High Dimensional Behaviour of Some Two-Sample Tests Based on Ball Divergence

Bilol Banerjee and Anil K. Ghosh

Indian Statistical Institute, Kolkata

Supplementary Material

This supplementary material has two parts. Appendix A contains proofs of the main theorems, lemmas and propositions stated in the manuscript. It also contains some new lemmas. Appendix B contains some numerical results on strongly spiked eigenvalue (SSE) models.

Appendix A: Proofs and Mathematical details

Proof of Lemma 2.1 . Consider a random permutations π of $\{1, 2, \ldots, n+m\}$ and let $T^{\pi}_{n,m}$ be the permuted test statistic (permutation analogs of $T_{n,m}$). Let $c_{1-\alpha}$ be the upper α-th quantile of the distribution of $T^{\pi}_{n,m}$ given the pooled data $\mathcal{U} := \{X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m\}$. $c_{1-\alpha}$ is permutation invariant and under $H_0 : F = G$, $(T_{n,m}, c_{1-\alpha})$ and $(T^{\pi}_{n,m}, c_{1-\alpha})$ are identically distributed, irrespective of the values of n, m and d. Hence, we have

$$\mathbb{P}[p_{n,m} < \alpha] = \mathbb{P}[T_{n,m} > c_{1-\alpha}] = \mathbb{P}\left[T^{\pi}_{n,m} > c_{1-\alpha}\right]$$
$$= \mathbb{E}\left[\mathbb{P}\left\{T^{\pi}_{n,m} > c_{1-\alpha} \mid \mathcal{U}\right\}\right] \leq \mathbb{E}\left[\alpha\right] \leq \alpha$$

The second last inequality follows from the definition of $c_{1-\alpha}$. Hence the level of the permutation tests are controlled at a desired level α in HDLSS scenario and when n, m and d all simultaneously diverge to infinity. $\square$

Lemma A.1. *If* $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{iid}{\sim} F$ *and* $\mathbf{Y}_1, \ldots, \mathbf{Y}_m \overset{iid}{\sim} G$ *are independent, then* $\mathbb{E}\{T^{\rho}_{n,m}\} = \frac{1}{6}\left(\frac{1}{n-2} + \frac{1}{m-2}\right) + \frac{1}{m}(p_0 - p_1) + \frac{1}{n}(p_2 - p_3) + \Theta^2_{\rho}(F,G)$, *where* $\Theta^2_{\rho}(F,G)$ *is the ball divergence measure defined in Section 2,*

$$p_0 = \mathbb{P}\big\{\rho(Y_1, X_1) \le \rho(X_2, X_1)\big\}$$

$$p_1 = \mathbb{P}\big\{\rho(Y_1, X_1) \le \rho(X_2, X_1); \rho(Y_2, X_1) \le \rho(X_2, X_1)\big\},$$

$$p_2 = \mathbb{P}\big\{\rho(X_1, Y_1) \le \rho(Y_2, Y_1)\big\} \quad \textit{and}$$

$$p_3 = \mathbb{P}\big\{\rho(X_1, Y_1) \le \rho(Y_2, Y_1); \rho(X_2, Y_1) \le \rho(Y_2, Y_1)\big\}.$$

Proof. Note that $T^{\rho}_{n,m}$ can be written as $T^{\rho}_{n,m} = V_1 + V_2$, where

$$V_1 = \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} \left\{ \frac{1}{n-2} \sum_{k=1, k \ne i,j}^{n} \delta(\mathbf{X}_k, \mathbf{X}_j, \mathbf{X}_i) - \frac{1}{m} \sum_{k=1}^{m} \delta(\mathbf{Y}_k, \mathbf{X}_j, \mathbf{X}_i) \right\}^2,$$

$$V_2 = \frac{1}{m(m-1)} \sum_{1 \le i \ne j \le m} \left\{ \frac{1}{n} \sum_{k=1}^{n} \delta(\mathbf{X}_k, \mathbf{Y}_j, \mathbf{Y}_i) - \frac{1}{m-2} \sum_{k=1, k \ne i,j}^{m} \delta(\mathbf{Y}_k, \mathbf{Y}_j, \mathbf{Y}_i) \right\}^2,$$

and $\delta(\mathbf{s}, \mathbf{u}, \mathbf{v}) = \mathbb{1}[\rho(\mathbf{s}, \mathbf{v}) \le \rho(\mathbf{u}, \mathbf{v})]$.

Therefore, we have

$$\mathbb{E}\{V_1\} = \frac{1}{(n-2)^2} \mathbb{E}\Big\{ \big(\sum_{k=3}^{n} \delta(\mathbf{X}_k, \mathbf{X}_2, \mathbf{X}_1) \big)^2 \Big\} + \frac{1}{m^2} \mathbb{E}\Big\{ \big(\sum_{k=1}^{m} \delta(\mathbf{Y}_k, \mathbf{X}_2, \mathbf{X}_1) \big)^2 \Big\}$$
$$- \frac{2}{m(n-2)} \mathbb{E}\Big\{ \big(\sum_{k=3}^{n} \delta(\mathbf{X}_k, \mathbf{X}_2, \mathbf{X}_1) \big) \big(\sum_{k=1}^{m} \delta(\mathbf{Y}_k, \mathbf{X}_2, \mathbf{X}_1) \big) \Big\}$$

$$
\begin{aligned}
&= \frac{1}{(n-2)^2}\mathbb{E}\Big\{\sum_{k=3}^{n}\delta(\mathbf{X}_k,\mathbf{X}_2,\mathbf{X}_1) + \sum_{k,l=3,k\neq l}^{n}\delta(\mathbf{X}_k,\mathbf{X}_2,\mathbf{X}_1)\delta(\mathbf{X}_l,\mathbf{X}_2,\mathbf{X}_1)\Big\} \\
&\quad + \frac{1}{m^2}\mathbb{E}\Big\{\sum_{k=3}^{m}\delta(\mathbf{Y}_k,\mathbf{X}_2,\mathbf{X}_1) + \sum_{k,l=3,k\neq l}^{m}\delta(\mathbf{Y}_k,\mathbf{X}_2,\mathbf{X}_1)\delta(\mathbf{Y}_l,\mathbf{X}_2,\mathbf{X}_1)\Big\} \\
&\quad - \frac{2}{m(n-2)}\mathbb{E}\Big\{\sum_{k=3}^{n}\sum_{l=1}^{m}\delta(\mathbf{X}_k,\mathbf{X}_2,\mathbf{X}_1)\delta(\mathbf{Y}_l,\mathbf{X}_2,\mathbf{X}_1)\Big\} \\
&= \frac{1}{n-2}\ \mathbb{P}\Big\{\rho(\mathbf{X}_3,\mathbf{X}_1)\le\rho(\mathbf{X}_2,\mathbf{X}_1)\Big\} \\
&\quad + \frac{n-3}{n-2}\ \mathbb{P}\Big\{\rho(\mathbf{X}_3,\mathbf{X}_1)\le\rho(\mathbf{X}_2,\mathbf{X}_1);\rho(\mathbf{X}_4,\mathbf{X}_1)\le\rho(\mathbf{X}_2,\mathbf{X}_1)\Big\} \\
&\quad + \frac{1}{m}\ \mathbb{P}\Big\{\rho(\mathbf{Y}_1,\mathbf{X}_1)\le\rho(\mathbf{X}_2,\mathbf{X}_1)\Big\} \\
&\quad + \frac{m-1}{m}\ \mathbb{P}\Big\{\rho(\mathbf{Y}_1,\mathbf{X}_1)\le\rho(\mathbf{X}_2,\mathbf{X}_1);\rho(\mathbf{Y}_2,\mathbf{X}_1)\le\rho(\mathbf{X}_2,\mathbf{X}_1)\Big\} \\
&\quad - 2\ \mathbb{P}\Big\{\rho(\mathbf{X}_3,\mathbf{X}_1)\le\rho(\mathbf{X}_2,\mathbf{X}_1);\rho(\mathbf{Y}_1,\mathbf{X}_1)\le\rho(\mathbf{X}_2,\mathbf{X}_1)\Big\} \\
&= \frac{1}{(n-2)}\Big\{\frac{1}{2}+(n-3)\frac{1}{3}\Big\} + \frac{1}{m}\Big\{p_0+(m-1)p_1\Big\} - 2p_4 \\
&= \frac{1}{3} + \frac{1}{6(n-2)} + \frac{1}{m}(p_0-p_1) + p_1 - 2p_4,
\end{aligned}
$$

where $p_4 = \mathbb{P}\{\rho(\mathbf{X}_3,\mathbf{X}_1)\le\rho(\mathbf{X}_2,\mathbf{X}_1);\rho(\mathbf{Y}_1,\mathbf{X}_1)\le\rho(\mathbf{X}_2,\mathbf{X}_1)\}$.

Similarly one can show that

$$
\mathbb{E}\{V_2\} = \frac{1}{3} + \frac{1}{6(m-2)} + \frac{1}{n}(p_2-p_3) + p_3 - 2p_5,
$$

where $p_5 = \mathbb{P}\{\rho(\mathbf{Y}_3,\mathbf{Y}_1)\le\rho(\mathbf{Y}_2,\mathbf{Y}_1);\rho(\mathbf{X}_1,\mathbf{Y}_1)\le\rho(\mathbf{Y}_2,\mathbf{Y}_1)\}$.

Hence, we have

$$
\mathbb{E}\{T^{\rho}_{n,m}\} = \frac{1}{6(n-2)} + \frac{1}{6(m-2)} + \frac{1}{m}(p_0-p_1) + \frac{1}{n}(p_2-p_3) + \frac{2}{3} + p_1 - 2p_4 + p_3 - 2p_5.
$$

Now observe that

$$\begin{aligned}
\Theta_\rho^2(F,G) = & \int \{F(B(u,\rho(v,u))) - G(B(u,\rho(v,u)))\}^2 \left[dF(u)dF(v) + dG(u)dG(v)\right] \\
= & \int F^2(B(u,\rho(v,u))dF(u)dF(v) \\
& - 2\int F(B(u,\rho(v,u))G(B(u,\rho(v,u))dF(u)dF(v) \\
& + \int G^2(B(u,\rho(v,u))dF(u)dF(v) + \int F^2(B(u,\rho(v,u))dG(u)dG(v) \\
& - 2\int F(B(u,\rho(v,u))G(B(u,\rho(v,u))dG(u)dG(v) \\
& + \int G^2(B(u,\rho(v,u))dG(u)dG(v) \\
= & \frac{1}{3} - 2p_4 + p_1 + p_3 - 2p_5 + \frac{1}{3} = p_1 + p_3 - 2p_4 - 2p_5 + \frac{2}{3}.
\end{aligned}$$

Hence, $\mathbb{E}\{T^\rho_{n,m}\} = \frac{1}{6}\left(\frac{1}{n-2} + \frac{1}{m-2}\right) + \frac{1}{m}(p_0 - p_1) + \frac{1}{n}(p_5 - p_3) + \Theta^2_\rho(F,G)$.

This completes the proof. □

Lemma A.2. *If $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n \overset{iid}{\sim} F$ and $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m \overset{iid}{\sim} G$ are independent, then $Var(T^\rho_{n,m}) \leq C_1\Theta^2_\rho(F,G)\left(\frac{1}{n}+\frac{1}{m}\right) + C_2\left(\frac{1}{n}+\frac{1}{m}\right)^2$, where the constants C_1 and C_2 are independent of the dimension d.*

Proof. Note that $\Theta^2_\rho(F,G)$ can be written as $\Theta^2_\rho(F,G) = A_1 + A_2$, where

$$\begin{aligned}
A_1 &= \int\int \{F(\mathbb{B}(\mathbf{u},\rho(\mathbf{v},\mathbf{u})) - G(\mathbb{B}(\mathbf{u},\rho(\mathbf{v},\mathbf{u}))\}^2 dF(\mathbf{u})dF(\mathbf{v}) \\
&= \mathbb{E}\Big(F(B(\mathbf{X}_1,\rho(\mathbf{X}_2,\mathbf{X}_1))) - G(B(\mathbf{X}_1,\rho(\mathbf{X}_2,\mathbf{X}_1)))\Big)^2_, \text{ and}
\end{aligned}$$

$$\begin{aligned}
A_2 &= \int\int \{F(\mathbb{B}(\mathbf{u},\rho(\mathbf{v},\mathbf{u})) - G(\mathbb{B}(\mathbf{u},\rho(\mathbf{v},\mathbf{u}))\}^2 dG(\mathbf{u})dG(\mathbf{v}) \\
&= \mathbb{E}\Big(F(B(\mathbf{Y}_1,\rho(\mathbf{Y}_2,\mathbf{Y}_1))) - G(B(\mathbf{Y}_1,\rho(\mathbf{Y}_2,\mathbf{Y}_1)))\Big)^2.
\end{aligned}$$

It can be verified that V_1 (as defined in the proof of Lemma A.1) can be expressed as

$$V_1 = \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \left\{ \frac{1}{n-2} \sum_{k=1,k\ne i,j}^{n} \delta(\mathbf{X}_k, \mathbf{X}_j, \mathbf{X}_i) - \frac{1}{m} \sum_{k=1}^{m} \delta(\mathbf{Y}_k, \mathbf{X}_j, \mathbf{X}_i) \right\}^2$$

$$= \frac{1}{n(n-1)(n-2)^2 m^2} \sum_{i\ne j=1}^{n} \sum_{u,u'\ne i,j}^{n} \sum_{v,v'=1}^{m} \Big\{ \delta(\mathbf{X}_u, \mathbf{X}_j, \mathbf{X}_i)\delta(\mathbf{X}_{u'}, \mathbf{X}_j, \mathbf{X}_i)$$

$$+ \delta(\mathbf{Y}_v, \mathbf{X}_j, \mathbf{X}_i)\delta(\mathbf{Y}_{v'}, \mathbf{X}_j, \mathbf{X}_i) - \delta(\mathbf{X}_{u'}, \mathbf{X}_j, \mathbf{X}_i)\delta(\mathbf{Y}_v, \mathbf{X}_j, \mathbf{X}_i)$$

$$- \delta(\mathbf{X}_u, \mathbf{X}_j, \mathbf{X}_i)\delta(\mathbf{Y}_{v'}, \mathbf{X}_j, \mathbf{X}_i) \Big\}$$

$$= \frac{1}{n(n-1)(n-2)^2 m^2} \sum_{i\ne j=1}^{n} \sum_{u,u'\ne i,j}^{n} \sum_{v,v'=1}^{m} \psi_{A_1}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_u, \mathbf{X}_{u'}; \mathbf{Y}_v, \mathbf{Y}_{v'}), \text{ say.}$$

Clearly, V_1 can be written as a linear combination of U-statistics of different degrees. Let $\hat{U}_{A_1}^{(4,2)}$ be the U-statistic with the kernel function $\psi_{A_1}(\mathbf{X}_i, \mathbf{X}_j, \mathbf{X}_u, \mathbf{X}_{u'}; \mathbf{Y}_v, \mathbf{Y}_{v'})$, which has the same degree as V_1. So, it determines the order of $Var(V_1)$. More specifically, we get

$$V_1 = \frac{1}{(n-2)^2 m^2} \left\{ 4\binom{n-2}{2}\binom{m}{2} \hat{U}_{A_1}^{(4,2)} \right\} + O_P\left(\left(\frac{1}{n} + \frac{1}{m}\right)\right) \text{ and}$$

$$Var(V_1) = \frac{\sigma_{1,0}^2(A_1)}{n} + \frac{\sigma_{0,1}^2(A_1)}{m} + C_2\left(\left(\frac{1}{n} + \frac{1}{m}\right)^2\right).$$

Note that here $\sigma_{1,0}^2(A_1) = Var(\psi_{A_1,1,0}^s(\mathbf{X}))$ and $\sigma_{0,1}^2(A_1) = Var(\psi_{A_1,0,1}^s(\mathbf{X}))$, where we have $\psi_{A_1,1,0}^s(\mathbf{x}) = E[\psi_{A_1}(\mathbf{x}, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4; \mathbf{Y}_1, \mathbf{Y}_2)]$ and $\psi_{A_1,0,1}^s(\mathbf{y}) = E[\psi_{A_1}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4; \mathbf{y}, \mathbf{Y}_2)]$. Since ψ_{A_1} is uniformly bounded, the constant C_2 does not depend on d.

Now, we have

$$\begin{aligned}
\psi^s_{A_1,1,0}(\mathbf{x}) &= \mathbb{E}\Big\{\frac{1}{4!2!}\sum_{\pi\in S_4}\sum_{\gamma\in S_2}\psi_{A_1}(\mathbf{X}_{\pi(1)},\mathbf{X}_{\pi(2)},\mathbf{X}_{\pi(3)},\mathbf{X}_{\pi(4)};\mathbf{Y}_{\gamma(1)},\mathbf{Y}_{\gamma(2)}) \mid \mathbf{X}_1=\mathbf{x}\Big\}\\
&= \frac{1}{4}\mathbb{E}\Big\{\psi_{A_1}(\mathbf{x},\mathbf{X}_2,\mathbf{X}_3,\mathbf{X}_4;\mathbf{Y}_1,\mathbf{Y}_2)+\psi_{A_1}(\mathbf{X}_i,\mathbf{x},\mathbf{X}_3,\mathbf{X}_4;\mathbf{Y}_1,\mathbf{Y}_2)\\
&\qquad +\psi_{A_1}(\mathbf{X}_i,\mathbf{X}_2,\mathbf{x},\mathbf{X}_4;\mathbf{Y}_1,\mathbf{Y}_2)+\psi_{A_1}(\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3,\mathbf{x};\mathbf{Y}_1,\mathbf{Y}_2)\Big\}\\
&= \frac{1}{4}\left\{\mathbb{E}\Big(F(B(\mathbf{x},\rho(\mathbf{X}_2,\mathbf{x})))-G(B(\mathbf{x},\rho(\mathbf{X}_2,\mathbf{x})))\Big)^2\right\}\\
&\quad+\frac{1}{4}\left\{\mathbb{E}\Big(F(B(\mathbf{X}_1,\rho(\mathbf{x},\mathbf{X}_1)))-G(B(\mathbf{X}_1,\rho(\mathbf{x},\mathbf{X}_1)))\Big)^2\right\}\\
&\quad+\frac{1}{2}\mathbb{E}\Big(\delta(\mathbf{x},\mathbf{X}_2,\mathbf{X}_1)-G(B(\mathbf{X}_1,\rho(\mathbf{X}_2,\mathbf{X}_1)))\Big)\\
&\qquad\quad\times\Big(F(B(\mathbf{X}_1,\rho(\mathbf{X}_2,\mathbf{X}_1)))-G(B(\mathbf{X}_1,\rho(\mathbf{X}_2,\mathbf{X}_1)))\Big)\\
&\quad= g_1(\mathbf{x})+g_2(\mathbf{x})+g_3(\mathbf{x}),\quad\text{say.}
\end{aligned}$$

Therefore, we have $\sigma^2_{1,0}=\mathbb{E}\big\{\psi^s_{A_1,1,0}(\mathbf{X})-\mathbb{E}(\psi^s_{A_1,1,0}(\mathbf{X}))\big\}^2=\mathbb{E}\big\{\psi^s_{A_1,1,0}(\mathbf{X})-A_1\big\}^2=\mathbb{E}\big\{g_1(\mathbf{X})+g_2(\mathbf{X})+g_3(\mathbf{X})-A_1\big\}^2$. So, using the inequality, $E(\sum_{i=1}^p Z_1)^2\le p\ E(\sum_{i=1}^p Z_i^2)$ and the fact that $0\le g_1(\mathbf{x}),g_2(\mathbf{x})\le 1/4$ for all $\mathbf{x}$, we get

$$\begin{aligned}
\sigma^2_{1,0} &\le 4\Big\{\mathbb{E}g_1^2(\mathbf{X})+\mathbb{E}g_2^2(\mathbf{X})+\mathbb{E}g_3^2(\mathbf{X})+A_1^2\Big\}\\
&\le 4\Big\{\frac{1}{4}\mathbb{E}g_1(\mathbf{X})+\frac{1}{4}\mathbb{E}g_2(\mathbf{X})+\mathbb{E}g_3^2(\mathbf{X})+A_1\Big\}.
\end{aligned}$$

Now note that $\mathbb{E}g_1(\mathbf{X})=\mathbb{E}g_2(\mathbf{X})=\frac{1}{4}A_1$. Also, using Cauchy-Schwartz inequality on $g_3(\mathbf{X})$, we get

$$\mathbb{E}g_3^2(\mathbf{X})\le\frac{1}{4}\mathbb{E}\Big(\delta(\mathbf{X},\mathbf{X}_2,\mathbf{X}_1)-G(B(\mathbf{X}_1,\rho(\mathbf{X}_2,\mathbf{X}_1)))\Big)^2A_1\le\frac{1}{4}A_1.$$

Hence, we have $\sigma^2_{1,0}(A_1)\le\frac{11}{2}A_1$. Similarly, we can also show that $\sigma^2_{0,1}(A_1)\le$

$\frac{9}{2}A_1$. Combining these, we get

$$Var(V_1) \leq \frac{11}{2}A_1\Big(\frac{1}{n}+\frac{1}{m}\Big)+C_2\left(\Big(\frac{1}{n}+\frac{1}{m}\Big)^2\right).$$

Using the same set of arguments, we also have

$$Var(V_2) \leq \frac{11}{2}A_2\Big(\frac{1}{n}+\frac{1}{m}\Big)+C_2\left(\Big(\frac{1}{n}+\frac{1}{m}\Big)^2\right).$$

This completes the proof. $\square$

Lemma A.3. *Consider a random permutation π of $\{1,2,\ldots,n+m\}$. If $T^{\rho}_{n,m,\pi}$ denotes the permuted test statistic (the permutation analog of $T^{\rho}_{n,m}$), given the pooled sample $\mathcal{U}=\{\mathbf{U}_1,\mathbf{U}_2,\ldots,\mathbf{U}_{n+m}\}$, the conditional expectation $T^{\rho}_{n,m,\pi}$ is given by $\mathbb{E}\{T^{\rho}_{n,m,\pi}\,|\,\mathcal{U}\}=\frac{1}{6}\left(\frac{1}{n}+\frac{1}{m}+\frac{1}{n-2}+\frac{1}{m-2}\right)$.*

Proof. For any random permutation π of $\{1,2,\ldots,n+m\}$, we have

$$\begin{aligned}
T^{\rho}_{n,m,\pi} = \frac{1}{n(n-1)}\sum_{1\leq i\neq j\leq n}&\left\{\frac{1}{n-2}\sum_{k=1,k\neq i,j}^{n}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(j)},\mathbf{U}_{\pi(i)})-\frac{1}{m}\sum_{k=n+1}^{n+m}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(j)},\mathbf{U}_{\pi(i)})\right\}^2\\
+\frac{1}{m(m-1)}\sum_{n+1\leq i\neq j\leq n+m}&\left\{\frac{1}{n}\sum_{k=1}^{n}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(j)},\mathbf{U}_{\pi(i)})\right.\\
&\left.-\frac{1}{m-2}\sum_{k=n+1,k\neq i,j}^{n+m}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(j)},\mathbf{U}_{\pi(i)})\right\}^2,
\end{aligned}$$

So, the conditional expectation of $T^{\rho}_{n,m,\pi}$ for any given $\mathcal{U}$ is given by

$$\begin{aligned}
\mathbb{E}\{T^{\rho}_{n,m,\pi}\,|\,\mathcal{U}\} = \;&\mathbb{E}\left\{\left\{\frac{1}{n-2}\sum_{k=3}^{n}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})-\frac{1}{m}\sum_{k=n+1}^{n+m}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\right\}^2\Bigg|\;\mathcal{U}\right\}\\
&+\mathbb{E}\left\{\left\{\frac{1}{n}\sum_{k=1}^{n}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})-\frac{1}{m-2}\sum_{k=n+1,k\neq i,j}^{n+m}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\right\}^2\Bigg|\;\mathcal{U}\right\}.
\end{aligned}$$

Now note that

$$\mathbb{E}\left\{\left\{\frac{1}{n-2}\sum_{k=3}^{n}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})-\frac{1}{m}\sum_{k=n+1}^{n+m}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\right\}^2\Big|\ \mathcal{U}\right\}$$

$$=\frac{1}{(n-2)^2}\mathbb{E}\Big\{\sum_{k=3}^{n}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\mid\mathcal{U}\Big\}$$

$$+\frac{1}{(n-2)^2}\sum_{k,l=3,k\neq l}^{n}\mathbb{E}\Big\{\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\delta(\mathbf{U}_{\pi(l)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\mid\mathcal{U}\Big\}$$

$$+\frac{1}{m^2}\sum_{k=n+1}^{n+m}\mathbb{E}\Big\{\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\mid\mathcal{U}\Big\}$$

$$+\frac{1}{m^2}\sum_{k,l=n+1,k\neq l}^{n+m}\mathbb{E}\Big\{\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\delta(\mathbf{U}_{\pi(l)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\mid\mathcal{U}\Big\}$$

$$-\frac{2}{m(n-2)}\sum_{k=3}^{n}\sum_{l=n+1}^{n+m}\mathbb{E}\Big\{\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\delta(\mathbf{U}_{\pi(l)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\mid\mathcal{U}\Big\}$$

$$=\frac{(n-2)q_1}{(n-2)^2}+\frac{(n-2)(n-3)q_2}{(n-2)^2}+\frac{mq_1}{m^2}+\frac{m(m-1)q_2}{m^2}-\frac{2m(n-2)q_2}{m(n-2)}$$

$$=(q_1-q_2)\Big(\frac{1}{n-2}+\frac{1}{m}\Big),$$

where $q_1=\mathbb{E}\{\delta(\mathbf{U}_{\pi(1)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(3)})\mid\mathcal{U}\}$ and $q_2=\mathbb{E}\{\delta(\mathbf{U}_{\pi(1)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(3)})$ $\delta(\mathbf{U}_{\pi(4)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(3)})\mid\mathcal{U}\}$.

Similarly, we can also show that

$$\mathbb{E}\left\{\left\{\frac{1}{n}\sum_{k=1}^{n}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})-\frac{1}{m-2}\sum_{k=n+1,k\neq i,j}^{n+m}\delta(\mathbf{U}_{\pi(k)},\mathbf{U}_{\pi(2)},\mathbf{U}_{\pi(1)})\right\}^2\Big|\ \mathcal{U}\right\}$$
$$=(q_1-q_2)\Big(\frac{1}{m-2}+\frac{1}{n}\Big).$$

But given the pooled sample $\mathcal{U}$, the random variables $\{\mathbf{U}_{\pi(i)}\}_{i=1}^{n+m}$ are exchangeable.

So, we must have $q_1 = 1/2$ and $q_2 = 1/3$. Hence, we have

$$\mathbb{E}\{T^{\rho}_{n,m,\pi} \mid \mathcal{U}\} = \frac{1}{6}\left(\frac{1}{n} + \frac{1}{m} + \frac{1}{n-2} + \frac{1}{m-2}\right)$$

□

Proof of Lemma 2.2. Here, we are interested in the quantiles of the conditional distribution of $T^{\rho}_{n,m,\pi}$ given the pooled sample $\mathcal{U}$. Since $T^{\rho}_{n,m,\pi}$ is non-negative, using Markov's inequality on the conditional random variable, we get

$$\mathbb{P}\left\{T^{\rho}_{n,m,\pi} \geq \frac{1}{\alpha}\mathbb{E}\{T^{\rho}_{n,m,\pi} \mid \mathcal{U}\} \;\middle|\; \mathcal{U}\right\} \leq \alpha.$$

Therefore, from the definition of the quantile $c_{1-\alpha}$, we have $c_{1-\alpha} \leq \frac{1}{\alpha}\mathbb{E}\{T^{\rho}_{n,m,\pi} \mid \mathcal{U}\}$, which holds with probability one. From Lemma A.3, we also have

$$\mathbb{E}\{T^{\rho}_{n,m,\pi} \mid \mathcal{U}\} = \frac{1}{6}\left(\frac{1}{n} + \frac{1}{m} + \frac{1}{n-2} + \frac{1}{m-2}\right) \leq \frac{2}{3(\min\{n,m\}-2)}.$$

This completes the proof. □

Proof of Theorem 2.1. In view of Lemma A.1 and Lemma A.2, as $\min\{n,m\}$ grows to infinity, $T^{\rho}_{n,m}$ converges in probability to $\Theta^2_{\rho}(F,G)$. So, if $\Theta^2_{\rho}(F,G) > 0$, under $H_1 : F \neq G$, $T^{\rho}_{n,m}$ converges in probability to a positive number. On the other hand, Lemma 2.1 shows that the cut-off value of the permutation test $c_{1-\alpha}$ goes to zero almost surely. Therefore, the power of the permutation test converges to one as $\min\{n,m\}$ grows to infinity. □

Proof of Lemma 2.3. To prove this lemma, we shall use the idea of Corollary 6.1 of Kim (2021). First, let us define

$$F(t) = \frac{1}{N!}\left\{\sum_{\pi\in\mathcal{S}_N} \mathbb{1}[\hat{\zeta}_{\pi_i} \le t]\right\} \quad \textit{and} \quad F_B(t) = \frac{1}{B}\left\{\sum_{i=1}^{B} \mathbb{1}[\hat{\zeta}_{\pi_i} \le t]\right\}.$$

where F and F_B are distribution functions conditioned on the observed pooled data $\mathcal{U}$. Then,

$$\begin{aligned}
|p_{n,m} - p_{n,m,B}| &= \left|\frac{1}{N!}\left\{\sum_{\pi\in\mathcal{S}_N} \mathbb{1}[\hat{\zeta}_{\pi_i} \ge \hat{\zeta}_{n,m}]\right\} - \frac{1}{B+1}\left\{\sum_{i=1}^{B} \mathbb{1}[\hat{\zeta}_{\pi_i} \ge \hat{\zeta}_{n,m}] + 1\right\}\right| \\
&= \left|\frac{1}{N!}\left\{\sum_{\pi\in\mathcal{S}_N} \mathbb{1}[\hat{\zeta}_{\pi_i} < \hat{\zeta}_{n,m}]\right\} - \frac{1}{B+1}\left\{\sum_{i=1}^{B} \mathbb{1}[\hat{\zeta}_{\pi_i} < \hat{\zeta}_{n,m}]\right\}\right| \\
&= |F(\hat{\zeta}_{n,m}-) - \frac{B}{B+1}F_B(\hat{\zeta}_{n,m}-)| \\
&\le |F(\hat{\zeta}_{n,m}-) - F_B(\hat{\zeta}_{n,m}-)| + |\frac{F_B(\hat{\zeta}_{n,m}-)}{B+1}| \\
&\le \sup_{t\in\mathbb{R}} |F(t) - F_B(t)| + \frac{1}{B+1}
\end{aligned}$$

Conditioned on the pooled data $\mathcal{U}$, the Dvoretzky-Keifer-Wolfwitz inequality (see, e.g., Massart, 1990) gives us $\mathbb{P}\{\sup_{t\in\mathbb{R}} |F(t) - F_B(t)| > \epsilon\} \le 2e^{-2B\epsilon^2}$. Hence, conditioned on $\mathcal{U}$, as B grows to infinity, the randomized p-value $p_{n,m,B}$ converges almost surely to $p_{n,m}$. □

Proof of Lemma 3.1. For $\mathbf{W} = \mathbf{X}_1 - \mathbf{X}_2, \mathbf{Y}_1 - \mathbf{Y}_2$ or $\mathbf{X}_1 - \mathbf{Y}_1$, under (A1), $\frac{1}{d}\Big|\|W\|^2 - \mathbb{E}(\|W\|^2)\Big| \xrightarrow{P} 0$ as $d \to \infty$. Again under (A2), as $d \to \infty$, $\frac{1}{d}E(\|W\|^2)$ converges to $2\sigma_F^2$, $2\sigma_F^2$ and $\sigma_F^2 + \sigma_G^2 + \nu^2$ in these three cases. spectively, as d grows to infinity. The result follows from these two facts. □

Lemma A.4. *Suppose that* $\mathbf{X}_1, \mathbf{X}_2 \sim F$, $\mathbf{Y}_1, \mathbf{Y}_2 \sim G$ *and they are independent. For a distance function* ρ, *assume that* $\rho(\mathbf{X}_1, \mathbf{X}_2) \xrightarrow{P} \theta_1$, $\rho(\mathbf{Y}_1, \mathbf{Y}_2) \xrightarrow{P} \theta_2$ *and* $\rho(\mathbf{X}_1, \mathbf{Y}_2) \xrightarrow{P} \theta_3$ *as* $d \to \infty$. *If* $\theta_3 > \min\{\theta_1, \theta_2\}$, *then* $P(T^{\rho}_{n,m} > 1/3) \to 1$ *as* d *diverges to infinity.*

Proof. Note that $T^{\rho}_{n,m}$ involves the terms $\delta(\mathbf{U}_k, \mathbf{U}_j, \mathbf{U}_i)$'s for different choices of $\mathbf{U}_i, \mathbf{U}_j, \mathbf{U}_k$ from the pooled sample. So, the behaviour of $T^{\rho}_{n,m}$ can be studied using the convergence of the $\delta(\mathbf{U}_k, \mathbf{U}_j, \mathbf{U}_i)$'s.

First, consider the case $\min\{\theta_1, \theta_2\} < \theta_3 < \max\{\theta_1, \theta_2\}$. Let us assume that $\theta_1 > \theta_3 > \theta_2$. In such a situation, we have $\lim_{d\to\infty} \mathbb{P}[\rho(\mathbf{X}_2, \mathbf{Y}_1) \le \rho(\mathbf{Y}_2, \mathbf{Y}_1)] = 0$ and $\lim_{d\to\infty} \mathbb{P}[\rho(\mathbf{Y}_1, \mathbf{X}_1) \le \rho(\mathbf{X}_2, \mathbf{X}_1)] = 1$. Hence, we get

$$
\begin{aligned}
V_1 &= \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \left\{ \frac{1}{n-2} \sum_{k=1,k\ne i,j}^{n} \delta(\mathbf{X}_k, \mathbf{X}_j, \mathbf{X}_i) - \frac{1}{m} \sum_{k=1}^{m} \delta(\mathbf{Y}_k, \mathbf{X}_j, \mathbf{X}_i) \right\}^2 \\
&\xrightarrow{P} \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \left\{ \frac{1}{n-2} \sum_{k=1,k\ne i,j}^{n} \delta(\mathbf{X}_k, \mathbf{X}_j, \mathbf{X}_i) - 1 \right\}^2 \\
&= \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \left\{ \frac{-1}{n-2} \sum_{k=1,k\ne i,j}^{n} \mathbb{1}[\rho(\mathbf{X}_k, \mathbf{X}_i) > \rho(\mathbf{X}_j, \mathbf{X}_i)] \right\}^2 \\
&= \frac{1}{n(n-1)(n-2)^2} \Bigg\{ \sum_{1\le i\ne j\ne k\le n} \mathbb{1}[\rho(\mathbf{X}_k, \mathbf{X}_i) > \rho(\mathbf{X}_j, \mathbf{X}_i)] \\
&\quad + \sum_{1\le i\ne j\le n} \sum_{k=1,k\ne i,j}^{n} \sum_{l=1,l\ne i,j,k}^{n} \mathbb{1}[\rho(\mathbf{X}_k, \mathbf{X}_i) > \rho(\mathbf{X}_j, \mathbf{X}_i)] \mathbb{1}[\rho(\mathbf{X}_l, \mathbf{X}_i) > \rho(\mathbf{X}_j, \mathbf{X}_i)] \Bigg\} \\
&= \frac{1}{n(n-1)(n-2)^2} \left\{ \binom{n}{1}\binom{n-1}{2} + 2\binom{n}{1}\binom{n-1}{3} \right\} = \frac{1}{3} + \frac{1}{6(n-2)}.
\end{aligned}
$$

Similarly, as d diverges to infinity, we have

$$V_2 \xrightarrow{P} \frac{1}{m(m-1)} \sum_{n+1 \le i \ne j \le n+m} \left\{ \frac{1}{m-2} \sum_{k=1, k \ne i,j}^{m} \delta(\mathbf{Y}_k, \mathbf{Y}_j, \mathbf{Y}_i) \right\}^2 = \frac{1}{3} + \frac{1}{6(m-2)}.$$

Thus, $T^{\rho}_{n,m} \xrightarrow{P} \frac{2}{3} + \frac{1}{6}\left(\frac{1}{n-2} + \frac{1}{m-2}\right)$. The same result holds for $\theta_1 < \theta_3 < \theta_2$ as well.

Now consider the case, $\theta_3 = \max\{\theta_1, \theta_2\}$. Assume that $\theta_2 < \theta_1 = \theta_3$. In this case, the convergence of $\mathbb{P}[\rho(\mathbf{Y}_1, \mathbf{X}_1) \le \rho(\mathbf{X}_2, \mathbf{X}_1)]$ is not clear, but $\mathbb{P}[\rho(\mathbf{X}_1, \mathbf{Y}_1) \le \rho(\mathbf{Y}_2, \mathbf{Y}_1)]$ converges to zero as d diverges to infinity. Hence, V_2 converges to $1/3 + 1/6(m-2)$ in probability, and V_1 converges in probability to a non-negative random variable. Therefore, $P(T^{\rho}_{n,m} > 1/3)$ converges to one. Similar arguments can be given for $\theta_1 < \theta_2 = \theta_3$ as well.

Finally, consider the case $\theta_3 > \max\{\theta_1, \theta_2\}$. In this case, we have $\lim_{d\to\infty} \mathbb{P}[\rho(\mathbf{X}_2, \mathbf{Y}_1) \le \rho(\mathbf{Y}_2, \mathbf{Y}_1)] = 0$ and $\lim_{d\to\infty} \mathbb{P}[\rho(\mathbf{Y}_1, \mathbf{X}_1) \le \rho(\mathbf{X}_2, \mathbf{X}_1)] = 0$. Hence as d diverges to infinity, V_1 and V_2 converge in probability to $1/3 + 1/6(n-2)$ and $1/3 + 1/6(m-2)$, respectively. Thus, $T^{\rho}_{n,m}$ converges in probability to $2/3 + 1/6\,(1/(n-2) + 1/(m-2))$.

These three cases together imply that if $\theta_1 > \min\{\theta_1, \theta_2\}$, $P(T^{\rho}_{n,m} > 1/3) \to 1$ as $d \to \infty$. $\square$

Proof of Lemma 3.2. Here, we have $d^{-1/2}\|\mathbf{X}_1 - \mathbf{X}_2\| \xrightarrow{P} \sigma_F\sqrt{2}$, $d^{-1/2}\|\mathbf{Y}_1 - \mathbf{Y}_2\| \xrightarrow{P} \sigma_G\sqrt{2}$ and $d^{-1/2}\|\mathbf{X}_1 - \mathbf{Y}_1\| \xrightarrow{P} \sqrt{\sigma_F^2 + \sigma_G^2 + \nu^2}$ as $d \to \infty$ (see

Lemma 3.1). Let these limiting values be denoted by θ_1, θ_2, and θ_3, respectively. If $\nu^2 + (\sigma_F - \sigma_G)^2 > 0$, one can check that $\theta_3 > \sqrt{\theta_1\theta_2} \geq \min\{\theta_1, \theta_2\}$. Hence the proof follows from Lemma A.4. $\square$

Proof of Theorem 3.1. It follows from Lemma A.4 that under the condition $\nu^2+(\sigma_F-\sigma_G)^2 > 0$, $P(T^{\ell_2}_{n,m} > 1/3)$ converges to 1 as d tends to infinity. We have also seen that the cut-off of the permutation test $c_{1-\alpha}$ has an upper bound $2/\{3\alpha(\min\{n, m\} - 2)\}$, which does not depend on the dimension d. Therefore, if $\min\{n, m\} \geq 2 + 2/\alpha$, the test based on $T^{\ell_2}_{n,m}$ rejects H_0 with probability tending to 1 as d grows to infinity. $\square$

Proof of Lemma 3.3. This lemma is taken from Sarkar and Ghosh (2018). The proof can be found on page 5 (see Lemma 1) of that article. $\square$

Proof of Theorem 3.2 . We use a sub-sequence argument to prove this theorem. Let $\{d_k\}$ be an arbitrary sub-sequence of the sequence of natural numbers. Under (A4) and $\liminf_{d\to\infty} e_{h,\psi}(F, G) > 0$, there exists a further subsequence $\{d'_k\}$ such that $\lim_{d'_k\to\infty} e_{h,\psi}(F, G) > 0$, and the corresponding limits of the three terms in $e_{h,\psi}(F, G)$ exist. Let θ_1, θ_2 and θ_3 be the limiting values of $d^{-1}\sum_{q=1}^{d}\{\psi(|X_1^{(q)} - X_2^{(q)}|^2), d^{-1}\sum_{q=1}^{d}\{\psi(|Y_1^{(q)} - Y_2^{(q)}|^2)$ and $d^{-1}\sum_{q=1}^{d}\{\psi(|X_1^{(q)} - Y_1^{(q)}|^2)$, respectively, along the sub-sequence $\{d'_k\}$.

Since $\lim_{d'_k\to\infty} e_{h,\psi}(F,G) > 0$, we have $2\theta_3 > \theta_1 + \theta_2$. Hence, using Lemma A.4, we get $P(T^{h,\psi}_{n,m} > 1/3) \to 1$ as $d'_k \to \infty$. Since $\{d'_k\}$ is the sub-sequence of an arbitrary sequence $\{d_k\}$, we can conclude that $P(T^{h,\psi}_{n,m} > 1/3) \to 1$ as $d \to \infty$. Now, using arguments similar to those in the proof of Theorem 3.1, one can establish the consistency of the level α test when $\min\{n,m\} \geq 2 + 2/\alpha$. □

Lemma A.5. *If* $\mathbf{X}_1, \mathbf{X}_2 \overset{iid}{\sim} F$ *and* $\mathbf{Y}_1, \mathbf{Y}_2 \overset{iid}{\sim} G$ *are independent random vectors, then* $\Theta^2_{\ell_2}(F,G) \geq \{\mathbb{P}\{\|\mathbf{X}_1 - \mathbf{Y}_1\| \leq \|\mathbf{Y}_2 - \mathbf{Y}_1\|\} - 1/2\}^2$
$+ \{\mathbb{P}\{\|\mathbf{Y}_1 - \mathbf{X}_1\| \leq \|\mathbf{X}_2 - \mathbf{X}_1\|\} - 1/2\}^2$.

Proof. We know that for any random variable Z, $(\mathbb{E}\{Z\})^2 \leq \mathbb{E}\{Z^2\}$. Using this fact twice, we get

$$\begin{aligned}
\Theta^2_{\ell_2}(F,G) &= \int \{F(B(u,\ell_2(v,u))) - G(B(u,\ell_2(v,u)))\}^2 \, (dF(u)dF(v) + dG(u)dG(v)) \\
&\geq \left\{\int F(B(u,\ell_2(v,u))) - G(B(u,\ell_2(v,u)))\, dF(u)dF(v)\right\}^2 \\
&\qquad + \left\{\int F(B(u,\ell_2(v,u))) - G(B(u,\ell_2(v,u)))\, dG(u)dG(v)\right\}^2 \\
&= \{\mathbb{P}\{\|\mathbf{Y}_1 - \mathbf{X}_1\| \leq \|\mathbf{X}_2 - \mathbf{X}_1\|\} - 1/2\}^2 + \{\mathbb{P}\{\|\mathbf{X}_1 - \mathbf{Y}_1\| \leq \|\mathbf{Y}_2 - \mathbf{Y}_1\|\} - 1/2\}^2
\end{aligned}$$

The last equality follows from the fact that if $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3 \overset{iid}{\sim} F_0$, where F_0 is a continuous distribution, we have $\int F_0(B(u,\ell_2(v,u)))\, dF_0(u)dF_0(v) = \mathbb{P}\{\|\mathbf{X}_3 - \mathbf{X}_1\| \leq \|\mathbf{X}_2 - \mathbf{X}_1\|\} = \frac{1}{2}$. □

Lemma A.6. *Consider two d-dimensional random variables $\mathbf{X} = (\xi_1, 0, \ldots, 0)^\top$ and $\mathbf{Y} = (\xi_2, 0, \ldots, 0)^\top$, where $\xi_1 \sim N(\mu_1, 1)$ and $\xi_2 \sim N(\mu_2, 1)$ are independent. Let P_1 and P_2 denote the distributions of $\mathbf{X}$ and $\mathbf{Y}$, respectively. If $\mu_1 = cn^{-1/2}$ and $\mu_2 = -cm^{-1/2}$ for some $c > 0$, then there exists a constant $C > 0$ independent of the dimension d such that $\Theta^2(P_1, P_2) \geq C\big(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\big)^2$. This lower bound is tight up to a constant factor.*

Proof. Let $\xi_{11}, \xi_{12} \stackrel{iid}{\sim} N(\mu_1, 1)$ and $\xi_{21}, \xi_{22} \stackrel{iid}{\sim} N(\mu_2, 1)$ be independent random variables. In view of Lemma A.5, for $\mathbf{X}_1 = (\xi_{11}, 0, \ldots, 0)^\top, \mathbf{X}_2 = (\xi_{12}, 0, \ldots, 0)^\top \sim P_1$ and $\mathbf{Y}_1 = (\xi_{21}, 0, \ldots, 0)^\top$, $\mathbf{Y}_2 = (\xi_{22}, 0, \ldots, 0)^\top \sim P_2$, it is enough to prove that

$$[\mathbb{P}\{\|\mathbf{X}_1 - \mathbf{Y}_1\| \leq \|\mathbf{Y}_2 - \mathbf{Y}_1\|\} - 1/2]^2 + [\mathbb{P}\{\|\mathbf{Y}_1 - \mathbf{X}_1\| \leq \|\mathbf{X}_2 - \mathbf{X}_1\|\} - 1/2]^2 \geq C\Big(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\Big)^2,$$

Now, we derive the lower bounds for these two terms separately. Note that

$$\begin{aligned}\Big|\mathbb{P}\{\|\mathbf{X}_1 - \mathbf{Y}_1\| \leq \|\mathbf{Y}_2 - \mathbf{Y}_1\|\} - 1/2\Big| &= \Big|\mathbb{P}\{|\xi_{11} - \xi_{21}| \leq |\xi_{22} - \xi_{22}|\} - 1/2\Big| \\ &= \Big|\mathbb{P}\{|\xi_{11} - \xi_{21}|^2 - |\xi_{22} - \xi_{21}|^2 \leq 0\} - 1/2\Big| \\ &= \Big|\mathbb{P}\{(\xi_{11} + \xi_{22} - 2\xi_{21})(\xi_{11} - \xi_{22}) \leq 0\} - 1/2\Big|.\end{aligned}$$

So, taking $T_1 = \xi_{11} - \xi_{22}$ and $S_1 = \xi_{11} + \xi_{22} - 2\xi_{21}$, we get

$$\begin{aligned}\Big|\mathbb{P}\{\|\mathbf{X}_1 - \mathbf{Y}_1\| \leq \|\mathbf{Y}_2 - \mathbf{Y}_1\|\} - 1/2\Big| &= \Big|\mathbb{P}\{S_1 T_1 \leq 0\} - 1/2\Big| \\ &= \Big|\mathbb{E}\{\mathbb{P}\{S_1 T_1 \leq 0 \mid S_1\} - 1/2\}\Big|.\end{aligned}$$

Here, T_1 and S_1 jointly follow a bivariate normal distribution with $E(T_1) = c(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}})$, $E(S_1) = c(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}})$, $Var(T_1) = 2$, $Var(S_1) = 6$ and $Cov(T_1, S_1) = 0$. Therefore,

$$\begin{aligned}\Big|\mathbb{E}\big\{\mathbb{P}\{S_1T_1 \le 0 \mid S_1\} - 1/2\big\}\Big| &= \Big|\mathbb{E}\Big\{\Phi\Big(-c\Big(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\Big)\frac{S_1}{|S_1|\sqrt{2}}\Big) - 1/2\Big\}\Big| \\ &\ge \frac{c}{\sqrt{2}}\Big(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\Big)\phi(\sqrt{2}c),\end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density function and the distribution function of the standard normal variate, respectively. Here, the last inequality is obtained by using the mean value theorem and the fact that $\phi(t)$ is decreasing in $|t|$. So, we get

$$\Big|\mathbb{P}\{\|\mathbf{X}_1 - \mathbf{Y}_1\| \le \|\mathbf{Y}_2 - \mathbf{Y}_1\|\} - 1/2\Big| \ge \frac{c}{\sqrt{2}}\Big(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\Big)\phi(\sqrt{2}c).$$

Similarly, we can derive the same lower bound for $\Big|\mathbb{P}\{\|\mathbf{Y}_1 - \mathbf{X}_1\| \le \|\mathbf{X}_2 - \mathbf{X}_1\|\} - 1/2\Big|$ as well. So, we can find a constant $C > 0$ independent of the dimension d such that

$$\Theta^2_{\ell_2}(P_1, P_2) \ge C\Big(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\Big)^2.$$

To show that this lower bound is tight, notice that

$$\begin{aligned}\Theta^2(P_1, P_2) &\le \sup_A |P_1(A) - P_2(A)| \le KL(P_1, P_2) \\ &= \frac{c^2}{2}(\mu_1 - \mu_2)^2 = \frac{c^2}{2}\Big(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\Big)^2,\end{aligned}$$

where $KL(\cdot, \cdot)$ denotes the Kullback-Leibler divergence between two probability measures. Here, the first inequality follows trivially, and the second one is known as Pinsker's inequality (see Lemma 2.5 in Tsybakov, 2009). Hence the lower bound is tight up to a constant term. □

Proof of Theorem 4.1 . The minimax lower bound can be obtained based on the standard application of Neyman-Pearson lemma (see Baraud, 2002; Kim et al., 2020). Let the distributions of the sample under the null and alternative hypotheses be denoted as Q_0 and Q_1 respectively. Then following our notations, we have

$$R_{n,m,d}(\epsilon) \geq 1 - \alpha - \sup_A |Q_0(A) - Q_1(A)| \geq 1 - \alpha - \sqrt{\frac{1}{2} KL(Q_0, Q_1)},$$

where the second inequality is obtained from Pinsker's inequality (see Tsybakov, 2009). Now suppose that P_1 and P_2 are the distributions corresponding to $\mathbf{X} = (\xi_1, 0, \ldots, 0)^\top$ and $\mathbf{Y} = (\xi_2, 0, \ldots, 0)^\top$, where ξ_1 and ξ_2 are independent random variables following normal distributions with the unit variance and means

$$\mu_1 = \frac{\sqrt{2}(1-\alpha-\zeta)}{\sqrt{n}} \qquad \text{and} \qquad \mu_2 = -\frac{\sqrt{2}(1-\alpha-\zeta)}{\sqrt{m}},$$

respectively. Let P_0 be the distribution of $(\xi, 0, \ldots, 0)^\top$ where ξ is a standard normal random variable. Define $k(\alpha, \zeta) := (1-\alpha-\zeta)^2 \Big(\phi\big(\sqrt{2}(1-\alpha-\zeta)\big)\Big)^2$. Then by Lemma A.5, $(P_1, P_2) \in \mathcal{F}(c\lambda(n, m))$ for all $0 < c < k(\alpha, \zeta)$. Now taking $Q_0 = P_0^{(n+m)}$ and $Q_1 = P_1^n P_2^m$, we have

$$KL(Q_0, Q_1) = \frac{n}{2}\mu_1^2 + \frac{m}{2}\mu_2^2 = 2(1-\alpha-\zeta)^2.$$

Therefore, $R_{n,m,d}(c\lambda(n,m)) \geq \zeta$ for all $0 < c < k(\alpha,\zeta)$. Since ζ and $k(\alpha,\zeta)$ do not depend on n, m and d, this trivially satisfies the condition $\liminf_{n,m,d\to\infty} R_{n,m,d}(c\lambda(n,m)) \geq \zeta$ for all $0 < c < k(\alpha,\zeta)$. □

Proof of Theorem 4.2 . Here we want to show that for every positive α and ζ, there exists a constant $K(\alpha,\zeta)$ such that

$$\limsup_{n,m,d\to\infty} \sup_{(F,G)\in\mathcal{F}(c\lambda(n,m))} \mathbb{P}_{F,G}^{n,m}\{T_{n,m} \leq c_{1-\alpha}\} \leq \zeta$$

for all $c > K(\alpha,\zeta)$. Let us first choose a constant K_1 such that

$$K_1\Big(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{m}}\Big)^2 \geq \frac{1}{\alpha}\mathbb{E}\{T_{n,m}^{\pi} \mid \mathcal{U}\} = \frac{1}{6\alpha}\left(\frac{1}{n} + \frac{1}{m} + \frac{1}{n-2} + \frac{1}{m-2}\right).$$

Now, take any $(F,G) \in \mathcal{F}(c\lambda(n,m))$ such that $c > K_1$. Using the fact that $c_{1-\alpha} \leq \frac{1}{\alpha}\mathbb{E}\{T_{n,m}^{\pi} \mid \mathcal{U}\}$ (see the proof of Lemma 2.1), we get

$$\begin{aligned}\mathbb{P}_{F,G}^{n,m}\{T_{n,m} \leq c_{1-\alpha}\} &\leq \mathbb{P}_{F,G}^{n,m}\{T_{n,m} \leq \frac{1}{\alpha}\mathbb{E}\{T_{n,m}^{\pi} \mid \mathcal{U}\}\} \\ &= \mathbb{P}_{F,G}^{n,m}\{-T_{n,m} + \mathbb{E}_{F,G}\{T_{n,m}\} \geq \mathbb{E}_{F,G}\{T_{n,m}\} - \frac{1}{\alpha}\mathbb{E}\{T_{n,m}^{\pi} \mid \mathcal{U}\}\}.\end{aligned}$$

Since $\mathbb{E}_{F,G}\{T_{n,m}\} \geq \Theta_{\rho}^2(F,G) \geq c\lambda(n,m) \geq K_1\lambda(n,m) \geq \frac{1}{\alpha}\mathbb{E}\{T_{n,m,\pi}^{\rho} \mid \mathcal{U}\}$, using the Chebyshev's inequality, one gets

$$\begin{aligned}
&\mathbb{P}^{n,m}_{F,G}\{T_{n,m} \leq c_{1-\alpha}\} \\
&\leq \mathbb{P}^{n,m}_{F,G}\{-T_{n,m} + \mathbb{E}_{F,G}\{T_{n,m}\} \geq \mathbb{E}_{F,G}\{T_{n,m}\} - \frac{1}{\alpha}\mathbb{E}\{T^{\pi}_{n,m} \mid \mathcal{U}\}\} \\
&\quad \leq \frac{Var_{F,G}(T_{n,m})}{\left(\mathbb{E}_{F,G}\{T_{n,m}\} - \frac{1}{\alpha}\mathbb{E}\{T^{\pi}_{n,m} \mid \mathcal{U}\}\right)^2} \\
&\quad \leq \frac{C_1\Theta^2_{\rho}(F,G)\left(\frac{1}{n}+\frac{1}{m}\right) + C_2\left(\frac{1}{n}+\frac{1}{m}\right)^2}{\left(\frac{1}{6}\left(\frac{1}{n-2}+\frac{1}{m-2}\right) + \frac{1}{m}(p_0-p_1) + \frac{1}{n}(p_2-p_3) + \Theta^2_{\rho}(F,G) - \frac{1}{6\alpha}\left(\frac{1}{n}+\frac{1}{m}+\frac{1}{n-2}+\frac{1}{m-2}\right)\right)^2} \\
&\quad \leq \frac{C_1\Theta^2_{\rho}(F,G)\left(\frac{1}{n}+\frac{1}{m}\right) + C_2\left(\frac{1}{n}+\frac{1}{m}\right)^2}{\left(\Theta^2_{\rho}(F,G) - \frac{1}{6\alpha}\left(\frac{1}{n}+\frac{1}{m}+\frac{1}{n-2}+\frac{1}{m-2}\right)\right)^2},
\end{aligned}$$

This implies that

$$\limsup_{n,m,d\to\infty} \sup_{(F,G)\in\mathcal{F}(c\lambda(n,m))} \mathbb{P}^{n,m}_{F,G}\{T_{n,m} \leq c_{1-\alpha}\} \leq (C_1c + C_2)/\left(c - \frac{1}{3\alpha}\right)^2.$$

One can notice that this upper bound is a decreasing function in c, and as c grows to infinity, it goes to zero. Hence, for any $0 < \zeta < 1 - \alpha$, there exists a constant $K_2 > 0$ such that this upper bound is smaller than ζ. Now let $K(\alpha, \zeta) = \max\{K_1, K_2\}$. Then for $c > K(\alpha, \zeta)$ the maximum type II error rate is asymptotically upper bounded by ζ. This establishes the theorem. □

Proof of Theorem 4.3 . If F and G are such that $\lim\limits_{d\to\infty} \Theta^2_{\ell_2}(F,G)/\lambda(n,m) = \infty$, then from Theorem 4.2, we have $\lim_{d\to\infty} \mathbb{P}^{n,m}_{F,G}\{T_{n,m} \leq c_{1-\alpha}\} = 0$. Hence the power of the test converges to 1. □

Proof of Theorem 4.4 . Using similar arguments as in the proofs of Theorems 4.1 and 4.2, one can show that if h and ψ are strictly increasing functions, then for testing $H_0 : \Theta^2_{\varphi_{h,\psi}}(F,G) = 0$ against $H_1 : \Theta^2_{\varphi_{h,\psi}}(F,G) > \epsilon$, the minimax rate of separation is $\lambda(n,m) = (1/\sqrt{n} + 1/\sqrt{m})^2$, and the permutation test based on $T^{h,\psi}_{n,m}$ is minimax rate optimal. Hence, one gets a similar conclusion as in Theorem 4.3. □

Proof of Proposition 4.1 . Here, we use Lemma A.5 to establish the condition of Theorem 4.3 for different β. Assume that $\mathbf{X}_1, \mathbf{X}_2 \sim F = \prod_{i=1}^d \mathcal{N}_1(1/d^\beta, 1)$, $\mathbf{Y}_1, \mathbf{Y}_2 \sim G = \prod_{i=1}^d \mathcal{N}_1(-1/d^\beta, 1)$, and they are independent. Then

$$\begin{aligned}\Theta^2_{\ell_2}(F,G) &\geq \Big[\mathbb{P}\big(\|\mathbf{X}_1 - \mathbf{Y}_1\| \leq \|\mathbf{Y}_2 - \mathbf{Y}_1\|\big) - 1/2\Big]^2 \\ &\qquad + \Big[\mathbb{P}\big(\|\mathbf{Y}_1 - \mathbf{X}_1\| \leq \|\mathbf{X}_2 - \mathbf{X}_1\|\big) - 1/2\Big]^2 \\ &= \Big[\mathbb{P}\big(\|\mathbf{X}_1 - \mathbf{Y}_1\|^2 - \|\mathbf{Y}_2 - \mathbf{Y}_1\|^2 \leq 0\big) - 1/2\Big]^2 \\ &\qquad + \Big[\mathbb{P}\big(\|\mathbf{Y}_1 - \mathbf{X}_1\|^2 - \|\mathbf{X}_2 - \mathbf{X}_1\|^2 \leq 0\big) - 1/2\Big]^2 \\ &= \left[\mathbb{P}\left(\sum_{i=1}^d T_i S_i \leq 0\right) - \frac{1}{2}\right]^2 + \left[\mathbb{P}\left(\sum_{i=1}^d T_i' S_i' \leq 0\right) - \frac{1}{2}\right]^2,\end{aligned}$$

where $T_i = X_{1i} - Y_{2i}$, $S_i = X_{1i} + Y_{2i} - 2Y_{1i}$, $T_i' = Y_{1i} - X_{2i}$ and $S_i' = Y_{1i} + X_{2i} - 2X_{1i}$ $(i = 1, 2, \ldots, d)$. Clearly, T_i, S_i are independent, and so are T_i', S_i'. Here $S_1, S_2, \ldots, S_d \overset{iid}{\sim} N(\frac{2}{d^\beta}, 6)\big)$, and S_i' has the same distribution as

$-S_i$ for all $i = 1, 2, \ldots, d$. Now,

$$\Big[\mathbb{P}\big(\|\mathbf{X}_1 - \mathbf{Y}_1\| \leq \|\mathbf{Y}_2 - \mathbf{Y}_1\|\big) - 1/2\Big]^2 + \Big[\mathbb{P}\big(\|\mathbf{Y}_1 - \mathbf{X}_1\| \leq \|\mathbf{X}_2 - \mathbf{X}_1\|\big) - 1/2\Big]^2$$
$$= \left[\mathbb{E}\left\{\Phi\left(-\frac{2}{d^\beta}\frac{\sum_{i=1}^d S_i}{\sqrt{2\sum_{i=1}^d S_i^2}}\right)\right\} - \frac{1}{2}\right]^2 + \left[\mathbb{E}\left\{\Phi\left(\frac{2}{d^\beta}\frac{\sum_{i=1}^d S_i'}{\sqrt{2\sum_{i=1}^d S_i'^2}}\right)\right\} - \frac{1}{2}\right]^2.$$
$$= 2\left[\mathbb{E}\left\{\Phi\left(-\frac{2}{d^\beta}\frac{\sum_{i=1}^d S_i}{\sqrt{2\sum_{i=1}^d S_i^2}}\right)\right\} - \frac{1}{2}\right]^2$$

Hence, studying the behaviour of $Z(\beta) = \big(2\sum_{i=1}^d S_i\big)/\big(d^\beta\sqrt{2\sum_{i=1}^d S_i^2}\big)$ for different values of β will yield the conditions for the consistency of our test.

Note that $\sum_{i=1}^d S_i/d^{\beta+1/2} \sim N(2/d^{2\beta-1/2}, 6/d^{2\beta})$. Hence for $\beta < 1/4$, $\frac{1}{d^{\beta+1/2}}\sum_{i=1}^d S_i \xrightarrow{P} \infty$ and $\sum_{i=1}^d S_i^2/d \xrightarrow{P} 6$. So, for $\beta < 1/4$, $Z(\beta) \xrightarrow{P} \infty$. For $\beta = 1/4$, $\sum_{i=1}^d S_i/d^{\beta+1/2} \xrightarrow{P} 2$. So, $Z(\beta) \xrightarrow{P} 2/\sqrt{3}$. Therefore, for $\beta \leq 1/4$, we have

$$\liminf_{d\to\infty}\left[\mathbb{E}\left\{\Phi\left(-\frac{2}{d^\beta}\frac{\sum_{i=1}^d S_i}{\sqrt{2\sum_{i=1}^d S_i^2}}\right)\right\} - \frac{1}{2}\right]^2 > 0,$$

which in turn implies that $\liminf_{d\to\infty}\Theta^2_{\ell_2}(F,G) > 0$. This proves Proposition 4.1(a).

For $1/4 < \beta < 1/2$, notice that $d^{2\beta-1/2}\sum_{i=1}^d S_i/d^{\beta+1/2} \sim N(2, 6d^{2\beta-1})$. So, as d tends to infinity, $d^{2\beta-1/2}\sum_{i=1}^d S_i/d^{\beta+1/2} \xrightarrow{P} 2$. Now, if we take $n \asymp m \asymp d^\gamma$, to match this convergence rate so that $\Theta^2_{\ell_2}(F,G)/\lambda(n,m)$

diverges to infinity, we require the following

$$\lim_{d\to\infty} d^{\gamma}\left[\mathbb{E}\left\{\Phi\left(-\frac{2}{d^{2\beta-1/2}}\frac{d^{\beta-1}\sum_{i=1}^{d}S_i}{\sqrt{2\sum_{i=1}^{d}S_i^2/d}}\right)\right\}-\frac{1}{2}\right]^2=\infty.$$

This is possible when $\gamma > 4\beta - 1$. Also, note that for $\beta = 1/2$, $\sum_{i=1}^{d} S_i/d^{1/2}$ forms a tight sequence. In this case, we need

$$\lim_{d\to\infty} d^{\gamma}\left[\mathbb{E}\left\{\Phi\left(-\frac{2}{d^{\beta}}\frac{d^{-1/2}\sum_{i=1}^{d}S_i}{\sqrt{2\sum_{i=1}^{d}S_i^2/d}}\right)\right\}-\frac{1}{2}\right]^2=\infty,$$

which is satisfied when $\gamma > 1 = 4\beta - 1$. This proves Proposition 4.1(b).

Now for $\beta > 1/2$, we have $\sum_{i=1}^{d} S_i/d^{1/2} \sim N(2/d^{\beta-1/2}, 6)$, and hence it is a tight sequence of random variables. In this scenario we require

$$\lim_{d\to\infty} d^{\gamma}\left[\mathbb{E}\left\{\Phi\left(-\frac{2}{d^{\beta}}\frac{d^{-1/2}\sum_{i=1}^{d}S_i}{\sqrt{2\sum_{i=1}^{d}S_i^2/d}}\right)\right\}-\frac{1}{2}\right]^2=\infty,$$

which is satisfied if $\gamma > 2\beta$. Also notice that when $\beta > 1/2$ and $\gamma < 2\beta - 1$, the Kullback-Leibler Divergence ($KL(Q_1, Q_0) \asymp d^{\gamma-2\beta+1}$) converges to zero with increasing dimensions. Hence in this scenario, the asymptotic type II error rate of any test remains bounded below by $1-\alpha$, i.e., no tests have asymptotic power more than the nominal level α. This completes the proof of Proposition 4.1(c). $\square$

Appendix B: Numerical Results on SSE Models

In the main part of the manuscript, we analyzed several simulated data sets belonging to non-strongly spike eigenvalue (NSSE) models to study the high dimensional behaviour of our tests. Here we consider two examples involving data sets generated using strongly spiked eigenvalue (SSE) models (see Aoshima and Yata, 2018) and analyze them. In particular, we consider a scale problem (Example A1) and a location problem (Example A2) and investigate the performance of different methods when the sample sizes remain fixed (50 from each distribution) and the dimension increases.

Example E1: *Two probability distributions* $F = N_d(\mathbf{0}_d, \mathbf{\Sigma}_d^{\circ}(1.1))$ *and* $G = N_d(0, \mathbf{\Sigma}_d^{\circ}(1.5))$ *differ only in the scale of the first coordinate variable. Here* $\mathbf{\Sigma}_d^{\circ}(\gamma)$ *denotes the* $d \times d$ *diagonal matrix with the first diagonal entry* d^{γ} *for some* $\gamma > 0$ *and the rest equal to unity.*

Example E2: *Here* $F = N_d(\mathbf{0}_d, \mathbf{\Sigma}_d^{\circ}(1.5))$ *and* $G = N_d(0.5\mathbf{1}_d, \mathbf{\Sigma}_d^{\circ}(1.5))$ *differ only in location. Note that* $\mathbf{\Sigma}_d^{\circ}(\gamma)$ *has the same meaning as in Example E1, and here* $\mathbf{1}_d$ *is a* d *dimensional vector with all entries being unity.*

Since $\lim \lambda_{max}^2(\mathbf{\Sigma}_d^{\circ}(\gamma))/trace(\mathbf{\Sigma}_d^{\circ^2}(\gamma)) = \lim d^{2\alpha}/(d^{2\alpha}+d-1) = 1$ for any $\gamma > 1$, both examples belong to SSE models. For each of these examples, we considered 10 different choices of d (2^i for $i = 1, \ldots, 10$), and in each case, we repeated the experiment 500 times to estimate the power of the tests by the proportion of times they rejected the null hypothesis ($H_0 : F = G$). These results are summarized in Figure F1.

Note that in Example E1, $(X_{11} - X_{21}) \sim N(0, 2d^{1.1})$ and $(X_{1i} - X_{2i}) \sim N(0, 2)$ for $i = 2, \ldots, d$. Hence, as $d \to \infty$, $\|X_1 - X_2\|^2/2d^{1.1} \xrightarrow{L} \chi^2_1$ (converges in distribution to a chi-square random variable with one degree of freedom). Similarly, as $d \to \infty$, we have $\|Y_1 - Y_2\|^2/2d^{1.5} \xrightarrow{L} \chi^2_1$, $\|X_1 - Y_1\|^2/d^{1.5} \xrightarrow{L} \chi^2_1$, and hence $\mathbb{P}\{\|X_1 - X_2\| \leq \|X_1 - Y_1\|\} \to 1$. Thus by Lemma A.4, the HDLSS consistency of the BD-ℓ_2 test holds. One can use similar arguments to show the consistency of BD-ℓ_1 test. This was the reason behind the excellent performance by these tests. In this example, the BG test performed best followed by our BD-ℓ_2 and BD-

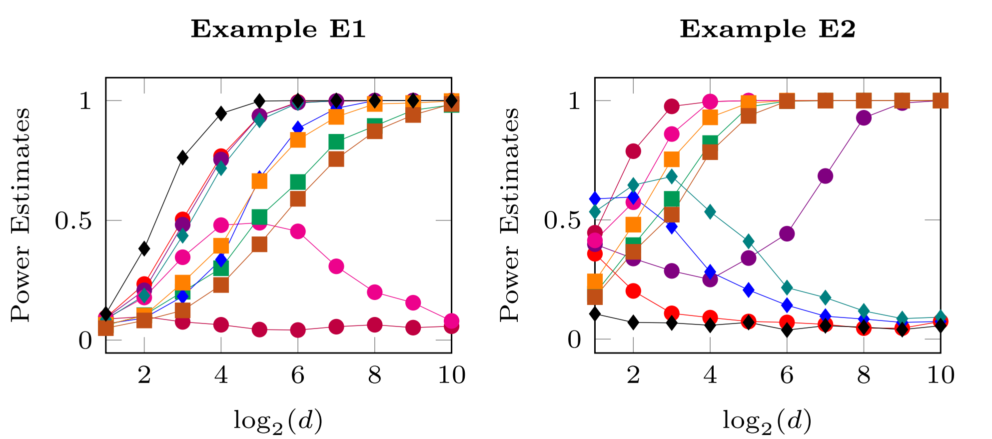

Figure F1: Powers of BD-ℓ_2 (●), BD-ℓ_1 (●), BD-exp (●), BD-log (●), FR (■), BF (♦), NN (■), MMD (♦), SHP (■), BG (♦) tests in Examples E1 and E2.

ℓ_1 tests. Except for BD-exp and BD-log, powers of all other tests also converged to 1 as the dimension increased. In cases of these two tests, the pairwise distances $\sum_{i=1}^{d} \psi\big((X_{1i} - X_{2i})^2\big)/d, \sum_{i=1}^{d} \psi\big((Y_{1i} - Y_{2i})^2/2\big)/d$ and $\sum_{i=1}^{d} \psi\big((X_{1i} - Y_{1i})^2/2\big)/d$ converges in probability to the same positive constant as d goes to infinity. So, unlike BD-ℓ_2 and BD-ℓ_1, they were unable to extract substantial discriminatory information from the first coordinate.

In Example E2, powers of BD-ℓ_2, BF, BG and MMD tests dropped down as d increased, but those for the graph-based tests increased steadily. Note that while the performance of the graph-based tests depends on the ordering of the pairwise distances, those of the above mention four tests depends on their magnitudes. In this example, though the inter-sample distances had a tendency to take higher values than the intra-sample distances, as d grows to infinity, $\|X_1 - X_2\|^2/2d^{1.5}$, $\|Y_1 - Y_2\|^2/2d^{1.5}$ and $\|X_1 - Y_1\|^2/2d^{1.5}$ all converge in distribution to a chi-square random variable with one degree of freedom, and that is why they failed to discriminate among the two populations in higher dimensions. However, BD-exp and BD-log tests outperformed all graph-based tests in this example. Note that here $\sum_{i=1}^{d} \psi\big((X_{1i} - X_{2i})^2\big)/d$ and $\sum_{i=1}^{d} \psi\big((Y_{1i} - Y_{2i})^2/2\big)/d$ converge in probability to the same limit but $\sum_{i=1}^{d} \psi\big((X_{1i} - Y_{1i})^2/2\big)/d$ converges in probability to a limit higher than that. This explains the excellent performance of these tests. Beacuse of the same reason the BD-ℓ_1 also had increasing power. In this example, the first coordinate difference was a dominating term in the ℓ_1 distance, and along that coordinate, we had very little difference between the two populations. This affected the performance of the BD-ℓ_1 test. In this case, coordinate-wise standardization of the variables may improve the performance of this test. Similar coordinatewise standardization may lead to better performance by the BD$-\ell_2$ test as well.

Bibliography

Aoshima, M. and K. Yata (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statistica Sinica 28*(1), 43–62.

Baraud, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli 8*(5), 577–606.

Kim, I. (2021). Comparing a large number of multivariate distributions. *Bernoulli 27*(1), 419–441.

Kim, I., S. Balakrishnan, and L. Wasserman (2020). Robust multivariate nonparametric tests via projection averaging. *Ann. Statist. 48*(6), 3417–3441.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab. 18*(3), 1269–1283.

Sarkar, S. and A. K. Ghosh (2018). On some high-dimensional two-sample tests based on averages of inter-point distances. *Stat 7*, e187, 16.

Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation.* Springer, New York.

Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Kolkata.

E-mail: banerjeebilol@outlook.com

E-mail: anilkghosh@gmail.com