{\Gamma}^{JK}_m, \quad \tilde{\Gamma}^{JK}_n := \frac{(n-2)^2}{n(n-1)}\hat{\Gamma}^{JK}_n, \text{ and } \hat{\Delta}_{m,n} := \big\|(\tilde{\Gamma}^{JK}_m - \Gamma^X) + \delta^2_{m,n}(\tilde{\Gamma}^{JK}_n - \Gamma^Y)\big\|_\infty. \tag{3.6}$$

For any two random vectors $\xi,\zeta$, the notation $\xi|\zeta$ denotes the conditional distribution of $\xi$, given $\zeta$. Note that, $W^{eX}_m|X^m$ is $\mathcal{N}_d(0,\tilde{\Gamma}^{JK}_m)$ and $W^{eY}_n|Y^n$ is $\mathcal{N}_d(0,\tilde{\Gamma}^{JK}_n)$. We are ready to state the following lemma which plays a crucial role towards obtaining the bootstrap approximation result. Its proof appears in Section 8.

**Lemma 3.1.** *Let $Z^X, Z^Y$ be two independent r.v.'s such that $Z^X|X^m \sim_D \mathcal{N}_d(0, \tilde{\Gamma}_m^{JK})$ and $Z^Y|Y^n \sim_D \mathcal{N}_d(0, \tilde{\Gamma}_n^{JK})$. Then, for some constant $0 < C < \infty$ and every sequence of real numbers $\bar{\Delta}_{m,n} > 0$, on the event $\{\hat{\Delta}_{m,n} \le \bar{\Delta}_{m,n}\}$,*

$$\sup_{A\in\mathcal{A}^{Re}} \Big|\mathbb{P}\Big(Z^X + \delta_{m,n} Z^Y \in A|X^m, Y^n\Big) - \mathbb{P}\Big(T_m^{G_1} + \delta_{m,n} T_n^{G_2} \in A\Big)\Big| \le C(\bar{\Delta}_{m,n})^{1/2} \log d.$$

To proceed further, let $\mathbb{P}_{\boldsymbol{e}}$ denote the conditional distribution of $\boldsymbol{e} := \{e_1, e_2, \cdots, e_{m+n}\}$, given $X^m, Y^n$ and define

$$\rho_{m,n}^{JK} := \sup_{A\in A^{Re}} \big|\mathbb{P}_{\boldsymbol{e}}(W_m^{eX} + \delta_{m,n} W_n^{eY} \in A) - \mathbb{P}(T_m^{G_1} + \delta_{m,n} T_n^{G_2} \in A)\big|.$$

Lemma 3.1 applied with $Z^X = W_m^{eX}$, $Z^Y = W_n^{eY}$, yields a preliminary upper bound for $\rho_{m,n}^{JK}$, which is instrumental in obtaining an improved rate of bootstrap approximation as is evidenced by the next theorem, under the following additional assumption.

(e) There exists a sequence of constant $\gamma_{m,n} \in (0, e^{-1})$ such that for all sufficiently large $d \wedge m \wedge n$,

$$m^{-1} B_{m,n}^2 \log^5(md) \log^2(1/\gamma_{m,n}) \le 1, \qquad n^{-1} B_{m,n}^2 \log^5(nd) \log^2(1/\gamma_{m,n}) \le 1. \tag{3.7}$$

We are ready to state the following theorem.

**Theorem 3.2.** *Under the above setup and assumptions (a)–(c) and (e), the following holds. For a $\gamma_{m,n} < 1/56$, with probability at least $1 - 56\gamma_{m,n}$,*

$$\rho_{m,n}^{JK} \lesssim \Big(\frac{B_{m,n}^2 \log^5(md) \log^2(1/\gamma_{m,n})}{m}\Big)^{1/4} + \Big(\frac{B_{m,n}^2 \log^5(nd) \log^2(1/\gamma_{m,n})}{n}\Big)^{1/4}.$$

The entity $\rho_{m,n}^{JK}$ provides an upper bound to the error of approximation of the bootstrap distribution of the sequence of test statistics $W_m^{eX} + \delta_{m,n} W_n^{eY}$ by the Gaussian counterpart. Theorem 3.2 provides a theoretical guarantee towards the Gaussian approximation term and its jackknife covariance multiplier bootstrap counterpart. It shows that the rate of bootstrap approximation has improved from the rate $\big(\frac{\log^5(nd)}{n}\big)^{1/6}$ given in Chen [4] to $\big(\frac{\log^5(nd)}{n}\big)^{1/4}$.

**Remark 2.** One can choose a sequence $\gamma_{m,n}$ such that $\sum_{m,n} \gamma_{m,n} < \infty$. Then, by the Borel-Cantelli Lemma, the bootstrap convergence result holds almost surely. For example, if $\gamma_{m,n} = (m(\log m))^{-2}$ and $m = n$, then for $m \ge 11$, $\gamma_{m,n} < 1/56$. Then, the choice of $B_{m,n} = Cn^{1/6}$, for some $C > 0$ and $p = e^{n^{1/10}}$ will yield the condition (e).

## 4. Test Procedure

In this section, we shall describe the multiplier bootstrap distribution of the test statistic $\|T_{m,n}\|_\infty$ of (3.1). This is facilitated by Theorems 3.1 and 3.2. The proposed test rejects $H_0$ whenever $\|T_{m,n}\|_\infty$ is large. To implement the test in the large samples, we propose to use multiplier bootstrap version of the $T_{m,n}$ given by $T_{m,n}^{JK} = W_m^{eX} - \sqrt{\frac{m}{n}} W_n^{eY}$, where

$$W_m^{eX} := \sqrt{m}\Big(\frac{1}{m}\sum_{i=1}^m \Big[\frac{1}{m-1}\sum_{j\ne i=1}^m \Big(\frac{\mathrm{vec}((X_i - X_j)(X_i - X_j)^T)}{2} - U_m^X\Big)\Big] e_i\Big),$$
$$W_n^{eY} := \sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n \Big[\frac{1}{n-1}\sum_{j\ne i=1}^n \Big(\frac{\mathrm{vec}((Y_i - Y_j)(Y_i - Y_j)^T)}{2} - U_n^Y\Big)\Big] e_{i+m}\Big).$$

Let

$$c_B(\alpha) = \inf\left\{t \in \mathbb{R} : \mathbb{P}_{\boldsymbol{e}}\left(\left\|W_m^{eX} - \sqrt{\frac{m}{n}} W_n^{eY}\right\|_\infty \le t\right) \ge 1 - \alpha\right\}, \quad 0 < \alpha < 1.$$

Corollary 4.1 below show that the test that rejects $H_0$ whenever $\|T_{m,n}\|_\infty > c_B(\alpha)$ is of the asymptotic size $\alpha$.

From now onwards, for the sake of brevity, let $h(\xi_1, \xi_2) := \dfrac{\mathrm{vec}((\xi_1 - \xi_2)(\xi_1 - \xi_2)^T)}{2}$, $\xi_1, \xi_2 \in \mathbb{R}^d$. The following theorem along with the corollary provides the guarantee of the asymptotic level of the above mentioned test, under the following assumptions.

(a′) For some universal constants $0 < c_1 < c_2 < 1$, $\frac{m}{m+n} \in (c_1, c_2), \forall\, m \wedge n \geq 2$.

(b′) There exists a constant $b > 0$ such that $\mathbb{E}[g_a^2(X) + \delta_{m,n}^2 \ell_a^2(Y)] \geq b$, for all $1 \leq a \leq d$.

(c′) There exists a sequence of constants $B_{m,n} \geq 1$ such that for $l = 1, 2$,

$$\max_{1\leq a\leq d} \mathbb{E}\Big[\big|(\text{vec}(h(X_1, X_2)))_a\big|^{2+l}\Big] \leq B_{m,n}^l, \quad \max_{1\leq a\leq d} \mathbb{E}\Big[\big|(\text{vec}(h(Y_1, Y_2)))_a\big|^{2+l}\Big] \leq B_{m,n}^l,$$

(d′) $\max_{1\leq a\leq d} \big\|(\text{vec}(h(X_1, X_2)))_a\big\|_{\psi_1} \leq B_{m,n}, \quad \max_{1\leq a\leq d} \big\|(\text{vec}(h(Y_1, Y_2)))_a\big\|_{\psi_1} \leq B_{m,n}$,

(e′) $\max\left(\dfrac{B_{m,n}^2 \log^7(dm)}{m}, \dfrac{B_{m,n}^2 \log^7(dn)}{n}\right) \to 0$, as $m \wedge n \to \infty$.

For brevity, let $\mathcal{D} = \Sigma_1 - \Sigma_2$. The Kolmogorov distance between the two distributions of suitably centered $T_{m,n}$ and $T_{m,n}^{JK}$ is defined to be

$$KD\Big(T_{m,n}, T_{m,n}^{JK}\Big) = \sup_{t\geq 0} \left| P\Big(\Big\|\frac{\sqrt{m}(V_m^X - V_n^Y) - \sqrt{m}\,\text{vec}(\mathcal{D})}{2}\Big\|_\infty \leq t\Big) - \mathbb{P}_{\boldsymbol{e}}\Big(\Big\|W_m^{eX} - \sqrt{\frac{m}{n}} W_n^{eY}\Big\|_\infty \leq t\Big)\right|.$$

We are ready to state the following theorem.

**Theorem 4.1.** *Suppose the conditions (a′)–(e′) hold. Then for any non-negative definite matrices $\Sigma_1$ and $\Sigma_2$ of real numbers, $KD(T_{m,n}, T_{m,n}^{JK}) \to 0$, almost surely.*

**Remark 3.** The above conditions (a′)–(e′) are analogous to the assumptions (a)–(e) suitable for the $h$ of the theorem. Condition (a′) specifies that the ratio of the sample sizes can reside in any open interval. Condition (b′) ensures the non-degeneracy of the sample observations. This condition is less restrictive than the minimum eigen-value condition considered in Li and Chen [11], Cai et al. [2] and Chang et al.[3]. Condition (c′) allows the bounds on the third and fourth order moments to grow with the sample sizes $m, n$, unlike as in Cai et al. [2], Li and Chen [11] and Chang et al. [3]. In these papers the moments appearing in (c′) are assumed to be bounded from the above by a fixed constant, for all sample sizes. Condition (d′) allows the sub-exponential tails to grow freely with the sample sizes, which is also advantageous than the analogous conditions in Cai et al. [2] and Chang et al. [3]. The condition (e′) specifies the ultra-high dimensional regime of the test.

On a similar note, in Cai et al. [2], for the convergence of the null distribution of their test statistic to an extreme Type-I distribution or to a normal distribution as in Li and Chen [11], they assumed the sparsity or weak correlation structure among the individual components of the observed random vectors and the corresponding covariance matrices. The jackknifed multiplier bootstrap test proposed in this paper if free of any such correlational assumptions.

The proposed test rejects $H_0 : \Sigma_1 = \Sigma_2$ versus $H_{alt} : \Sigma_1 \neq \Sigma_2$, at the significance level $\alpha \in (0, 1)$, whenever $\varphi_\alpha = 1$, where $\varphi_\alpha = \text{I}(\|T_{m,n}\|_\infty > c_B(\alpha))$. The following corollary is an immediate consequence of Theorem 4.1 and the definition of $c_B(\alpha)$.

**Corollary 4.1.** *Under the conditions of Theorem 4.1 and under $H_0$,*

$$\mathbb{P}\Big(\Big\|\frac{\sqrt{m}(V_m^X - V_n^Y)}{2}\Big\|_\infty \geq c_B(\alpha)\Big) \to \alpha. \tag{4.1}$$

An immediate consequence of this corollary is the formulation of the following confidence region for $\text{vec}(\Sigma_1 - \Sigma_2)$ of the asymptotic confidence level $(1 - \alpha)$:

$$CR_{1-\alpha} := \left\{\text{vec}(\Sigma_1 - \Sigma_2) : \left\|T_{m,n} - \frac{\sqrt{m}\text{vec}(\Sigma_1 - \Sigma_2)}{2}\right\|_\infty \leq c_B(\alpha)\right\}.$$

**A computing procedure for $c_B(\alpha)$.** The multiplier bootstrapped version of the critical value $c_B(\alpha)$ is quite advantageous in terms of faster computation. A procedure for computing is as follows.

**Step 1:** Generate $N$ sets of size $m+n$ i.i.d. $\mathcal{N}(0,1)$ r.v.'s. Denote them by $\boldsymbol{e}_1^*, \cdots, \boldsymbol{e}_N^*$. Treat $\boldsymbol{e}_j^*$ as a copy of $\boldsymbol{e} = \{e_1, e_2, \cdots, e_{m+n}\}$, $1 \le j \le N$.

**Step 2:** Keeping $X^m$ and $Y^n$ fixed, using the vectors $\boldsymbol{e}_j^*$'s of Step 1, compute the bootstrapped version of the test statistic $\|T_{m,n}\|_\infty$ $N$ times, viz., calculate $\|T_{m,n}^{JK}\|_\infty$ $N$ times, $j$th time with $\boldsymbol{e}$ replaced by $\boldsymbol{e}_j^*$, $1 \le j \le N$. Denote these N values by $\{T_{mn1}^{JK}, T_{mn2}^{JK}, \cdots, T_{mnN}^{JK}\}$.

**Step 3:** The $(1-\alpha)^{\text{th}}$ quantile of $\{T_{mn1}^{JK}, T_{mn2}^{JK}, \cdots, T_{mnN}^{JK}\}$ would be treated as an approximate value for $c_B(\alpha)$.

A general criticism of the maximum norm-based statistic of Cai et al. [2] is that the convergence of the null distribution to Gumbel requires relatively large sample sizes, which in turn poses some computational challenges in terms of size and power when the sample sizes are moderate to small. Fan et al. [9] suggested power enhancement techniques in this context. In contrast, the above multiplier bootstrap method makes the power computation a lot faster even without this power enhancement technique.

## 5. Consistency

In this section, we show that the proposed jackknifed multiplier bootstrap based test is consistent against a sequence of shrinking nonparametric alternatives. The power function of the test is

$$\mathbb{P}_{H_{alt}}(\varphi_\alpha = 1) = \mathbb{P}\left(\left\|\frac{\sqrt{m}(V_m^X - V_n^Y)}{2}\right\|_\infty \ge c_B(\alpha) | H_{alt}\right).$$

This power function is an abstract quantity because the respective covariance matrices $\Gamma^X$ and $\Gamma^Y$ of $V_m^X$ and $V_n^Y$ are unknown in practice. To circumvent this problem we define jackknifed multiplier bootstrap based power function

$$\mathbb{P}_{H_{alt}}^*(\varphi_\alpha = 1) := \mathbb{P}_{\boldsymbol{e}^*}\left(\left\|W_m^{e^*X} - \sqrt{\frac{m}{n}} W_n^{e^*Y} + \frac{\sqrt{m}(\text{vec}(\Sigma_1 - \Sigma_2))}{2}\right\|_\infty \ge c_B(\alpha) | H_{alt}\right),$$

where $\mathbb{P}_{\boldsymbol{e}^*}(.)$ denotes the conditional distribution of $\boldsymbol{e}^*$, given all the other r.v.'s. Before exploring the asymptotic theoretical aspects of the power function we shall describe the multiplier bootstrap procedure in the context of approximating the true power function as follows.

Step 1: Generate $\{\boldsymbol{e}_1^*, \boldsymbol{e}_2^*, \cdots, \boldsymbol{e}_{m+n}^*\}$ independent of $\boldsymbol{e}$ which has been used previously to calculate $c_B(\alpha)$.

Step 2: Use the $c_B(\alpha)$ from Step 1 to compute the bootstrap power function for the proposed test by computing:

$$\mathbb{P}_{\boldsymbol{e}^*}\left(\left\|W_m^{e^*X} - \sqrt{\frac{m}{n}} W_n^{e^*Y} + \frac{\sqrt{m}(\text{vec}(\Sigma_1 - \Sigma_2))}{2}\right\|_\infty \ge c_B(\alpha)\right).$$

The following theorem establishes consistency of the proposed test by approximating the true power function $\mathbb{P}_{H_{alt}}(\varphi_\alpha = 1)$, by its jackknifed multiplier bootstrap counterpart $P_{H_{alt}}^*(\varphi_\alpha = 1)$. Its proof also appears in Section 8.

Recall $\mathcal{D} := \Sigma_1 - \Sigma_2$. Let $C > 0$ be a constant and define the sequence of alternatives

$$\mathcal{M}_{m,n,d} = \left\{\mathcal{D} \in \mathbb{R}^{p\times p} : \|\text{vec}(\mathcal{D})/2\|_\infty \ge C(B_{m,n}\log(md)/m)^{1/2}\right\}.$$

**Theorem 5.1.** *Suppose the conditions for Theorem 4.1 hold. Then, for all $\mathcal{D} \in \mathcal{M}_{m,n,d}$,*

$$\mathbb{P}\left(\left\|\frac{\sqrt{m}(V_m^X - V_n^Y)}{2}\right\|_\infty \ge c_B(\alpha)\right) \to 1, \text{ as } n \wedge m \wedge d \to \infty$$

**Remark 4.** Cai et al. [2] and Chang et al. [3] obtained similar consistency results for their test statistics under a class of sparse alternatives. The above theorem generalizes their result in the sense that it is valid for general alternatives where $B_{m,n}$ possibly diverges to infinity. This can be understood by noting that the class $\mathcal{M}_{m,n,d}$ is constructed in a manner such that $\Sigma_1$ and $\Sigma_2$ are separated by a lower bound $K\left(\sqrt{\frac{B_{m,n}\log(md)}{m}}\right)$. Theorem 4 in Cai et al. [2] obtained a similar bound treating $B_{m,n}$ as a constant and their separation parameter was bounded from below by $C\left(\sqrt{\frac{\log d}{m}}\right)$ for some universal constant $C > 0$ under the class of sparse alternatives.

## 6. Simulation studies

In this section, we report the findings of a finite sample study that compares the empirical level and power of the proposed test with those of the popular four tests for testing the equality of the two high dimensional covariance matrices.

We chose the following four competitors to contrast with the performance of the proposed test (CKK test henceforth): They are the tests of Chang et al. [3], Cai et al. [2], Schott [12] and Li and Chen [11] denoted by CZZW, CLX, Sc and LC test, respectively. These tests are currently popular in the existing literature for comparing the covariance matrices in the high-dimensional set up. The CZZW test has similar flavor to ours as they derive their critical values using multiplier bootstrap without jackknifing. The CLX test uses the critical values obtained from its asymptotic Gumbel distribution. The Sc and LC tests are based on the Frobenius norm or the vectorized $l_2$ norm. These tests use critical values obtained from their respective asymptotic normal distributions.

The finite sample performances of the five tests are compared for the following choices of the two covariance matrices, where $\mathbf{1}$ denotes the $p \times 1$ vector of 1's. For $\Sigma_1$ we used the two models M1 and M2, where

$$\text{M1: } \Sigma_{1,M1} = 0.1 I_{p\times p} + 0.9\mathbf{1}\mathbf{1}^T, \qquad \text{M2: } \Sigma_{1,M2} = \left(\left(\rho_1^{|i-j|^{\rho_2}}\right)\right)_{ij} \text{ with } \rho_1 = 0.99, \rho_2 = 0.5.$$

The covariance matrix $\Sigma_2$ is either $\Sigma_1$ under $H_0$ or some small perturbation from $\Sigma_1$ under $H_1$, to be made precise below.

For the two sample sizes and the dimension, we choose $n = 60, m = 60, p = 100$. In our case, $d = (100 \times 101)/2 = 5050$, which falls under the ultra-high dimensional regime. For the distributions of the two samples, we used the following three choices.

- D1: $(X_0)_{ij}, (Y_0)_{ij} \overset{iid}{\sim} N(0, 1)$ and $X = \Sigma_1^{1/2} X_0,\ Y = \Sigma_2^{1/2} Y_0$.
- D2: $(X_0)_{ij}, (Y_0)_{ij} \overset{iid}{\sim} t_{10}$ and $X = \Sigma_1^{1/2} X_0, Y = \Sigma_2^{1/2} Y_0$.
- D3: $(X_0)_{ij}, (Y_0)_{ij} \overset{iid}{\sim} \chi^2_{10} - 10$ and $X = \Sigma_1^{1/2} X_0, Y = \Sigma_2^{1/2} Y_0$.

To assess the sensitivity of these tests to the variances in the base distributions, we chose the variances of the base distributions in D1, D2 and D3 to be 1, 1.25 and 20, respectively. For the evaluation of the empirical sizes of the tests, we chose $\Sigma_1 = \Sigma_2 = \Sigma_{1,M1}$(or $\Sigma_{1,M2}$). For the evaluation of the empirical powers, we chose to explore alternative hypothesis from both the light of how sparse the difference is and how large the signal is. Thus for each of the setup combinations, we generated $\Sigma_2$ by perturbing $\Sigma_1$ as follows, where $0 < \beta < 1$, $v_j$ is the $j^{\text{th}}$ component of a vector $v \in \mathbb{R}^p$ and for any $x \in \mathbb{R}$, $[x]$ is its integer part:

$$\Sigma_2 := \Sigma_1 + \delta v v^T, \text{ with } v_j \overset{iid}{\sim} \text{Unif}(-1, 1), \text{ for } 1 \le j \le [\beta p] \text{ and } v_j = 0, \text{ for } [\beta p] < j \le p.$$

The variable $\beta$ is the sparsity parameter with smaller values indicating the larger sparsity and the larger values correspond to the dense scenarios while $\delta$ measures the distance of the alternative from the null. Larger the value of $\delta$ farther is the alternative from the null.

In the simulation, we chose $\beta = 0.2, 0.4, 0.6, 0.8, 1$ and $\delta = 0, 0.2, 0.4, 0.6, 0.8, 1$. The entries in Tables 1 and 2 below for $\delta = 0\,(\delta > 0)$ represent the empirical levels (the empirical powers) of these tests. These entries are based 500 iterations. For the CKK and CZZW tests, in each empirical iteration a bootstrap sample of size 1000 was used. The nominal level of significance used in the simulation is 0.05.

We now discuss the findings of the simulation reported in Tables 1 and 2, which pertain to the two models M1 and M2, respectively. The empirical level of the CKK test is somewhat conservative for the distribution (D3) and model M1 and somewhat liberal for the distribution (D2) and the model M2. In all other chosen cases of the sparsity parameter values $\beta$, distributions and models, it is close to the nominal level of 0.05. In contrast, other than the CZZW test, the empirical levels of all other tests are quite far off. The CLX test fails to reject almost everywhere whereas the Sc and LC tests tend to be liberal, i.e., their empirical level is significantly higher than the nominal level of 0.05, almost everywhere. The CZZW test is the most competitive with the CKK test in terms of the closeness of the empirical level to the nominal level 0.05 in this simulation.

As far as the empirical power is concerned, we observe that for every chosen value of $\beta$, the empirical power of all tests, except that of the Sc test, increases with the increasing chosen $\delta$ values for all distributions (D1)–(D3) and models M1 and M2.

The empirical power of the CKK test is much higher than that of CLX, Sc and LC tests. Remaining consistent with the low empirical level, the CLX test has the lowest empirical power for every setting. For the Sc and LC tests, one sees that despite starting with relatively higher empirical levels, their empirical powers do not scale properly as $\delta$ increases and remain below those of the CZZW and CKK tests, uniformly for all chosen distributions (D1)–(D3), sparsity levels $\beta$, distance $\delta$ from the null and the models M1 and M2.

Coming to the CZZW test, the empirical power of the CKK test is larger than that of the CZZW test for all choices of the distributions (D1)–(D3), sparsity parameter $\beta$ and models M1 and M2, when $\delta = 0.4, 0.6, 0.8, 1$. For $\delta = 0.2$, this continues to be the case for (D2) while for (D1) and (D3), the empirical power of the CZZW test is at least as large as that of the CKK test, for all the choices of $\beta$ and the models M1 and M2.

Finally, the empirical levels and powers of all the tests used in this study appear to be robust against the chosen three distribution models (D1)–(D3). The tails or the variances of the underlying distributions do not appear to have any effect on the empirical levels and powers of these tests in this simulation study.

## 7. Discussion

This paper proposes a test for testing the equality of the two population covariance matrices in a ultra-high dimensional regime, where the dimension is generally much larger than the sample sizes. The proposed test is based on the maximum of the absolute differences between the entries of the multiplier bootstrap Jackknifed estimators of the two population covariance matrices. The paper contains the proof of the asymptotic normality of the test statistic under the null hypothesis and the consistency of the sequence of the proposed tests against a sequence of shrinking alternatives. A finite sample simulation exhibits some superiority of the proposed test in terms of the empirical level and power, compared to the currently popular four tests. Several further works can be considered following the spirit of the current paper. Sometimes researchers are interested in testing the equality of correlation matrices via Kendall's tau for two populations see for instance Zhou, Han, Zhang and Liu [16]. The methodology proposed in this paper can be extended to those situations, which we leave for future research.

## 8. Some Useful Auxillary Lemmas

Before stating the next lemma we need some definitions. For any functions $f, q$ from $\mathbb{R}^p \times \mathbb{R}^p$ to $\mathbb{R}^p \times \mathbb{R}^p$, define

$$\mathcal{V}_m^X = \frac{1}{m(m-1)} \sum_{1\le i\ne j\le m} f(X_i, X_j), \qquad \mathcal{V}_n^Y = \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} q(Y_i, Y_j),$$

$$M^X = \max_{1\le i\ne j\le m} \max_{1\le a\le d} \left|f_a(X_i, X_j)\right|, \qquad M^Y = \max_{1\le i\ne j\le n} \max_{1\le a\le d} \left|q_a(Y_i, Y_j)\right|,$$

$$D_r^X = \max_{1\le a\le d} \left(\mathbb{E}|f_a(X_1, X_2)|^r\right)^{\frac{1}{r}}, \qquad D_r^Y = \max_{1\le a\le d} \left(\mathbb{E}|q_a(Y_1, Y_2)|^r\right)^{\frac{1}{r}}, \quad r > 0.$$

The following lemma will provide a bound for $R_{m,n}$. The claim (8.1) of this lemma is Theorem 5.1 of Chen [4] while (8.2) follows from (8.1) by applying it to each of the two samples.

| | D1 | | | | | D2 | | | | | D3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta$ | 0.2 | 0.4 | 0.6 | 0.8 | 1 | 0.2 | 0.4 | 0.6 | 0.8 | 1 | 0.2 | 0.4 | 0.6 | 0.8 | 1 |
| CKK | | | | | | | | | | | | | | | |
| $\delta = 0$ | 0.04 | 0.046 | 0.04 | 0.042 | 0.05 | 0.042 | 0.05 | 0.062 | 0.05 | 0.04 | 0.036 | 0.03 | 0.042 | 0.042 | 0.044 |
| $\delta = 0.2$ | 0.06 | 0.08 | 0.09 | 0.096 | 0.096 | 0.072 | 0.094 | 0.112 | 0.1 | 0.096 | 0.062 | 0.074 | 0.076 | 0.086 | 0.080 |
| $\delta = 0.4$ | 0.17 | 0.242 | 0.24 | 0.286 | 0.268 | 0.188 | 0.246 | 0.244 | 0.28 | 0.244 | 0.158 | 0.226 | 0.244 | 0.224 | 0.232 |
| $\delta = 0.6$ | 0.402 | 0.514 | 0.53 | 0.518 | 0.53 | 0.386 | 0.47 | 0.448 | 0.564 | 0.528 | 0.372 | 0.46 | 0.474 | 0.508 | 0.516 |
| $\delta = 0.8$ | 0.648 | 0.774 | 0.786 | 0.790 | 0.778 | 0.616 | 0.722 | 0.722 | 0.796 | 0.774 | 0.634 | 0.726 | 0.752 | 0.766 | 0.774 |
| $\delta = 1$ | 0.830 | 0.898 | 0.918 | 0.926 | 0.918 | 0.776 | 0.882 | 0.886 | 0.914 | 0.924 | 0.812 | 0.872 | 0.906 | 0.904 | 0.916 |
| CZZW | | | | | | | | | | | | | | | |
| $\delta = 0$ | 0.082 | 0.052 | 0.058 | 0.07 | 0.048 | 0.056 | 0.044 | 0.058 | 0.044 | 0.058 | 0.054 | 0.062 | 0.06 | 0.05 | 0.048 |
| $\delta = 0.2$ | 0.076 | 0.092 | 0.092 | 0.092 | 0.128 | 0.076 | 0.076 | 0.088 | 0.096 | 0.11 | 0.072 | 0.092 | 0.094 | 0.112 | 0.114 |
| $\delta = 0.4$ | 0.18 | 0.214 | 0.216 | 0.27 | 0.3 | 0.176 | 0.204 | 0.242 | 0.272 | 0.286 | 0.132 | 0.188 | 0.236 | 0.276 | 0.262 |
| $\delta = 0.6$ | 0.35 | 0.422 | 0.504 | 0.528 | 0.544 | 0.324 | 0.412 | 0.486 | 0.478 | 0.496 | 0.318 | 0.458 | 0.514 | 0.544 | 0.546 |
| $\delta = 0.8$ | 0.532 | 0.648 | 0.702 | 0.726 | 0.772 | 0.548 | 0.64 | 0.72 | 0.752 | 0.758 | 0.51 | 0.682 | 0.708 | 0.802 | 0.754 |
| $\delta = 1$ | 0.694 | 0.848 | 0.882 | 0.904 | 0.932 | 0.726 | 0.836 | 0.9 | 0.906 | 0.924 | 0.724 | 0.816 | 0.876 | 0.916 | 0.904 |
| CLX | | | | | | | | | | | | | | | |
| $\delta = 0$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.002 | 0 | 0 | 0 | 0 | 0.002 | 0.002 | 0.002 |
| $\delta = 0.2$ | 0.002 | 0.002 | 0.004 | 0.002 | 0 | 0.002 | 0 | 0.004 | 0.004 | 0 | 0.004 | 0.004 | 0.002 | 0 | 0.002 |
| $\delta = 0.4$ | 0 | 0.01 | 0.016 | 0.012 | 0.014 | 0.002 | 0.008 | 0.014 | 0.016 | 0.02 | 0.002 | 0.006 | 0.014 | 0.01 | 0.012 |
| $\delta = 0.6$ | 0.022 | 0.024 | 0.04 | 0.062 | 0.062 | 0.018 | 0.04 | 0.04 | 0.04 | 0.064 | 0.016 | 0.038 | 0.048 | 0.072 | 0.05 |
| $\delta = 0.8$ | 0.044 | 0.102 | 0.104 | 0.124 | 0.166 | 0.068 | 0.094 | 0.078 | 0.096 | 0.14 | 0.058 | 0.114 | 0.124 | 0.138 | 0.132 |
| $\delta = 1$ | 0.1 | 0.17 | 0.236 | 0.29 | 0.32 | 0.1 | 0.176 | 0.224 | 0.224 | 0.306 | 0.128 | 0.166 | 0.218 | 0.254 | 0.27 |
| Sc | | | | | | | | | | | | | | | |
| $\delta = 0$ | 0.126 | 0.074 | 0.078 | 0.082 | 0.064 | 0.088 | 0.07 | 0.09 | 0.078 | 0.084 | 0.086 | 0.096 | 0.08 | 0.07 | 0.088 |
| $\delta = 0.2$ | 0.094 | 0.1 | 0.08 | 0.074 | 0.09 | 0.104 | 0.078 | 0.082 | 0.076 | 0.088 | 0.084 | 0.108 | 0.1 | 0.082 | 0.088 |
| $\delta = 0.4$ | 0.112 | 0.094 | 0.078 | 0.088 | 0.09 | 0.09 | 0.072 | 0.086 | 0.08 | 0.092 | 0.074 | 0.068 | 0.086 | 0.088 | 0.096 |
| $\delta = 0.6$ | 0.114 | 0.1 | 0.108 | 0.106 | 0.148 | 0.07 | 0.088 | 0.09 | 0.078 | 0.134 | 0.08 | 0.096 | 0.106 | 0.134 | 0.118 |
| $\delta = 0.8$ | 0.096 | 0.102 | 0.11 | 0.158 | 0.256 | 0.112 | 0.094 | 0.078 | 0.13 | 0.246 | 0.1 | 0.098 | 0.094 | 0.168 | 0.23 |
| $\delta = 1$ | 0.08 | 0.092 | 0.152 | 0.232 | 0.434 | 0.068 | 0.114 | 0.158 | 0.216 | 0.478 | 0.09 | 0.114 | 0.15 | 0.264 | 0.388 |
| LC | | | | | | | | | | | | | | | |
| $\delta = 0$ | 0.156 | 0.1 | 0.092 | 0.106 | 0.072 | 0.1 | 0.094 | 0.108 | 0.09 | 0.102 | 0.104 | 0.112 | 0.098 | 0.098 | 0.114 |
| $\delta = 0.2$ | 0.116 | 0.112 | 0.096 | 0.096 | 0.108 | 0.118 | 0.088 | 0.096 | 0.09 | 0.102 | 0.11 | 0.124 | 0.118 | 0.112 | 0.104 |
| $\delta = 0.4$ | 0.128 | 0.12 | 0.104 | 0.1 | 0.122 | 0.116 | 0.09 | 0.102 | 0.106 | 0.114 | 0.094 | 0.086 | 0.106 | 0.11 | 0.11 |
| $\delta = 0.6$ | 0.14 | 0.118 | 0.132 | 0.13 | 0.18 | 0.1 | 0.112 | 0.118 | 0.096 | 0.166 | 0.1 | 0.118 | 0.138 | 0.172 | 0.152 |
| $\delta = 0.8$ | 0.114 | 0.112 | 0.146 | 0.18 | 0.33 | 0.132 | 0.106 | 0.1 | 0.166 | 0.298 | 0.12 | 0.132 | 0.126 | 0.212 | 0.29 |
| $\delta = 1$ | 0.104 | 0.128 | 0.188 | 0.298 | 0.518 | 0.088 | 0.136 | 0.196 | 0.268 | 0.554 | 0.102 | 0.144 | 0.184 | 0.31 | 0.47 |

**Table 1:** Empirical level and power of 5 tests, for $n = 60, m = 60, p = 100$ and model M1

| | D1 | | | | | D2 | | | | | D3 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\beta$ | 0.2 | 0.4 | 0.6 | 0.8 | 1 | 0.2 | 0.4 | 0.6 | 0.8 | 1 | 0.2 | 0.4 | 0.6 | 0.8 | 1 |
| CKK | | | | | | | | | | | | | | | |
| $\delta = 0$ | 0.048 | 0.038 | 0.042 | 0.046 | 0.048 | 0.052 | 0.052 | 0.058 | 0.06 | 0.05 | 0.048 | 0.04 | 0.046 | 0.05 | 0.042 |
| $\delta = 0.2$ | 0.086 | 0.102 | 0.104 | 0.102 | 0.114 | 0.092 | 0.136 | 0.138 | 0.126 | 0.132 | 0.092 | 0.098 | 0.098 | 0.098 | 0.098 |
| $\delta = 0.4$ | 0.23 | 0.32 | 0.292 | 0.316 | 0.326 | 0.252 | 0.316 | 0.29 | 0.342 | 0.326 | 0.234 | 0.3 | 0.286 | 0.298 | 0.298 |
| $\delta = 0.6$ | 0.476 | 0.574 | 0.554 | 0.594 | 0.59 | 0.47 | 0.532 | 0.54 | 0.624 | 0.574 | 0.502 | 0.544 | 0.588 | 0.592 | 0.6 |
| $\delta = 0.8$ | 0.712 | 0.816 | 0.816 | 0.828 | 0.824 | 0.69 | 0.798 | 0.782 | 0.84 | 0.814 | 0.716 | 0.776 | 0.824 | 0.816 | 0.828 |
| $\delta = 1$ | 0.858 | 0.906 | 0.942 | 0.942 | 0.93 | 0.852 | 0.918 | 0.916 | 0.958 | 0.936 | 0.878 | 0.9 | 0.946 | 0.94 | 0.946 |
| CZZW | | | | | | | | | | | | | | | |
| $\delta = 0$ | 0.04 | 0.056 | 0.038 | 0.056 | 0.054 | 0.05 | 0.048 | 0.066 | 0.05 | 0.058 | 0.056 | 0.052 | 0.046 | 0.048 | 0.066 |
| $\delta = 0.2$ | 0.096 | 0.122 | 0.112 | 0.126 | 0.118 | 0.098 | 0.08 | 0.11 | 0.116 | 0.12 | 0.078 | 0.072 | 0.102 | 0.13 | 0.128 |
| $\delta = 0.4$ | 0.198 | 0.24 | 0.268 | 0.316 | 0.322 | 0.196 | 0.224 | 0.284 | 0.312 | 0.298 | 0.19 | 0.268 | 0.292 | 0.322 | 0.326 |
| $\delta = 0.6$ | 0.394 | 0.52 | 0.532 | 0.582 | 0.582 | 0.364 | 0.482 | 0.574 | 0.598 | 0.596 | 0.388 | 0.488 | 0.558 | 0.556 | 0.56 |
| $\delta = 0.8$ | 0.62 | 0.73 | 0.798 | 0.8 | 0.816 | 0.612 | 0.734 | 0.806 | 0.792 | 0.828 | 0.638 | 0.714 | 0.77 | 0.796 | 0.826 |
| $\delta = 1$ | 0.794 | 0.874 | 0.916 | 0.932 | 0.908 | 0.818 | 0.888 | 0.922 | 0.918 | 0.904 | 0.812 | 0.906 | 0.924 | 0.932 | 0.928 |
| CLX | | | | | | | | | | | | | | | |
| $\delta = 0$ | 0 | 0.002 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\delta = 0.2$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.002 | 0.002 | 0 | 0.002 |
| $\delta = 0.4$ | 0.004 | 0 | 0.014 | 0.016 | 0.006 | 0.002 | 0.002 | 0 | 0.004 | 0.006 | 0.006 | 0.006 | 0.004 | 0.004 | 0.004 |
| $\delta = 0.6$ | 0.006 | 0.008 | 0.018 | 0.032 | 0.036 | 0.014 | 0.016 | 0.032 | 0.034 | 0.036 | 0.014 | 0.022 | 0.034 | 0.028 | 0.04 |
| $\delta = 0.8$ | 0.03 | 0.044 | 0.07 | 0.088 | 0.122 | 0.036 | 0.066 | 0.08 | 0.076 | 0.092 | 0.046 | 0.058 | 0.082 | 0.088 | 0.088 |
| $\delta = 1$ | 0.084 | 0.128 | 0.13 | 0.216 | 0.208 | 0.106 | 0.108 | 0.198 | 0.176 | 0.21 | 0.068 | 0.13 | 0.15 | 0.198 | 0.194 |
| Sc | | | | | | | | | | | | | | | |
| $\delta = 0$ | 0.068 | 0.098 | 0.082 | 0.088 | 0.092 | 0.068 | 0.076 | 0.098 | 0.082 | 0.096 | 0.08 | 0.074 | 0.074 | 0.082 | 0.088 |
| $\delta = 0.2$ | 0.108 | 0.092 | 0.08 | 0.09 | 0.078 | 0.088 | 0.088 | 0.086 | 0.08 | 0.084 | 0.078 | 0.076 | 0.078 | 0.078 | 0.078 |
| $\delta = 0.4$ | 0.088 | 0.086 | 0.106 | 0.096 | 0.11 | 0.076 | 0.068 | 0.086 | 0.098 | 0.102 | 0.086 | 0.11 | 0.1 | 0.098 | 0.102 |
| $\delta = 0.6$ | 0.08 | 0.1 | 0.096 | 0.086 | 0.134 | 0.064 | 0.092 | 0.084 | 0.156 | 0.152 | 0.08 | 0.108 | 0.102 | 0.098 | 0.13 |
| $\delta = 0.8$ | 0.078 | 0.072 | 0.126 | 0.138 | 0.214 | 0.09 | 0.108 | 0.128 | 0.134 | 0.234 | 0.082 | 0.094 | 0.11 | 0.156 | 0.198 |
| $\delta = 1$ | 0.074 | 0.118 | 0.11 | 0.232 | 0.358 | 0.088 | 0.092 | 0.154 | 0.188 | 0.362 | 0.098 | 0.092 | 0.114 | 0.23 | 0.36 |
| LC | | | | | | | | | | | | | | | |
| $\delta = 0$ | 0.09 | 0.118 | 0.09 | 0.11 | 0.12 | 0.082 | 0.098 | 0.12 | 0.104 | 0.104 | 0.104 | 0.094 | 0.096 | 0.106 | 0.106 |
| $\delta = 0.2$ | 0.124 | 0.106 | 0.094 | 0.106 | 0.098 | 0.118 | 0.104 | 0.106 | 0.088 | 0.106 | 0.09 | 0.082 | 0.104 | 0.104 | 0.1 |
| $\delta = 0.4$ | 0.108 | 0.096 | 0.122 | 0.124 | 0.13 | 0.1 | 0.1 | 0.114 | 0.134 | 0.118 | 0.102 | 0.136 | 0.116 | 0.132 | 0.122 |
| $\delta = 0.6$ | 0.098 | 0.112 | 0.11 | 0.122 | 0.176 | 0.09 | 0.112 | 0.102 | 0.174 | 0.17 | 0.11 | 0.136 | 0.126 | 0.116 | 0.158 |
| $\delta = 0.8$ | 0.092 | 0.094 | 0.15 | 0.184 | 0.26 | 0.114 | 0.132 | 0.152 | 0.166 | 0.284 | 0.102 | 0.108 | 0.134 | 0.196 | 0.244 |
| $\delta = 1$ | 0.09 | 0.146 | 0.146 | 0.298 | 0.438 | 0.12 | 0.126 | 0.186 | 0.266 | 0.434 | 0.116 | 0.122 | 0.152 | 0.276 | 0.456 |

**Table 2:** Empirical level and power of 5 tests, for $n = 60, m = 60, p = 100$ and model M2

**Lemma 8.1.** *Let $X^m = (X_1, \cdots, X_m)$ and $Y^n = (Y_1, \cdots, Y_n)$ be two independent random samples from $F_1, F_2$, respectively. Let $f, q : \mathbb{R}^p \times \mathbb{R}^p \mapsto \mathbb{R}^d$ be measurable symmetric functions such that $\mathbb{E}\big|f_a(X_1, X_2)\big| + \mathbb{E}\big|q_a(Y_1, Y_2)\big| < \infty$. If $2 \le d \le \exp(\underline{b}(m \vee n))$, for some constant $\underline{b} > 0$, then $\exists$ a constant $0 < C^X < \infty$ such that*

$$\mathbb{E}\big\|\mathcal{V}_m^X\big\|_\infty \le C^X(1 + \sqrt{\underline{b}})\Big[\Big(\frac{\log(d)}{m}\Big)^{\frac{3}{2}}\|M^X\|_4 + \frac{\log(d)}{m}D_2^X + \Big(\frac{\log(d)}{m}\Big)^{\frac{5}{4}}D_4^X\Big]. \tag{8.1}$$

*Consequently, with $C = \max\{C^X, C^Y\} > 0$, we obtain that*

$$\begin{aligned}
&\mathbb{E}\Big[\big\|\mathcal{V}_m^X - \delta_{m,n}\mathcal{V}_n^Y\big\|_\infty\Big] \qquad (8.2)\\
&\le K(1 + \sqrt{\underline{b}})\left(\left(\frac{\log d}{m}\right)^{\frac{3}{2}}\|M^X\|_4 + \frac{\log(d)}{m}D_2^X + \left(\frac{\log(d)}{m}\right)^{\frac{5}{4}}D_4^X\right)\\
&\quad + \delta_{m,n}\left(\left(\frac{\log(d)}{n}\right)^{\frac{3}{2}}\|M^Y\|_4 + \frac{\log(d)}{n}D_2^Y + \left(\frac{\log(d)}{n}\right)^{\frac{5}{4}}D_4^Y\right).
\end{aligned}$$

To proceed further we need more notation. For $r > 0$ and any sequences of real numbers $\phi_m, \phi_n \ge 1$, define

$$\begin{aligned}
&D_{g,r}^X = \max_{1\le a\le d}\mathbb{E}|g_a(X - \mu^X)|^r, \qquad D_{\ell,r}^Y = \max_{1\le a\le d}\mathbb{E}|\ell_a(Y - \mu^Y)|^r,\\
&M_{g,r}^X(\phi_m) = \mathbb{E}\Big[\max_{1\le a\le d}\big|g_a(X - \mu^X)\big|^r \,\mathrm{I}\Big(\max_{1\le a\le d}\big|g_a(X - \mu^X)\big| > \frac{\sqrt{n}}{4\phi_m\log d}\Big)\Big],\\
&M_{\ell,r}^Y(\phi_n) = \mathbb{E}\Big[\max_{1\le a\le d}\big|\ell_a(Y - \mu^Y)|^r \,\mathrm{I}\Big(\max_{1\le a\le d}\big|\ell_a(Y - \mu^Y)\big| > \frac{\sqrt{n}}{4\phi_n\log d}\Big)\Big],\\
&M_r^{G_1}(\phi_m) = \mathbb{E}\Big[\max_{1\le a\le d}|T_{ma}^{G_1}|^r \,\mathrm{I}\Big(\max_{1\le a\le d}|T_{ma}^{G_1}| > \frac{\sqrt{n}}{4\phi_m\log d}\Big)\Big], \quad M_r^{G_2}(\phi_n) = \mathbb{E}\Big[\max_{1\le a\le d}|T_{na}^{G_2}|^q \,\mathrm{I}\Big(\max_{1\le a\le d}|T_{na}^{G_2}| > \frac{\sqrt{n}}{4\phi_n\log d}\Big)\Big],\\
&M_r^X(\phi_m) = M_{g,r}^X(\phi_m) + M_r^{G_1}(\phi_m), \quad M_r^Y(\phi_n) = M_{\ell,r}^Y(\phi_n) + M_r^{G_2}(\phi_n),\\
&M_{h,r}^X = \mathbb{E}\Big[\max_{1\le i\ne j\le m}\max_{1\le a\le d}\Big|\big(\mathrm{vec}(h(X_i, X_j))\big)_a\Big|^r\Big], \quad M_{h,r}^Y = \mathbb{E}\Big[\max_{1\le i\ne j\le m}\max_{1\le a\le d}\Big|\big(\mathrm{vec}(h(Y_i, Y_j))\big)_a\Big|^r\Big].
\end{aligned}$$

We are ready to state the following lemma. Recall $\underline{b}$ appears in condition (a).

**Lemma 8.2.** *Suppose condition* (a) *holds and $\log(d) \le \bar{b}(m \vee n)$, for some constant $\bar{b} > 0$. Then, for some constants $C_i := C_i(\underline{b}, \bar{b}) > 0, i = 1, 2$ and for any two sequences $\bar{D}_{g,3}^X$ and $\bar{D}_{\ell,3}^Y$ of real numbers satisfying $D_{g,3}^X \le \bar{D}_{g,3}^X$ and $D_{\ell,3}^Y \le \bar{D}_{\ell,3}^Y$,*

$$\begin{aligned}
\rho_{m,n}^{**} \le C_3\Bigg[&\left(\frac{(\bar{D}_{g,3}^X)^2\log^7 d}{m}\right)^{\frac{1}{6}} + \frac{M_3^X(\phi_m)}{\bar{D}_{g,3}^X} + \left(\frac{(\bar{D}_{\ell,3}^Y)^2|\delta_{m,n}|^6\log^7 d}{n}\right)^{\frac{1}{6}} + \frac{M_3^Y(\phi_n)}{\bar{D}_{\ell,3}^Y}\\
&+ \phi^*\Big(\frac{\log^{3/2} d}{m}(M_{h,4}^X)^{1/4} + \frac{\log(d)}{m^{1/2}}(D_2^X)^{1/2} + \frac{\log^{5/4} d}{m^{3/4}}(D_4^X)^{1/4}\\
&+ \frac{\log^{3/2} d}{m}(M_{h,4}^Y)^{1/4} + \frac{\log(d)}{m^{1/2}}(D_2^Y)^{1/2} + \frac{\log^{5/4} d}{m^{3/4}}(D_4^Y)^{1/4}\Big)\Bigg].
\end{aligned} \tag{8.3}$$

*where, $C_3 = \max\{C_1, C_2\}$, $\phi^* := (\max\phi_m, \phi_n)$, with*

$$\phi_m = C_1\left(\frac{(\bar{D}_{g,3}^X)^2\log^4 d}{m}\right)^{-1/6}, \quad \phi_n = C\left(\frac{(\bar{D}_{\ell,3}^Y)^2\log^4 d}{n}\right)^{-1/6}. \tag{8.4}$$

**Proof.** This lemma is analogous to Proposition 5.3 of Chen [4]. We provide details to clearly address the additional changes needed in the proof of Proposition 5.3 to prove the stated lemma. Fix a $y \in \mathbb{R}^p$ and define

$$F_\beta(w) = \frac{1}{\beta}\log\Big(\sum_{j=1}^d\exp\big(\beta(w_j - y_j)\big)\Big), \quad \beta \in \mathbb{R}, w \in \mathbb{R}^p.$$

We shall often use this function with $\beta = \phi \log(d)$, where $\phi \geq 1$. In this case,

$$0 \leq F_\beta(w) - \max_{1\leq j\leq d}(w_j - y_j) \leq \frac{\log(d)}{\beta} = \phi^{-1}, \quad \forall\, w \in \mathbb{R}^d,\ \phi \geq 1.$$

Next, let $u_0 : \mathbb{R} \to [0, 1]$ be a function such that $u_0(t) = 1$, if $t < 0$, $u_0(t) = 0$, if $t > 1$ and $u_0(t)$, $t \in [0, 1]$, is five times continuously differentiable with bounded derivatives. Let

$$u(t) := u_0(\phi t), \qquad \Psi(w) = u(\phi F_\beta(w)), \qquad t \in \mathbb{R}, \phi \geq 1, w \in \mathbb{R}^p.$$

Note that, $\Psi(w) : \mathbb{R}^d \to [0, 1]$. For the later use, we note that when $\beta = \phi \log(d)$,

$$I(t \leq 0) \leq u(t) \leq I(t \leq \phi^{-1}), \quad t \in \mathbb{R}.$$

Let $G_{1i}, H_{1i}, 1 \leq i \leq m$ be i.i.d. $\mathcal{N}_d(0, \Gamma^X)$ r.v.'s and $G_{2j}, H_{2j}, 1 \leq j \leq n$ be i.i.d. $\mathcal{N}_d(0, \Gamma^Y)$ r.v.'s, independent of $G_{1i}, H_{1i}, 1 \leq i \leq m$, where $\Gamma^X = \mathrm{Cov}(g(X))$, $\Gamma^Y := \mathrm{Cov}(\ell(Y))$. Let

$$\begin{aligned}
Z_i^*(t) &:= \frac{1}{\sqrt{m}}\Big[\sqrt{t}\Big\{\sqrt{v} g(X_i) + \sqrt{1-v} G_{1i}\Big\} + \sqrt{1-t} H_{1i}\Big], \quad 1 \leq i \leq m,\\
Z_j^{**}(t) &:= \frac{1}{\sqrt{n}}\delta_{m,n}\Big[\sqrt{t}\Big\{\sqrt{v}\ell(Y_j) + \sqrt{1-v} G_{2j}\Big\} + \sqrt{1-t} H_{2j}\Big], \quad 1 \leq j \leq n,\\
Z^*(t) &:= \sum_{i=1}^m Z_i(t), \quad Z^{**}(t) := \sum_{j=1}^n Z_j^{**}(t), \quad Z(t) = Z^*(t) + Z^{**}(t), \quad v, t \in [0, 1].
\end{aligned}$$

Let

$$\begin{aligned}
I_{m,n} &:= \Psi\Big(\sqrt{v}\frac{1}{\sqrt{m}}\sum_{i=1}^n g(X_i) + \sqrt{1-v}\frac{1}{\sqrt{m}}\sum_{i=1}^n G_{1i} + \sqrt{v}\delta_{m,n}\frac{1}{\sqrt{n}}\sum_{j=1}^n \ell(Y_j)\\
&\qquad + \sqrt{1-v}\delta_{m,n}\frac{1}{\sqrt{n}}\sum_{j=1}^n G_{2j}\Big) - \Psi\Big(\frac{1}{\sqrt{m}}\sum_{i=1}^n H_{1i} + \delta_{m,n}\frac{1}{\sqrt{n}}\sum_{j=1}^n H_{2j}\Big)\\
&= \Psi(Z(1)) - \Psi(Z(0)).
\end{aligned}$$

Recall (8.4). From Xue and Yao [15], (Lemma 2, eqn 99) we obtain

$$\begin{aligned}
\Big|\mathbb{E}[I_{m,n}(v)]\Big| \lesssim C_1(\underline{b})\Big\{&\frac{\phi_m^2 \log^2 d}{\sqrt{m}}\Big[\phi_m D_{g,3}^X \rho_{m,n}^1 + D_{g,3}^X \sqrt{\log(d)} + \phi_m M_3^X(\phi_m)\Big]\\
&+ \frac{\phi_n^2 \log^2 d}{\sqrt{n}}|\delta_{m,n}|^3 \left(\phi_2 D_{\ell,3}^Y \rho_{m,n}^1 + D_{\ell,3}^Y \sqrt{\log(d)} + \phi_n M_3^Y(\phi_n)\right)\Big\}.
\end{aligned}$$

To proceed further, define

$$\begin{aligned}
\rho_{m,n}^1 := \sup_{v\in[0,1]}\sup_{y\in\mathbb{R}^d}\Big|\mathbb{P}\Big(&\sqrt{v}\Big\{\frac{1}{\sqrt{m}}\sum_{i=1}^n g(X_i) + \delta_{m,n}\frac{1}{\sqrt{n}}\sum_{j=1}^n \ell(Y_j)\Big\}\\
&+ \sqrt{1-v}\Big\{\frac{1}{\sqrt{m}}\sum_{i=1}^m G_{1i} + \delta_{m,n}\frac{1}{\sqrt{n}}\sum_{j=1}^n G_{2j}\Big\} \leq y\Big)\\
&- \mathbb{P}\Big(\frac{1}{\sqrt{m}}\sum_{i=1}^n G_{1i} + \delta_{m,n}\frac{1}{\sqrt{n}}\sum_{j=1}^n G_{2j} \leq y\Big)\Big|.
\end{aligned}$$

Note that

$$\rho_{m,n}^1 = \sup_{v\in[0,1]}\sup_{y\in\mathbb{R}^d}\Big|\mathbb{P}\Big(Z(1) \leq y\Big) - \mathbb{P}\Big(Z(0) \leq y\Big)\Big|.$$

Recall from (3.3) and (3.4) that $W_{m,n} - L_{m,n} = R_{m,n}$. Write $R_{m,n} = (R_{m,n,1}, \cdots, R_{m,n,d})^T$. By the Mean Value Theorem,

$$\Psi(W_{m,n}) - \Psi(L_{m,n}) = \sum_{a=1}^{d} \partial_a \Psi(\xi) R_{m,n,a} = \sum_{a=1}^{d} u'(F_\beta(\xi))\eta_a(\xi) R_{m,n,a}$$

where $\eta_a(w) = \partial F_\beta(w)/\partial w_a$ is defined to be the first order partial derivative of $F_\beta(w)$ w.r.t $w_a$ and $\eta := (\eta_1, \cdots, \eta_d)^T$ is a $d \times 1$ random vector on the line segment joining $L_{m,n}$ and $T_{m,n}$. Following the arguments in Xue and Yao [15], we can verify that $\eta_a(w) \geq 0, \sum_{a=1}^d \eta_a(w) = 1$, for any $w \in \mathbb{R}^d$ and there is a constat $K_1(\phi^*)$ such that $\sup_{t\in\mathbb{R}} |u'(t)| \leq K_1(\phi^*)$, where $\phi^* = \max\{\phi_m, \phi_n\}$. Therefore, with $|R_{m,n}|_\infty = \max_{1\leq a\leq d} |R_{m,n,a}|$, we obtain that

$$\Big|\mathbb{E}[\Psi(T_{m,n}) - \Psi(L_{m,n})]\Big| \leq K_1 \phi^* \mathbb{E}|R_{m,n}|_\infty.$$

Proceeding as in Xue and Yao [15] (Eqn (99)) with $\phi = \min\{\phi_m, \phi_n\}$, we conclude that

$$\mathbb{P}\Big(Z(1) \leq y - \phi^{-1}\Big) \leq \mathbb{P}\Big(Z(0) \leq y - \phi^{-1}\Big) + C(\underline{b})\phi^{-1}\sqrt{\log(d)} + |\mathbb{E}[I_{m,n}]| + K_1\phi^*\mathbb{E}\Big(|R_{m,n}|_\infty\Big),$$
$$\mathbb{P}\Big(Z(0) \leq y + \phi^{-1}\Big) \geq \mathbb{P}\Big(Z(1) \leq y + \phi^{-1}\Big) + C(\underline{b})\phi^{-1}\sqrt{\log(d)} + |\mathbb{E}[I_{m,n}]| + K_1\phi^*\mathbb{E}\Big(|R_{m,n}|_\infty\Big).$$

Combining these bounds with the previous equations, we conclude that

$$\begin{aligned}\rho^1_{m,n} \leq\ & K_1\phi^*\mathbb{E}|R_{m,n}|_\infty + C(\underline{b})\phi^{-1}\log^{\frac{1}{2}} d \\ & + C_1(\underline{b})\Big[\frac{(\phi_m)^2\log^2 d}{\sqrt{m}}\Big(\phi_m D^X_{g,3}\rho^1_{m,n} + D^X_{g,3}\sqrt{\log(d)} + \phi_m M^X_3(\phi_m)\Big) \\ & + \frac{|\delta_{m,n}|^3\phi_n^2\log^2 d}{\sqrt{n}}\Big(\phi_n D^Y_{\ell,3}\rho^1_{m,n} + D^Y_{\ell,3}\sqrt{\log(d)} + \phi_n M^Y_3(\phi_n)\Big)\Big].\end{aligned}$$

By similar arguments as used in Lemma 4 of Xue and Yao [15] and choosing $\phi^X_m, \phi^Y_n \geq 1$ we conclude that for any two sequence of real numbers $(\bar{D}^X_{g,3})^2, (\bar{D}^Y_{\ell,3})^2$ such that $(\bar{D}^X_{g,3})^2 \geq D^X_{g,3}$ and $(\bar{D}^Y_{\ell,3})^2 \geq D^Y_{\ell,3}$, $\rho^1_{m,n}$ is bounded from the above by $C_3(\underline{b})$ multiplied by

$$\left[\phi\,\mathbb{E}|R_{m,n}|_\infty + \left(\frac{(\bar{D}^X_{g,3})^2\log^7 d}{m}\right)^{\frac{1}{6}} + \frac{M^X_3(\phi_m)}{\bar{D}^X_{g,3}} + \left(\frac{(\bar{D}^Y_{\ell,3})^2|\delta_{m,n}|^6\log^7 d}{n}\right)^{\frac{1}{6}} + \frac{M^Y_3(\phi_n)}{\bar{D}^Y_{g,3}}\right].$$

By similar arguments as used in Chen [4], Lemma A.1 and Jensen's inequality, there exist universal positive constants $K_2, K_3$ such that the following inequalities hold.

$$\mathbb{E}\Big[\max_{1\leq a\leq d}\max_{1\leq i\neq j\leq m} f_a^4(X_i, X_j)\Big] \leq K_2\mathbb{E}\Big[\max_{1\leq a\leq d}\max_{1\leq i\neq j\leq n}\big(\mathrm{vec}(h(X_i, X_j))\big)^4_a\Big],$$
$$\mathbb{E}\Big[\max_{1\leq a\leq d}\max_{1\leq i\neq j\leq m} q_a^4(Y_i, Y_j)\Big] \leq K_3\mathbb{E}\Big[\max_{1\leq a\leq d}\max_{1\leq i\neq j\leq n}\big(\mathrm{vec}(h(Y_i, Y_j))\big)^4_a\Big].$$

By Lemma 8.1, we obtain that

$$\begin{aligned}\mathbb{E}[\|R_{m,n}\|_\infty] \leq K_3(\bar{b}^{\frac{1}{2}} + 1)\Big[&\frac{\log^{3/2} d}{m}(M^X_{h,4})^{1/4} + \frac{\log(d)}{m^{1/2}}(D^X_2)^{1/2} + \frac{\log^{5/4} d}{m^{3/4}}(D^X_4)^{1/4} \\ & + \frac{|\delta_{m,n}|\log^{3/2} d}{m}(M^Y_{h,4})^{1/4} + \frac{\log(d)}{m^{1/2}}(D^Y_2)^{1/2} + \frac{\log^{5/4} d}{m^{3/4}}(D^Y_4)^{1/4}\Big].\end{aligned}$$

Finally by using Xue and Yao [15] Lemma 3 and Lemma 4, we conclude the proof of this lemma, since

$$\begin{aligned}\rho^{**}_{m,n} \leq C_3\Big[&\left(\frac{(\bar{D}^X_{g,3})^2\log^7 d}{m}\right)^{\frac{1}{6}} + \frac{M^X_3(\phi_m)}{\bar{D}^X_{g,3}} + \left(\frac{(\bar{D}^Y_{\ell,3})^2|\delta_{m,n}|^6\log^7 d}{n}\right)^{\frac{1}{6}} + \frac{M^Y_3(\phi_n)}{\bar{D}^Y_{g,3}} \\ & + \phi^*\Big(\frac{\log^{3/2} d}{m}(M^X_{h,4})^{1/4} + \frac{\log(d)}{m^{1/2}}(D^X_2)^{1/2} + \frac{\log^{5/4} d}{m^{3/4}}(D^X_4)^{1/4} \\ & + \frac{\log^{3/2} d}{m}(M^Y_{h,4})^{1/4} + \frac{\log(d)}{m^{1/2}}(D^Y_2)^{1/2} + \frac{\log^{5/4} d}{m^{3/4}}(D^Y_4)^{1/4}\Big)\Big].\end{aligned}$$

$\square$

**Appendix**

This section contains the proof of theorems and lemmas. In the proofs below, $C$ denotes a large enough finite positive constant, not depending on $m, n, d$ and that may be different in different context.

**Proof of Theorem 3.1.** This theorem is a two sample version of the Theorem 2.1 of Chen [4]. Some detailed calculations are still needed so we provide the proof for the sake of completeness. The objective of the proof is to quantify the bounds obtained in (8.3) in terms of $d, m$ and $n$. First, we obtain explicit rates for each summand of the upper bound of (8.3). Then, these rates are combined to obtain an overall rate bound for $\rho^{**}_{m,n}$.

Recall $g_a(x) = \mathbb{E}[(\text{vec}(h(\xi,\xi_2)))_a|\xi = x],\ x \in \mathbb{R}$. For any random vectors $\xi, \xi_1, \xi_2$, let

$$D^{\xi}_2 := \max_{1\le a\le d} \mathbb{E}\big|(\text{vec}(h(\xi_1,\xi_2)))_a\big|^2,\ D^{\xi}_{g,3} := \max_{1\le a\le d} \mathbb{E}\big|g_a(\xi)\big|^3,\ D^{\xi}_4 := \max_{1\le a\le d} \mathbb{E}\big|(\text{vec}(h(\xi_1,\xi_2)))_a\big|^4.$$

The Jensen's inequality and the assumption (b) yield that

$$D^{\xi}_2 = \max_{1\le a\le d} \mathbb{E}\big|(\text{vec}(h(\xi_1,\xi_2)))_a\big|^2 \le \max_{1\le a\le d} (\mathbb{E}\big|(\text{vec}(h(\xi_1,\xi_2)))_a\big|^3)^{\frac{2}{3}} \le B^{\frac{2}{3}}_{m,n}.$$

To analyse $D^{\xi}_{g,3}$, assumption (b) implies that

$$\begin{aligned}
&\mathbb{E}\big|g_a(\xi)\big|^3 \le \mathbb{E}\Big(\mathbb{E}\big[\big|(\text{vec}(h(\xi_1,\xi_2)))_a\big|^3\big|\xi_1\big]\Big) = \mathbb{E}\big[\big|(\text{vec}(h(\xi_1,\xi_2)))_a\big|^3\big] \le B_{m,n},\ \forall\, 1\le a\le d,\\
&D^{\xi}_{g,3} \le B_{m,n}.
\end{aligned}$$

Again, by the condition (b), we readily obtain $D^{\xi}_4 \le B^2_{m,n}$.

Next, consider $M^{\xi}_{h,4}$. Using a property of Orlicz norm, see Van der Vaart and Wellner [14] (pg-96), we obtain

$$M^{\xi}_{h,4} = \mathbb{E}\Big[\max_{1\le i\ne j\le n}\max_{1\le a\le d}\big|(\text{vec}(h(\xi_i,\xi_j)))_a\big|^4\Big] \le C\Big\|\max_{1\le i\ne j\le n}\max_{1\le a\le d}(\text{vec}(h(\xi_i,\xi_j)))_a\Big\|^4_{\psi_1}.$$

This bound, Lemma 2.2.2 of Van der Vaart and Wellner [14] and (c) together yield that

$$M^{\xi}_{h,4} \le C\log^4(md)\Big[\max_{1\le i\ne j\le n}\max_{1\le a\le d}\big\|(\text{vec}(h(\xi_i,\xi_j)))_a\big\|_{\psi_1}\Big]^4 \le C\log^4(md)B^4_{m,n}. \tag{8.5}$$

We are now ready to obtain the overall rates for the upper bound of $\rho^{**}_{m,n}$ of (8.3). Recall $\phi_m$ from (8.4) and take $B_{m,n} = \bar{D}^{\xi}_{g,3}$. Let $\varpi^{\xi}_{m,n} := \Big(\frac{B^2_{m,n}\log^7(md)}{m}\Big)^{1/6}$, if $\xi = X$ and $\varpi^{\xi}_{m,n} := \Big(\frac{B^2_{m,n}\log^7(nd)}{n}\Big)^{1/6}$, if $\xi = Y$.

By (8.5) and the definitions of the entities involved,

$$\begin{aligned}
\frac{\phi_m(\log d)^{\frac{3}{2}}(M^{\xi}_{h,4})^{\frac{1}{4}}}{m} &\le C\Big(\frac{(\bar{D}^{\xi}_{g,3})^2\log^4 d}{m}\Big)^{-1/6}\Big(\frac{\log d}{m}\Big)^{\frac{3}{2}}(M^{\xi}_{h,4})^{\frac{1}{4}}\\
&\le C\Big(\frac{B^2_{m,n}(\log(md))^7}{m}\Big)^{\frac{1}{6}}\Big(\frac{\log(md)}{m}\Big)^{\frac{4}{6}} \le C\varpi^{\xi}_{m,n},\\
\phi_m\ \frac{\log(d)}{\sqrt{m}}D^{1/2}_2 &\le \Big(\frac{D^2_{g,3}\log^4 d}{m}\Big)^{-1/6}\ \frac{\log(d)}{\sqrt{m}}\ B^{1/3}_{m,n}\\
&\le \Big(\frac{B^2_{m,n}(\log(nd))^7}{n}\Big)^{1/6}\Big(\frac{1}{B^2_{m,n}\log^5(dm)}\Big)^{1/6} \le C\varpi^{\xi}_{m,n},\\
\phi_m\ \frac{\log^{5/4} d}{n^{3/4}}D^{1/4}_4 &\le C\Big(\frac{D^2_{g,3}\log^4 d}{m}\Big)^{-1/6}\ \frac{\log^{5/4} d}{m^{3/4}}\ B^{1/2}_{m,n}\\
&\le \Big(\frac{B^2_{m,n}(\log(nd))^7}{n}\Big)^{1/6}\ \Big(\frac{\log(d)}{m}\Big)^{7/12}\ \Big(\frac{1}{B^2_{m,n}}\Big)^{1/12} \le C\varpi^{\xi}_{m,n}.
\end{aligned}$$

Next, we shall obtain a bound for $M_3^{\xi}(\phi_m)$. By Lemma C.1 of Chernozhukov et al. [7], applied with $B_{m,n} = \bar{D}^{\xi}_{g,3}$, we obtain that for some universal constant $c^* > 0$,

$$M_3^{\xi}(\phi_m) \lesssim \Big(\frac{\sqrt{m}}{\phi_m \log(d)} + B_{m,n}\log(d)\Big)^3 + \exp\Big[-\frac{\sqrt{m}}{\big(4c^*\phi_m B_{m,n}(\log(d))^2\big)}\Big].$$

Since $\phi_m \geq 2$, $\dfrac{\sqrt{m}}{\phi_m \log(d)} \lesssim \frac{\sqrt{m}}{\log(d)} \lesssim \sqrt{m}$ and $B_{m,n}\log(d) \lesssim \sqrt{m}$ together yield that

$$\Big(\frac{\sqrt{m}}{4c^1\phi_m B_{m,n}(\log(d))^2}\Big) \gtrsim \bigg(\frac{(B_{m,n})^2(\log^7(dm))}{m}\bigg)^{-1/3}\log(dm) \gtrsim c^*\log(dm).$$

Combine these bounds to obtain that $M^{\xi}_{g,3}(\phi_m) \lesssim m^{3/2}(md)^{-c^*} \lesssim m^{-1/2}$. Similar arguments yield that $M_3^{G_1}(\phi_m) \lesssim m^{-1/2}$. The last two facts used with $\xi = X$ in turn yield that

$$M_3^X(\phi_m) = M^X_{g,3}(\phi_m) + M_3^{G_1}(\phi_m) \lesssim m^{-1/2}.$$

Finally we have that,

$$\Big(\frac{(\bar{D}^X_{g,3})^2\log^7 d}{m}\Big)^{1/6} + \frac{M_3^X(\phi_m)}{\bar{D}^X_{g,3}} \leq \Big(\frac{B^2_{m,n}\log^7 d}{m}\Big)^{1/6} + \frac{1}{\sqrt{m}B_{m,n}} \lesssim \Big(\frac{B^2_{m,n}\log^7(md)}{m}\Big)^{1/6}.$$

Similar calculations as the above used with $\xi = Y$ and the assumption (a) yield that

$$\Big(\frac{(\bar{D}^Y_{g,3})^2(\log^7 d)|\delta_{m,n}|^6}{n}\Big)^{1/6} + \frac{M_3^Y(\phi_n)}{\bar{D}^Y_{g,3}} \leq \Big(\frac{B^2_{m,n}(\log^7 d)|\delta_{m,n}|^6}{n}\Big)^{1/6} + \frac{1}{\sqrt{n}B_{m,n}} \lesssim \Big(\frac{B^2_{m,n}\log^7(nd)}{n}\Big)^{1/6}.$$

The above bounds combined with (8.3) readily yield the bound (3.5), thereby completing the proof of Theorem 3.1.

**Proof of Lemma 3.1.** The proof is an immediate consequence of Theorem 5.1, Chernozhukov et al. [8] applied with $Z = T_m^{G_1} + \delta_{m,n}T_n^{G_2} \sim \mathcal{N}_d(0, \Gamma^X + \delta^2_{m,n}\Gamma^Y)$ and $V = (Z_1^X + \delta_{m,n}Z_2^Y)\big|(X^m, Y^n) \sim \mathcal{N}_d(0, \tilde{\Gamma}^{JK}_m + \delta^2_{m,n}\tilde{\Gamma}^{JK}_n)$.

**Proof of Theorem 3.2.** Throughout the proof below, $m, n, d$ are large enough so that $\log(md) > 1, \log(nd) > 1$. Recall the definition of $\hat{\Delta}_{m,,n}$ from (3.6). By Lemma 3.1, for any sequence of constants $\bar{\Delta}_{m,n} > 0$, on the event $\{\hat{\Delta}_{m,n} \leq \bar{\Delta}_{m,n}\}$, $\rho^{JK}_{m,n} \lesssim (\bar{\Delta}_{m,n})^{1/2}\log d$.

The goal here is to find a real sequence $\bar{\Delta}_{m,n}$ such that $\mathbb{P}(\hat{\Delta}_{m,n} \geq \bar{\Delta}_{m,n}) \leq \gamma_{m,n}$, and then obtain a bound for the upper bound $(\bar{\Delta}_{m,n})^{1/2}\log d$. Towards this goal, we shall first bound obtain a rate bound for $\hat{\Delta}_{m,n}$.

For the sake of brevity, let $m_1 := m(m-1)^2$. To bound $\hat{\Delta}_{m,n}$, rewrite $\tilde{\Gamma}^{JK}_m$ as,

$$\begin{aligned}
\tilde{\Gamma}^{JK}_m &= \frac{1}{m_1}\sum_{i=1}^m\sum_{j\neq i}\sum_{k\neq i}\Big\{\mathrm{vec}(\tilde{h}(X_i,X_j)) - U^X_m\Big\}\Big\{\mathrm{vec}(\tilde{h}(X_i,X_k)) - U^X_m\Big\}^T,\\
&= \frac{1}{m_1}\sum_{i=1}^m\sum_{j\neq i}\sum_{k\neq i}\Big\{\mathrm{vec}(h(X_i,X_j)) - (U^X_m - \Sigma^X)\Big\}\Big\{\mathrm{vec}(h(X_i,X_k)) - (U^X_m - \Sigma^X)\Big\}^T,\\
&= \frac{1}{m_1}\Big(1-\frac{1}{m}\Big)\Bigg[\sum_{1\leq i\neq j\leq m}\{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_i,X_j))\}^T + \sum_{1\leq i\neq j\neq k\leq m}\{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_i,X_k))\}^T\Bigg]\\
&\quad - \frac{1}{m_1}\frac{1}{m}\Bigg[\sum_{1\leq i\neq j\neq l\leq m}\Big\{\mathrm{vec}(h(X_i,X_j))\Big\}\{\mathrm{vec}(h(X_l,X_j))\}^T + \sum_{1\leq i\neq j\leq m}\{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_i,X_j))\}^T\\
&\qquad + \sum_{1\leq i\neq j\neq k\leq m}\{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_j,X_k))\}^T + \sum_{1\leq i\neq j\neq l\leq m}\{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_i,X_l))\}^T\\
&\qquad + \sum_{1\leq i\neq j\neq l\neq k\leq m}\{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_l,X_k))\}^T\Bigg],
\end{aligned}$$

$$= \tilde{\Gamma}^{JK}_{m1} - \tilde{\Gamma}^{JK}_{m2}, \qquad \text{(say).}$$

Thus to obtain a bound for $\|\tilde{\Gamma}^{JK}_{m} - \Gamma^{X}\|_\infty$, it suffices to obtain bounds for $\|\tilde{\Gamma}^{JK}_{m2}\|_\infty$ and $\|\tilde{\Gamma}^{JK}_{m1} - \Gamma^{X}\|_\infty$.

Define the approximation rates

$$\varpi^{BX}_m(\gamma_{m,n}) := \Big(\frac{B^2_{m,n}\log^5(md)\log^2(1/\gamma_{m,n})}{m}\Big)^{1/4}, \qquad \varpi^{BY}_n(\gamma_{m,n}) := \Big(\frac{B^2_{m,n}\log^5(nd)\log^2(1/\gamma_{m,n})}{n}\Big)^{1/4}. \tag{8.6}$$

*Rate bound for* $\tilde{\Gamma}^{JK}_{m2}$. We begin with the decomposition

$$\tilde{\Gamma}^{JK}_{m2} = \frac{(m-2)(m-3)}{m(m-1)}\,\tilde{\Gamma}^{JK}_{m,2,4} + \frac{3(m-2)}{m(m-1)}\,\tilde{\Gamma}^{JK}_{m,2,3} + \frac{1}{m(m-1)}\,\tilde{\Gamma}^{JK}_{m,2,2}, \tag{8.7}$$

where

$$\begin{aligned}
\tilde{\Gamma}^{JK}_{m,2,4} &:= \frac{(m-4)!}{m!}\sum_{1\le i\ne j\ne k\ne l\le m}\{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_k,X_l))\}^T,\\
\tilde{\Gamma}^{JK}_{m,2,3} &:= \frac{(m-3)!}{m!}\sum_{1\le i\ne j\ne k\le m}\{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_i,X_k))\}^T,\\
\tilde{\Gamma}^{JK}_{m,2,2} &:= \frac{(m-2)!}{m!}\sum_{1\le i\ne j\le n}\{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_i,X_j))\}^T.
\end{aligned}$$

Consider the first term on the right hand side (r.h.s.) of (8.7). Let $H(x_1,x_2,x_3,x_4) = \mathrm{vec}(h(x_1,x_2))\mathrm{vec}(h(x_3,x_4))^T$. Then,

$$\tilde{\Gamma}^{JK}_{m,2,4} = \frac{(m-4)!}{m!}\sum_{1\le i\ne j\ne k\ne l\le m} H(X_i,X_j,X_k,X_l).$$

Note that, $\tilde{\Gamma}^{JK}_{m,2,4}$ is a U statistics of order four and $\mathbb{E}[\tilde{\Gamma}^{JK}_{m,2,4}] = 0$. Let $r = [m/4]$ and define

$$\begin{aligned}
&Z^X_{m,2,4} := r\|\tilde{\Gamma}^{JK}_{m,2,4}\|_\infty, \qquad M^X_{m,2,4} = \max_{0\le i\le r-1}\max_{1\le a\le d}\Big|H_a\Big(X_{4i+1},X_{4i+2},X_{4i+3},X_{4i+4}\Big)\Big|,\\
&(\bar{\varsigma}^X_{m,2,4})^2 := \max_a \sum_{i=0}^{r-1}\mathbb{E}\Big[H_a^2\Big(X_{4i+1},X_{4i+2},X_{4i+3},X_{4i+4}\Big)\Big],\\
&\bar{Z}^X_{m,2,4} := \max_a\Big|\sum_{i=0}^{r-1}\Big[\bar{H}_a(X_{4i+1},X_{4i+2},X_{4i+3},X_{4i+4}) - \mathbb{E}\bar{H}_a\Big]\Big|,\\
&\bar{H}_a\Big((X)_1^4\Big) := H_a((X)_1^4)I\Big(\max_a\Big|H_a((X)_1^4)\Big|\le\tau\Big), \quad \tau\ge 0,
\end{aligned}$$

where $a := (a_1,a_2)^T$, $a_1,a_2 = 1,\cdots,d$.

By Lemma (E.1) of Chen [4] applied with $\alpha = \frac{1}{2}, \eta = 1, \delta = \frac{1}{2}$ and $\tau = \mathbb{E}[M^X_{m,2,4}]$,

$$\mathbb{P}\Big(Z^X_{m,2,4} \ge \mathbb{E}[\bar{Z}^X_{m,2,4}] + t\Big) \le \exp\Big(-\frac{t^2}{3(\bar{\varsigma}^X_{m,2,4})^2}\Big) + 3\exp\Big[-\Big(\frac{t}{C\|M^X_{m,2,4}\|_{\psi_{\frac{1}{2}}}}\Big)^{1/2}\Big], \ \ \forall\, t>0.$$

Moreover,

$$\begin{aligned}
\mathbb{E}[\bar{Z}^X_{m,2,4}] \le C\Big\{&(\log(d))^{1/2}\Big[\max_a\sum_{i=0}^{r-1}\Big[\mathbb{E}\Big(\bar{H}_a(X_{4i+1},X_{4i+2},X_{4i+3},X_{4i+4}) - \mathbb{E}\bar{H}_a\Big)^2\Big]^{\frac{1}{2}}\\
&+ (\log(d))[\mathbb{E}\Big[\max_{i,a}\Big|\bar{H}_a(X_{4i+1},X_{4i+2},X_{4i+3},X_{4i+4}) - \mathbb{E}[\bar{H}_a]\Big|^2\Big]^{1/2}\Big\},\\
\le C\{&(\log(d))^{1/2}\bar{\varsigma}^X_{m,2,4} + (\log(d))\|M^X_{m,2,4}\|_{\psi_{1/2}}.
\end{aligned}$$

By the Cauchy-Schwarz inequality inequality and Condition (b),

$$\mathbb{E}\Big[H_a^2(X_{4i+1}, X_{4i+2}, X_{4i+3}, X_{4i+4})\Big] \le \Big[\mathbb{E}\Big\{\mathrm{vec}\Big(h(X_{4i+1}, X_{4i+2})\Big)\Big\}_{a_1}^4\Big]^{\frac{1}{2}}\Big[\mathbb{E}\Big\{\mathrm{vec}\Big(h(X_{4i+3}, X_{4i+4})\Big)\Big\}_{a_2}^4\Big]^{\frac{1}{2}} \le B_{m,n}^2.$$

Therefore,

$$\bar{\varsigma}_{m,2,4}^X \le \sqrt{r}B_{m,n} \le \sqrt{m}B_{m,n}.$$

By Condition (b) and Lemma 2.2.2 of Van der Waart and Wellner [14],

$$\|M_{m,2,4}^X\|_{\psi_{1/2}} \le C\log^2(rd)\max_{i,a}\Big\|\Big\{\mathrm{vec}\Big(h(X_{4i+1}, X_{4i+2})\Big)\Big\}_a\Big\|_{\psi_{1/2}}^2 \le CB_{m,n}^2\log(md)^2.$$

This bound together with condition (e) yield the following facts.

$$\begin{aligned}
&\mathbb{E}\bar{Z}_{m,2,4}^X \le C\Big\{(mB_{m,n}^2\log(d))^{1/2} + B_n^2(\log(d))\log(md)^2\Big\} \le 2C(mB_{m,n}^2\log(md))^{1/2},\\
&\mathbb{P}\Big(\|\tilde{\Gamma}_{m,2,4}^{JK}\|_\infty \ge C(m^{-1}B_{m,n}^2\log(md))^{1/2} + t)\Big)\\
&\le \exp\Big(-\frac{(mt)^2}{C3mB_{m,n}^2}\Big) + 3\exp\Big[-\Big(\frac{mt}{CB_{m,n}^2\log^2(md)}\Big)^{1/2}\Big],\\
&\le \exp\Big(-\frac{(mt)^2}{C3mB_{m,n}^2}\Big) + 3\exp\Big[-\Big(\frac{\sqrt{mt}}{CB_{m,n}\log(md)}\Big)\Big].
\end{aligned}$$

Let

$$t = C\sqrt{\frac{B_{m,n}^2\log(md)\log^2(\frac{1}{\gamma_{m,n}})}{m}}. \tag{8.8}$$

Apply the above bound with this $t$ to obtain that

$$\mathbb{P}\Big(\|\tilde{\Gamma}_{m,2,4}^{JK}\|_\infty \ge 2t\Big) \le \exp\Big(-C\log(md)\log^2\big(\frac{1}{\gamma_{m,n}}\big)\Big) + 3\exp\Big(-Cn^{1/4}\log^{\frac{1}{2}}(1/\gamma_{m,n})\log^{-\frac{3}{4}}(md)B_{m,n}^{-\frac{1}{2}}\Big). \tag{8.9}$$

Since $0 < \gamma_{m,n} < e^{-1}$, $\log(1/\gamma_{m,n}) > 1$ and $\log(md) > 1$. This fact and (e) reduces (8.9) to

$$\mathbb{P}\Big(\|\tilde{\Gamma}_{m,2,4}^{JK}\|_\infty \ge 2t\Big) \le 4\gamma_{m,n}^C \le 4\gamma_{m,n}. \tag{8.10}$$

Therefore,

$$\mathbb{P}\left(\|\tilde{\Gamma}_{m,2,4}^{JK}\|_\infty^{1/2} \ge C\left(\frac{B_{m,n}^2\log(md)\log^2(\frac{1}{\gamma_{m,n}})}{m}\right)^{1/4}\right) \le 4\gamma_{m,n}$$

which implies that,

$$\mathbb{P}\left(\|\tilde{\Gamma}_{m,2,4}^{JK}\|_\infty^{1/2}\log(md) \ge C\varpi_m^{BX}(\gamma_{m,n})\right) \le 4\gamma_{m,n}, \tag{8.11}$$

where $\varpi_m^{BX}$ is defined as in (8.6). In the rest of this proof, the $t$ is as in (8.8), unless specified otherwise and arguing as for (8.10), for the other terms in (8.7), to obtain the following bounds.

$$\mathbb{P}\Big(\frac{(m-2)}{m(m-1)}\|\tilde{\Gamma}_{m,2,3}^{JK}\|_\infty \ge 2t\Big) \le 4\gamma_{m,n} \quad \text{and} \quad \mathbb{P}\Big(\frac{1}{m(m-1)}\|\tilde{\Gamma}_{m,2,2}^{JK}\|_\infty \ge 2t\Big) \le 4\gamma_{m,n}. \tag{8.12}$$

Combine (8.11) and (8.12) with (8.7) we obtain that

$$\mathbb{P}\Big(\|\tilde{\Gamma}_{m2}^{JK}\|_\infty^{1/2}\log(md) \ge C\varpi_m^{BX}(\gamma_{m,n})\Big) \le 12\gamma_{m,n}. \tag{8.13}$$

*Rate bound for* $\tilde{\Gamma}^{JK}_{m1} - \Gamma^X$. Towards this goal, let

$$\tilde{\Gamma}^{JK}_{m,1,2} := \frac{(m-2)!}{m!} \sum_{1\le i\neq j\le n} \{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_i,X_j))\}^T,$$
$$\tilde{\Gamma}^{JK}_{m,1,3} := \frac{(m-3)!}{m!} \sum_{1\le i\neq j\neq k\le m} \{\mathrm{vec}(h(X_i,X_j))\}\{\mathrm{vec}(h(X_i,X_k))\}^T.$$

Then, we have the decomposition

$$\tilde{\Gamma}^{JK}_{m1} = \frac{1}{m}\,\tilde{\Gamma}^{JK}_{m,1,2} + \frac{(m-2)}{m}\,\tilde{\Gamma}^{JK}_{m,1,3}.$$

Write $\Gamma^X = \mathbb{E}\Big[\{\mathrm{vec}(h(X_1,X_2))\}\{\mathrm{vec}(h(X_1,X_3))\}^T\Big]$ and define

$$\Gamma^X_{1,2} := \mathbb{E}\Big[\{\mathrm{vec}(h(X_1,X_2))\}\{\mathrm{vec}(h(X_1,X_2))\}^T\Big].$$

The entities $\tilde{\Gamma}^{JK}_{m,1,3} - \Gamma^X$ and $\Gamma^{JK}_{m,1,2} - \Gamma^X_{1,2}$ are U statistics of degree three and two respectively. By using arguments similar to those used for $\tilde{\Gamma}^{JK}_{m,2,4}$ for U statistics of degree three and two, we obtain that

$$\mathbb{P}\Big(\Big(\|\tilde{\Gamma}^{JK}_{m,1,3} - \Gamma^X\|_\infty \log^2(md)\Big)^{1/2} \ge C\varpi^{BX}_m(\gamma_{m,n})\Big) \le 4\gamma_{m,n}, \tag{8.14}$$
$$\mathbb{P}\Big(\Big(\|\tilde{\Gamma}^{JK}_{m,1,2} - \Gamma^X_{1,2}\|_\infty \log^2(md)\Big)^{1/2} \ge C\varpi^{BX}_m(\gamma_{m,n})\Big) \le 4\gamma_{m,n}.$$

By the Cauchy-Schwarz and Lyapounov's inequalities and condition (b), $\|\Gamma^X_{1,2}\|_\infty \le B^{2/3}_{m,n}$, from which we obtain that $m^{-1}\|\Gamma^X_{1,2}\|_\infty \le t/2$. Hence, by the triangle inequality, $\|\tilde{\Gamma}^{JK}_{m,1,2}\|_\infty \le \|\tilde{\Gamma}^{JK}_{m,1,2} - \Gamma^X_{1,2}\|_\infty + \|\Gamma^X_{1,2}\|_\infty$, $\Big\{m^{-1}\|\tilde{\Gamma}^{JK}_{m,1,2}\|_\infty \ge \frac{3t}{2}\Big\} \subseteq \Big\{m^{-1}\|\tilde{\Gamma}^{JK}_{m,1,2} - \Gamma^X_{1,2}\|_\infty \ge t\Big\}$, and by (8.14),

$$\mathbb{P}\Big(m^{-1}\|\tilde{\Gamma}^{JK}_{m,1,2}\|_\infty \ge \frac{3t}{2}\Big) \le \mathbb{P}\Big(m^{-1}\|\tilde{\Gamma}^{JK}_{m,1,2} - \Gamma^X_{1,2}\|_\infty \ge t\Big) \le 4\gamma_{m,n}.$$

Finally by (8.13) and (8.14),

$$\mathbb{P}\Big(\|\tilde{\Gamma}^{JK}_m - \Gamma^X\|^{1/2}_\infty \log(md) \ge C\varpi^{BX}_1(\gamma_{m,n})\Big) \le 28\gamma_{m,n}.$$

Similarly for ${Y_1}^n$ we have the decomposition $\tilde{\Gamma}^{JK}_n = \tilde{\Gamma}^{JK}_{n1} - \tilde{\Gamma}^{JK}_{n2}$. Recall the definition of $\Gamma^Y$ from (3.2). By using the similar arguments as above,

$$\mathbb{P}\Big(\|\tilde{\Gamma}^{JK}_n - \Gamma^Y\|^{1/2}_\infty \log(nd) \ge C\varpi^{BY}_n(\gamma_{m,n})\Big) \le 28\gamma_{m,n}.$$

Now choose

$$\bar{\Delta}_{m,n} = C\left(\sqrt{\frac{B^2_{m,n}\log(md)\log^2(\frac{1}{\gamma_{m,n}})}{m}} + \sqrt{\frac{B^2_{m,n}\log(nd)\log^2(\frac{1}{\gamma_{m,n}})}{n}}\right).$$

Combining all the previous inequalities with this choice of $\bar{\Delta}_{m,n}$, it readily follows that

$$\begin{aligned}
\mathbb{P}(\hat{\Delta}_{m,n} \le \bar{\Delta}_{m,n}) &= \mathbb{P}\Big(\|\tilde{\Gamma}^{JK}_m - \Gamma^X + \delta^2_{m,n}(\tilde{\Gamma}^{JK}_n - \Gamma^Y)\|_\infty \le \bar{\Delta}_{m,n}\Big)\\
&= 1 - \mathbb{P}\Big(\|\tilde{\Gamma}^{JK}_m - \Gamma^X + \delta^2_{m,n}(\tilde{\Gamma}^{JK}_n - \Gamma^Y\|_\infty > \bar{\Delta}_{m,n}\Big)\\
&\ge 1 - \Big\{\mathbb{P}\Big(\|\tilde{\Gamma}^{JK}_m - \Gamma^X\|^{1/2}_\infty \log(md) \ge \frac{C}{2}\varpi^{BX}_1(\gamma_{m,n})\Big)\\
&\quad + \mathbb{P}\Big(\|\tilde{\Gamma}^{JK}_n - \Gamma^Y\|^{1/2}_\infty \log(nd) \ge \frac{C}{2}\varpi^{BY}_1(\gamma_{m,n})\Big)\Big\}\\
&\ge 1 - \Big(28\gamma_{m,n} + 28\gamma_{m,n}\Big) = 1 - 56\gamma_{m,n}.
\end{aligned}$$

Thus the final conclusion follows from Lemma 3.1, by setting $W_m^{eX}|X_1^m = Z_1^X|X^m$ and $W_n^{eY}|Y_1^n = Z_2^Y|Y^n$.
This also completes the proof of Theorem 3.2.

**Proof of Theorem 4.1.** The proof uses the results of the previous section with $\delta_{m,n} = -m^{1/2}n^{-1/2}$.
From condition (a′), we readily obtain the bounds

$$\Big\{\frac{c_1}{(1-c_1)}\Big\}^{1/2} = \delta_1 < |\delta_{m,n}| < \delta_2 = \Big\{\frac{c_2}{(1-c_2)}\Big\}^{1/2}, \tag{8.15}$$

It follows from assumption (b′) that

$$\min_{1\le a\le d} \mathbb{E}[g_a^2(X) + \delta_{m,n}^2 g_a^2(Y)] \ge \min\{1, \delta_1^2\} b. \tag{8.16}$$

Recall that, $T_{m,n}^G = T_m^{G_1} + \delta_{m,n} T_n^{G_2}$. By combining (8.15), (8.16), (a′), (e′) along with Theorem 3.1 with $\Omega^X = \text{vec}(\Sigma_1), \Omega^Y = \text{vec}(\Sigma_2)$ we obtain that

$$KD(T_{m,n}, T_{m,n}^G) \le \rho_{m,n}^{**} \lesssim \Big(\frac{B_{m,n}^2(\log^7(dm))}{m}\Big)^{1/6}, \tag{8.17}$$

where

$$KD(T_{m,n}, T_{m,n}^G) := \sup_{t\ge 0}\Big|P\Big(\Big\|\sqrt{m}(U_m^{\xi} - U_n^Y) - \sqrt{m}\,\text{vec}(\Sigma_1 - \Sigma_2)\Big\|_\infty \le 2t\Big) - \mathbb{P}\Big(\|T_m^{G_1} + \delta_{m,n}T_n^{G_2}\|_\infty \le t\Big)\Big|.$$

Choose $\gamma_{m,n}$ in condition (e) of Theorem 3.2 to be $\gamma_{m,n} = 1/(m(\log m)^2)$. From (a′) and (e′) it can be easily verified that

$$\frac{B_{m,n}^2 \log^5(dm)\log^2(1/\gamma_{m,n})}{m} \sim \frac{B_{m,n}^2 \log^5(dn)\log^2(1/\gamma_{m,n})}{n} \to 0, \quad \text{as } m,n\to\infty. \tag{8.18}$$

Combining (8.15), (8.16), (8.18) and Theorem 3.2 we conclude that,

$$KD(T_{m,n}^{JK}, T_{m,n}^G) \le \rho_{m,n}^{JK} \lesssim \Big\{\frac{B_{m,n}^2 \log^5(dm)\log^2(1/\gamma_{m,n})}{m}\Big\}^{1/4},$$

where

$$KD\Big(T_{m,n}^G, T_{m,n}^{JK}\Big) = \sup_{t\ge 0}\Big|\mathbb{P}_{\boldsymbol{e}}\Big(\|W_m^{eX} + \delta_{m,n}W_n^{eY}\|_\infty \le t\Big) - \mathbb{P}\Big(\|T_m^{G_1} + \delta_{m,n}T_n^{G_2}\|_\infty \le t\Big)\Big|.$$

The claim of the theorem follows from the triangle inequality and ($e'$), since

$$KD(T_{m,n}, T_{m,n}^{JK}) \le KD(T_{m,n}, T_{m,n}^G) + KD(T_{m,n}^G, T_{m,n}^{JK}).$$

This concludes the proof.

**Proof of Corollary 4.1.** The proof is an immediate consequence of Theorem 4.1 with the definition of $c_B(\alpha)$.

**Proof of Theorem 5.1.** In the proof below, $C$ and $c^*$ are positive and large enough universal constants, not depending on $m, n, d$, whose values keep changing depending on the context. We shall approximate the power function of the test with its bootstrap counterpart and prove the consistency of the proposed test. Recall $\mathcal{D} := \Sigma_1 - \Sigma_2$. Under $H_{alt}$,

$$\begin{aligned}
&\mathbb{P}\Big(\Big\|\frac{\sqrt{m}(U_m^X - U_n^Y)}{2}\Big\|_\infty \ge c_B(\alpha)\Big)\\
&\ge \mathbb{P}_{e^*}\Big(\Big\|W_m^{e^*X} - \sqrt{\frac{m}{n}}W_n^{e^*Y}\Big\|_\infty \le \Big\|\frac{\sqrt{m}\text{vec}(\mathcal{D})}{2}\Big\|_\infty - c_B(\alpha)\Big)\\
&\quad - \sup_{t\ge 0}\Big|\mathbb{P}\Big(\Big\|\frac{\sqrt{m}(U_m^X - U_n^Y) - \sqrt{m}\text{vec}(\mathcal{D})}{2}\Big\|_\infty \le t\Big) - \mathbb{P}_{e^*}\Big(\Big\|W_m^{e^*X} - \sqrt{\frac{m}{n}}W_n^{e^*Y}\Big\|_\infty \le t\Big)\Big|.
\end{aligned}$$

By arguing as in Theorem 4.1, we obtain that with probability tending to one, under $H_{alt}$,

$$\sup_{t\geq 0}\Big|\mathbb{P}\Big(\Big\|\frac{\sqrt{m}(U_m^X - U_n^Y) - \sqrt{m}\text{vec}(\mathcal{D})}{2}\Big\|_\infty \leq t\Big) - \mathbb{P}_{e^*}\Big(\Big\|W_m^{e^*X} - \sqrt{\frac{m}{n}}W_n^{e^*Y}\Big\|_\infty \leq t\Big)\Big| \lesssim \{B_{m,n}^2\log^7(dm)/m\}^{1/6}. \quad (8.19)$$

Next, we shall prove the consistency of the proposed test. Let $\eta_a, a = 1, 2, \cdots, d$ denote unit basis vectors in $\mathbb{R}^d$. Then, for any $t > 0$,

$$\mathbb{P}_{e^*}\Big(\|W_m^{e^*X} - \sqrt{\frac{m}{n}}W_n^{e^*Y}\|_\infty \geq t\Big) \leq \sum_{a=1}^{d}\mathbb{P}_{e^*}\Big(|T_{ma}^{e^*X} - \sqrt{\frac{m}{n}}T_{na}^{e^*Y}| \geq t\Big) \leq 2d\exp\Big[-\frac{t^2}{2\max_{1\leq a\leq d}\{\eta_a^T(\tilde{\Gamma}_m^{JK} + \frac{m}{n}\tilde{\Gamma}_n^{JK})\eta_a\}}\Big].$$

The last bound follows from Hoeffding's inequality for Gaussian variables. Now setting the above bound equal to $\alpha$ by plugging in $t = c_B(\alpha)$, for large enough $m$, we obtain that

$$c_B(\alpha) \leq \Big[2\log(2d/\alpha)\max_{1\leq a\leq d}\Big\{\eta_a^T(\tilde{\Gamma}_m^{JK} + \frac{m}{n}\tilde{\Gamma}_n^{JK})\eta_a\Big\}\Big]^{1/2} \leq \Big[4\log(dn)\max_{1\leq a\leq d}\Big\{\eta_a^T(\tilde{\Gamma}_m^{JK} + \frac{m}{n}\tilde{\Gamma}_n^{JK})\eta_a\Big\}\Big].$$

But,

$$\max_{1\leq a\leq d}\Big\{\eta_a^T(\tilde{\Gamma}_m^{JK} + \frac{m}{n}\tilde{\Gamma}_n^{JK})\eta_a\Big\} = \Big\|\tilde{\Gamma}_m^{JK} + \frac{m}{n}\tilde{\Gamma}_n^{JK}\Big\|_\infty \leq \Big\|\tilde{\Gamma}_m^{JK} - \Gamma^X + \frac{m}{n}(\tilde{\Gamma}_n^{JK} - \Gamma^Y)\Big\|_\infty + \Big\|\Gamma^X + \frac{m}{n}\Gamma^Y\Big\|_\infty.$$

From the bounds of $\hat{\Delta}_{m,n}$ in Theorem 3.2 with $\delta_{m,n} = \sqrt{m/n}$ and $\gamma_{m,n} = 1/(dm)$, it follows with probability tending to one that

$$\|\tilde{\Gamma}_m^{JK} - \Gamma^X + \frac{m}{n}(\tilde{\Gamma}_n^{JK} - \Gamma^Y)\|_\infty \lesssim \sqrt{\frac{B_{m,n}^2\log^3(md)}{m}}.$$

For the term, $\|\Gamma^X + \frac{m}{n}\Gamma^Y\|_\infty$, we note that, $m/n \leq c^*$ by the condition $(a')$ and using Holder's inequality and (c), it follows that

$$\begin{aligned}\|\Gamma^X + \frac{m}{n}\Gamma^Y\|_\infty \leq C\Big[&\max_{1\leq a_1\leq d}\Big\{\mathbb{E}(\text{vec}(h(X_1,X_2))_{a_1})^2\Big\}^{1/2}\max_{1\leq a_2\leq d}\Big\{\mathbb{E}(\text{vec}(h(X_1,X_3)_{a_2})^2\Big\}^{1/2}\\ &+ c^*\max_{1\leq a_1\leq d}\Big\{\mathbb{E}(\text{vec}(h(Y_1,Y_2)_{a_1})^2\Big\}^{1/2}\max_{1\leq a_2\leq d}\Big\{\mathbb{E}(\text{vec}(h(Y_1,Y_3)_{a_2})^2\Big\}^{1/2}\Big]\\ \leq B_{m,n}^{2/3} \leq B_{m,n}.&\end{aligned}$$

Therefore, with probability tending to one, we obtain that,

$$c_B(\alpha) \leq 4\log(dn)\max_{1\leq a\leq d}\{\eta_a^T(\tilde{\Gamma}_m^{JK} + \frac{m}{n}\tilde{\Gamma}_n^{JK})\eta_a\} \leq (8CB_{m,n}\log(dn))^{1/2}.$$

Upon choosing the constant $C$ in (f′) to be $K = \sqrt{8C}$, we obtain that

$$\big\|\sqrt{m}\text{vec}(\mathcal{D})/2\big\|_\infty - c_B(\alpha) \geq \{8CB_{m,n}\log(dm)\}^{1/2}.$$

Therefore, we conclude that as $m \wedge n \to \infty$ and $d \to \infty$, with probability tending to one,

$$\begin{aligned}&\mathbb{P}_{e^*}\Big(\|W_m^{e^*X} - \sqrt{\frac{m}{n}}W_n^{e^*Y} + \frac{\sqrt{m}\text{vec}(\mathcal{D})}{2}\|_\infty \geq c_B(\alpha)\Big)\\ &\geq \mathbb{P}_{e^*}\Big(\|W_m^{e^*X} - \sqrt{\frac{m}{n}}W_n^{e^*Y}\|_\infty \leq \|\frac{\sqrt{m}\text{vec}(\mathcal{D})}{2}\|_\infty - c_B(\alpha)\Big)\\ &\geq \mathbb{P}_{e^*}\Big(\|W_m^{e^*X} - \sqrt{\frac{m}{n}}W_n^{e^*Y}\|_\infty \leq \{8CB_{m,n}\log(dm)\}^{1/2}\Big)\\ &= 1 - \mathbb{P}_{e^*}\Big(\|W_m^{e^*X} - \sqrt{\frac{m}{n}}W_n^{e^*Y}\|_\infty \geq \{8CB_{m,n}\log(dm)\}^{1/2}\Big)\\ &\geq 1 - 2d\exp\Big(-4CB_{m,n}\log(dm)/B_{m,n}\Big) \geq 1 - \frac{2}{m} \to 1 \text{ as m,n,d} \to \infty.\end{aligned}$$

This bound together with (8.19) completes the proof.

## References


[1] T. Anderson, An introduction to statistical multivariate analysis, John-Wiley, 2003.

[2] T. Cai, W. Liu, Y. Xia, Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings, Journal of the American Statistical Association 108 (2013) 265–277.

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

[4] X. Chen, Gaussian and bootstrap approximations for high-dimensional u-statistics and their applications, The Annals of Statistics 46 (2018) 642–678.

[5] V. Chernozhukov, D. Chetverikov, K. Kato, Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors, The Annals of Statistics 41 (2013) 2786–2819.

[6] V. Chernozhukov, D. Chetverikov, K. Kato, Comparison and anti-concentration bounds for maxima of gaussian random vectors, Probability Theory and Related Fields 162 (2015) 47–70.

[7] V. Chernozhukov, D. Chetverikov, K. Kato, Central limit theorems and bootstrap in high dimensions, The Annals of Probability 45 (2017) 2309–2352.

[8] V. Chernozhuokov, D. Chetverikov, K. Kato, Y. Koike, Improved central limit theorem and bootstrap approximations in high dimensions, The Annals of Statistics 50 (2022) 2562–2586.

[9] J. Fan, Y. Liao, J. Yao, Power enhancement in high-dimensional cross-sectional tests, Econometrica 83 (2015) 1497–1541.

[10] J. He, S. X. Chen, High-dimensional two-sample covariance matrix testing via super-diagonals, Statistica Sinica 28 (2018) 2671–2696.

[11] J. Li, S. X. Chen, Two sample tests for high-dimensional covariance matrices, The Annals of Statistics 40 (2012) 908–940.

[12] J. R. Schott, A test for the equality of covariance matrices when the dimension is large relative to the sample sizes, Computational Statistics & Data Analysis 51 (2007) 6535–6542.

[13] M. S. Srivastava, H. Yanagihara, Testing the equality of several covariance matrices with fewer observations than the dimension, Journal of Multivariate Analysis 101 (2010) 1319–1329.

[14] A. W. Vaart, J. A. Wellner, Weak convergence, in: Weak convergence and empirical processes, Springer, 1996, pp. 16–28.

[15] K. Xue, F. Yao, Distribution and correlation-free two-sample test of high-dimensional means, The Annals of Statistics 48 (2020) 1304–1328.

[16] C. Zhou, F. Han, X.-S. Zhang, H. Liu, An extreme-value approach for testing the equality of large U-statistic based correlation matrices, Bernoulli 25 (2019) 1472 – 1503.