SUPPLEMENT TO "A KERNEL INDEPENDENCE TEST USING PROJECTION-BASED MEASURE IN HIGH-DIMENSION"

Yuexin Chen and Wangli Xu

Renmin University of China

Supplementary material includes the interpretations of Conditions (A1)–(A3), the detailed proofs of the main theorems and propositions, and the additional simulation results.

S1. Interpretation of Condition (A1)

In the high-dimensional regime, we allow the γ_1 and γ_2 to approach 0, as a function of the sample size n, projection number k and the dimensions p, q. Recall from Remark 4 that we assume $\gamma_1 p \to \infty$ and $\gamma_2 q \to \infty$ to make sure that γ_1 and γ_2 can not approach 0 arbitrarily fast.

We consider $(X, Y) \in \mathbb{R}^{p+q}$ are jointly Gaussian with covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix}.$$

Suppose X and Y have uniformly upper and lower bounded fourth moments, and the spectrum of Σ is contained in $[M^{-1}, M]$ for $M > 1$. Recall from Section 4.2 that $X_{1,\mathcal{S}}$ is defined as the subvector of X_1 which the coordinates belong to $\mathcal{S}$. Let $\tau^2_{X,\mathcal{S}} = E\|X_{1,\mathcal{S}} - X_{2,\mathcal{S}}\|^2 = 2\,\mathrm{tr}(\Sigma_{X,\mathcal{S}})$, $\tau^2_{Y,\mathcal{S}} = E\|Y_{1,\mathcal{S}} - Y_{2,\mathcal{S}}\|^2 = 2\,\mathrm{tr}(\Sigma_{Y,\mathcal{S}})$, where $\Sigma_{X,\mathcal{S}}$ and $\Sigma_{Y,\mathcal{S}}$ are the covariance matrices of $X_{1,\mathcal{S}}$ and $Y_{1,\mathcal{S}}$, respectively.

Under the Gaussian design, Han and Shen (2021) provided the precise mean and variance expansions for $\mathrm{HSIC}_n(X, Y)$ in the regime that $n \to \infty$ and $\min\{p, q\} \to \infty$. Next we apply these results to give the orders of the corresponding quantities in Condition (A1). Recall from Section 2 that the kernels can be represented compactly as

$$K(X_1, X_2) = f\left(\frac{\|X_1 - X_2\|}{\gamma_X}\right), L(Y_1, Y_2) = g\left(\frac{\|Y_1 - Y_2\|}{\gamma_Y}\right),$$

where $f(x)$, $g(x)$ are continuously differentiable functions. Suppose that $f(x)$ and $g(x)$ satisfy Assumption B in Han and Shen (2021), which can easily be checked out to be satisfied for the common used kernels.

First we deal with the term $E\{d^{\alpha}_K(X_1, X_2)^2\}$. Recall from (4.12) and (4.13) that when α and β follow the randomly sparsified Gaussian distribution with parameters γ_1 and γ_2, the projection kernels $\tilde{K}(X_1, X_2)$ and $\tilde{L}(Y_1, Y_2)$ are essentially the weighted sum of the kernels $K^*(X_{1,\mathcal{S}_1}, X_{2,\mathcal{S}_1})$ and $L^*(Y_{1,\mathcal{S}_2}, Y_{2,\mathcal{S}_2})$, respectively. Let $\rho_{X,\mathcal{S}} = \tau_{X,\mathcal{S}}/\gamma_{X,\mathcal{S}}$, $\rho_{Y,\mathcal{S}} = \tau_{Y,\mathcal{S}}/\gamma_{Y,\mathcal{S}}$ where $\gamma_{X,\mathcal{S}}$ and $\gamma_{Y,\mathcal{S}}$ denote the corresponding bandwidths. We also assume that $\tau_{X,\mathcal{S}}/\gamma_{X,\mathcal{S}}$ and $\tau_{Y,\mathcal{S}}/\gamma_{Y,\mathcal{S}}$ are contained in $[M^{-1}, M]$. Then using the mean expansion in Proposition E.8 of Han and Shen (2021) with $X = Y$,

$$\begin{aligned} E\{d_K^{\alpha}(X_1, X_2)^2\} &= \sum_{t=1}^{p} \gamma_1^t (1-\gamma_1)^{p-t} \sum_{\substack{\mathcal{S}_1 \subset \{1,\ldots,p\} \\ |\mathcal{S}_1| = t}} E\{d_{K^*}(X_{1,\mathcal{S}_1}, X_{2,\mathcal{S}_1})^2\} \\ &\leq C \sum_{t=1}^{p} \gamma_1^t (1-\gamma_1)^{p-t} \sum_{\substack{\mathcal{S}_1 \subset \{1,\ldots,p\} \\ |\mathcal{S}_1| = t}} \frac{f'(\rho_{X,\mathcal{S}_1})^2}{\gamma_{X,\mathcal{S}_1}^2} \frac{\|\Sigma_{X,\mathcal{S}_1}\|_F^2}{\tau_{X,\mathcal{S}_1}^2} \\ &= O\left\{ \sum_{t=1}^{p} \gamma_1^t (1-\gamma_1)^{p-t} \binom{p}{t} \frac{1}{t} \right\} = O\left(\frac{1}{p\gamma_1} \right), \end{aligned} \tag{S1.1}$$

where C is a positive constant, the second equality follows from the assumptions that Σ has bounded spectrum, the last equality follows from the order of the inverse moments of binomial distribution as $p \to \infty$ (see Example 2.1 in Hu et al. (2014)), and the summation $\sum_{\mathcal{S}_1 \subset \{1,\ldots,p\}, |\mathcal{S}_1| = t}$ is over all t-subsets of $\{1, 2, \ldots, p\}$.

On the other hand, it follows from (4.12) and (4.13) that

$$\begin{aligned} &E\left\{U(X_1, X_2)^2\right\} = E\{d_K^{\alpha_1}(X_1, X_2) d_K^{\alpha_2}(X_1, X_2)\} \\ &= \sum_{t=1}^{p} \sum_{s=1}^{p} \sum_{\substack{\mathcal{S}_1 \subset \{1,\ldots,p\} \\ |\mathcal{S}_1| = t}} \sum_{\substack{\mathcal{S}_2 \subset \{1,\ldots,p\} \\ |\mathcal{S}_2| = s}} \gamma_1^{t+s} (1-\gamma_1)^{2p-t-s} E\{d_{K^*}(X_{1,\mathcal{S}_1}, X_{2,\mathcal{S}_1}) d_{K^*}(X_{1,\mathcal{S}_2}, X_{2,\mathcal{S}_2})\} \end{aligned}$$

Note that for every set function $F(\mathcal{S}_1, \mathcal{S}_2)$, the summation in the equality can be expressed in the intersection form of $\mathcal{S}_1$ and $\mathcal{S}_2$, that is

$$\sum_{t=1}^{p} \sum_{s=1}^{p} \sum_{\substack{\mathcal{S}_1 \subset \{1,\ldots,p\} \\ |\mathcal{S}_1| = t}} \sum_{\substack{\mathcal{S}_2 \subset \{1,\ldots,p\} \\ |\mathcal{S}_2| = s}} F(\mathcal{S}_1, \mathcal{S}_2) = \sum_{k=0}^{p} \sum_{|\mathcal{S}_1 \cap \mathcal{S}_2| = k} F(\mathcal{S}_1, \mathcal{S}_2)$$

where the summation $\sum_{|\mathcal{S}_1 \cap \mathcal{S}_2| = k}$ is over all subsets $\mathcal{S}_1, \mathcal{S}_2 \subset \{1, 2, \ldots, p\}$

satisfying $|\mathcal{S}_1 \cap \mathcal{S}_2| = k$. Then we have

$$
\begin{aligned}
&E\left\{U(X_1, X_2)^2\right\} \\
&= \sum_{k=0}^{p} \sum_{|\mathcal{S}_1 \cap \mathcal{S}_2| = k} \gamma_1^{|\mathcal{S}_1| + |\mathcal{S}_2|} (1-\gamma_1)^{2p - |\mathcal{S}_1| - |\mathcal{S}_2|} E\{d_{K^*}(X_{1,\mathcal{S}_1}, X_{2,\mathcal{S}_1}) d_{K^*}(X_{1,\mathcal{S}_2}, X_{2,\mathcal{S}_2})\}
\end{aligned}
$$

Using the mean expansion in Proposition E.8 of Han and Shen (2021), we obtain

$$
\begin{aligned}
&E\left\{U(X_1, X_2)^2\right\} \\
&\geq C \sum_{k=0}^{p} \sum_{|\mathcal{S}_1 \cap \mathcal{S}_2| = k} \gamma_1^{|\mathcal{S}_1| + |\mathcal{S}_2|} (1-\gamma_1)^{2p - |\mathcal{S}_1| - |\mathcal{S}_2|} \frac{f'(\rho_{X,\mathcal{S}_1}) f'(\rho_{X,\mathcal{S}_2})}{\gamma_{X,\mathcal{S}_1} \gamma_{X,\mathcal{S}_2}} \frac{\|\Sigma_{\mathcal{S}_1,\mathcal{S}_2}\|_F^2}{\tau_{X,\mathcal{S}_1} \tau_{X,\mathcal{S}_2}} \\
&\geq C \sum_{k=0}^{p} \sum_{|\mathcal{S}_1 \cap \mathcal{S}_2| = k} \gamma_1^{|\mathcal{S}_1| + |\mathcal{S}_2|} (1-\gamma_1)^{2p - |\mathcal{S}_1| - |\mathcal{S}_2|} \frac{k}{|\mathcal{S}_1||\mathcal{S}_2|} \\
&= C \sum_{k=1}^{p} \binom{p}{k} \sum_{i=k}^{p} \binom{p-k}{i-k} \sum_{j=k}^{p-i+k} \binom{p-i}{j-k} \gamma_1^{i+j} (1-\gamma_1)^{2p-i-j} \frac{k}{ij} \\
&\geq C \sum_{k=1}^{p} \binom{p}{k} \sum_{i=k}^{p} \binom{p-k}{i-k} \gamma_1^{i} (1-\gamma_1)^{p-i} \frac{\gamma_1^{k-1} (1-\gamma_1)^{i-k}}{i(p-i+1)} \\
&\geq C \sum_{k=1}^{p} \binom{p}{k} (\gamma_1^2)^{k-1} (1-\gamma_1^2)^{p-k} (1+\gamma_1)^2 \frac{1}{(k+1)(p-k+1)(p-k+2)} \\
&= O\left(\frac{1}{p^3 \gamma_1^4}\right), \qquad \text{(S1.2)}
\end{aligned}
$$

where C in each step represents a sufficiently small positive constant, the third and the fourth inequalities utilize the fact that for $k = 1, \ldots, p$,

$$
\frac{j+1}{j+k} \geq \frac{1}{k}, \quad \frac{i+1}{i+k} \geq \frac{1}{k+1}.
$$

The last equality follows from the assumption that $\gamma_1 p \to \infty$. The first equality follows from the fact that for each function $F(|\mathcal{S}_1|, |\mathcal{S}_2|)$ that only depends on the size of the set,

$$\sum_{|\mathcal{S}_1\cap\mathcal{S}_2|=k} F(|\mathcal{S}_1|, |\mathcal{S}_2|) = \binom{p}{k}\sum_{i=k}^{p}\binom{p-k}{i-k}\sum_{j=k}^{p-i+k}\binom{p-i}{j-k}F(i,j).$$

Combining (S1.1) and (S1.2) leads to

$$\frac{E\{d_K^{\alpha}(X_1,X_2)^2\}}{E\left\{U(X_1,X_2)^2\right\}} = O\left(p^2\gamma_1^3\right).$$

Similar result applies to $E\{d_K^{\alpha}(X_1,X_2)^2\}$ and $E\left\{U(X_1,X_2)^2\right\}$. Thus under the independence of X and Y,

$$\frac{E\{d_K^{\alpha}(X_1,X_2)^2 d_L^{\beta}(Y_1,Y_2)^2\}}{(k\wedge n)\sigma_1^2} = O\left(\frac{p^2q^2\gamma_1^3\gamma_2^3}{k\wedge n}\right). \tag{S1.3}$$

S2. Interpretations of Conditions (A2)–(A3)

In this section, we present the interpretations of Conditions (A2)–(A3). Under the null hypothesis, Condition (A2) is satisfied when

$$\frac{E\{g_K(X_1,X_2,X_3,X_4)\}}{[E\{U(X_1,X_2)^2\}]^2} = o(1), \quad \frac{E\{g_L(X_1,X_2,X_3,X_4)\}}{[E\{V(Y_1,Y_2)^2\}]^2} = o(1). \tag{S2.4}$$

Let $L^2(P_X)$ be the space of functions such that $E\{f(X)^2\} \leq \infty$ for $f \in L^2(P_X)$. To provide some insight on Condition (A2), we define the integral operator T on $L^2(P_X)$ as $T(f)(x) = E\{U(X,x)f(X)\}$. Suppose that $E\{K(X_1,X_2)^4\} < \infty$. According to Theorems VI.16 and VI.23 of Reed (1972), the operator T is self-adjoint and Hilbert-Schmidt, then there exist a complete orthonornal basis $\{\phi_i\}_{i=1}^{\infty}$ such that $T(\phi_i) = \lambda_i\phi_i$. The eigenvalues

$\{\lambda_i\}_{i=1}^{\infty}$ associated with operator T satisfy $\sum_{i=1}^{\infty}\lambda_i^2 < \infty$ and $\lambda_i \to 0$ as $i \to \infty$. Notice that $E\{g_K(X_1,X_2,X_3,X_4)\} = \sum_{i=1}^{\infty}\lambda_i^4$ and $E\{U(X_1,X_2)^2\} = \sum_{i=1}^{\infty}\lambda_i^2$. Thus the condition $E\{g_K(X_1,X_2,X_3,X_4)\}/[E\{U(X_1,X_2)^2\}]^2 = o(1)$ boils down to $\sum_{i=1}^{\infty}\lambda_i^4/(\sum_{i=1}^{\infty}\lambda_i^2)^2 = o(1)$, which can be viewed as a generalization of the condition in the covariance matrix level (see Condition (3.6) of Chen and Qin (2010) and Condition (2.8) of Zhong and Chen (2011)) to the operator level.

Below, we give the orders of the associated moments in Condition (A2) when (X,Y) are jointly Gaussian. Under the setting of Section S1, from the residual moment estimates in Lemma 7.4 of Han and Shen (2021), we obtain

$$
\begin{aligned}
&E\{g_K(X_1,X_2,X_3,X_4)\} = E\{d_K^{\alpha_1}(X_1,X_2)d_K^{\alpha_2}(X_1,X_3)d_K^{\alpha_3}(X_2,X_4)d_K^{\alpha_4}(X_3,X_4)\}\\
&= \sum_{t,s,u,v=1}^{p}\sum_{|\mathcal{S}_1|=t,|\mathcal{S}_2|=s,|\mathcal{S}_3|=u,|\mathcal{S}_4|=v}\gamma_1^{t+s+u+v}(1-\gamma_1)^{4p-t-s-u-v}\\
&\quad E\{d_{K^*}(X_{1,\mathcal{S}_1},X_{2,\mathcal{S}_1})d_{K^*}(X_{1,\mathcal{S}_2},X_{3,\mathcal{S}_2})d_{K^*}(X_{2,\mathcal{S}_3},X_{4,\mathcal{S}_3})d_{K^*}(X_{3,\mathcal{S}_4},X_{4,\mathcal{S}_4})\}\\
&\le C\sum_{t,s,u,v=1}^{p}\sum_{|\mathcal{S}_1|=t,|\mathcal{S}_2|=s,|\mathcal{S}_3|=u,|\mathcal{S}_4|=v}\gamma_1^{t+s+u+v}(1-\gamma_1)^{4p-t-s-u-v}\\
&\quad \frac{f'(\rho_{X,\mathcal{S}})f'(\rho_{X,\mathcal{S}_2})f'(\rho_{X,\mathcal{S}_3})f'(\rho_{X,\mathcal{S}_4})}{\gamma_{X,\mathcal{S}_1}\gamma_{X,\mathcal{S}_2}\gamma_{X,\mathcal{S}_3}\gamma_{X,\mathcal{S}_4}}\frac{E\{X_{1,\mathcal{S}_1}^{\top}X_{2,\mathcal{S}_1}X_{1,\mathcal{S}_2}^{\top}X_{3,\mathcal{S}_2}X_{2,\mathcal{S}_3}^{\top}X_{4,\mathcal{S}_3}X_{3,\mathcal{S}_4}^{\top}X_{4,\mathcal{S}_4}\}}{\tau_{X,\mathcal{S}_1}\tau_{X,\mathcal{S}_2}\tau_{X,\mathcal{S}_3}\tau_{X,\mathcal{S}_4}}\\
&\le C\sum_{t,s,u,v=1}^{p}\sum_{|\mathcal{S}_1|=t,|\mathcal{S}_2|=s,|\mathcal{S}_3|=u,|\mathcal{S}_4|=v}\gamma_1^{t+s+u+v}(1-\gamma_1)^{4p-t-s-u-v}\frac{\mathrm{tr}(\Sigma_{\mathcal{S}_3,\mathcal{S}_1}\Sigma_{\mathcal{S}_1,\mathcal{S}_2}\Sigma_{\mathcal{S}_2,\mathcal{S}_4}\Sigma_{\mathcal{S}_4,\mathcal{S}_3})}{|\mathcal{S}_1||\mathcal{S}_2||\mathcal{S}_3||\mathcal{S}_4|},
\end{aligned}
$$

where the summation $\sum_{|\mathcal{S}_1|=t,|\mathcal{S}_2|=s,|\mathcal{S}_3|=u,|\mathcal{S}_4|=v}$ is over all t-subsets $\mathcal{S}_1$, s-

subsets $\mathcal{S}_2$, u-subsets $\mathcal{S}_3$ and v-subsets $\mathcal{S}_4 \subset \{1, 2 \ldots, p\}$. It follows from the bounded spectrum of Σ that

$$
\begin{aligned}
&E\{g_K(X_1, X_2, X_3, X_4)\} \\
&\leq C \sum_{t,s,u,v=1}^{p} \sum_{|\mathcal{S}_1|=t, |\mathcal{S}_2|=s, |\mathcal{S}_3|=u, |\mathcal{S}_4|=v} \gamma_1^{t+s+u+v}(1-\gamma_1)^{4p-t-s-u-v} \frac{\max\{|\mathcal{S}_1|, |\mathcal{S}_2|, |\mathcal{S}_3|, |\mathcal{S}_4|\}}{|\mathcal{S}_1||\mathcal{S}_2||\mathcal{S}_3||\mathcal{S}_4|} \\
&\leq C \sum_{t,s,u,v=1}^{p} \sum_{|\mathcal{S}_1|=t, |\mathcal{S}_2|=s, |\mathcal{S}_3|=u, |\mathcal{S}_4|=v} \gamma_1^{t+s+u+v}(1-\gamma_1)^{4p-t-s-u-v} \frac{|\mathcal{S}_1| + |\mathcal{S}_2| + |\mathcal{S}_3| + |\mathcal{S}_4|}{|\mathcal{S}_1||\mathcal{S}_2||\mathcal{S}_3||\mathcal{S}_4|} \\
&\leq C \sum_{t,s,u,v=1}^{p} \binom{p}{t}\binom{p}{s}\binom{p}{u}\binom{p}{v} \gamma_1^{t+s+u+v}(1-\gamma_1)^{4p-t-s-u-v} \frac{t+s+u+v}{tsuv} \\
&= O\left(\frac{1}{p^3\gamma_1^3}\right), \qquad \text{(S2.5)}
\end{aligned}
$$

where C in each step represents a sufficiently large positive constant, the last equality follows from the order of the inverse moments of binomial distribution as $p \to \infty$ (see Example 2.1 in Hu et al. (2014)). Combining (S1.2) and (S2.5) leads to

$$
\frac{E\{g_K(X_1, X_2, X_3, X_4)\}}{[E\{U(X_1, X_2)^2\}]^2} = O(p^3\gamma_1^5).
$$

Similar result applies to $E\{g_L(X_1, X_2, X_3, X_4)\}/[E\{V(Y_1, Y_2)^2\}]^2$. Thus (S2.4) can be satisfied when $p^3\gamma_1^5 = o(1)$ and $q^3\gamma_2^5 = o(1)$.

Now we turn to the interpretations of Condition (A3). Under the null hypothesis, Condition (A3) is satisfied when

$$
\frac{E\{U(X_1, X_2)\}^4}{[E\{U(X_1, X_2)^2\}]^2} = o(n^{1/2}), \quad \frac{E\{V(Y_1, Y_2)\}^4}{[E\{V(Y_1, Y_2)^2\}]^2} = o(n^{1/2}). \qquad \text{(S2.6)}
$$

By the similar calculations, we can deduce that

$$E\{U(X_1,X_2)^4\} = E\{d_K^{\alpha_1}(X_1,X_2)d_K^{\alpha_2}(X_1,X_2)d_K^{\alpha_3}(X_1,X_2)d_K^{\alpha_4}(X_1,X_2)\}$$

$$= \sum_{t,s,u,v=1}^{p} \sum_{|\mathcal{S}_1|=t,|\mathcal{S}_2|=s,|\mathcal{S}_3|=u,|\mathcal{S}_4|=v} \gamma_1^{t+s+u+v}(1-\gamma_1)^{4p-t-s-u-v}$$

$$E\{d_{K^*}(X_{1,\mathcal{S}_1},X_{2,\mathcal{S}_1})d_{K^*}(X_{1,\mathcal{S}_2},X_{2,\mathcal{S}_2})d_{K^*}(X_{1,\mathcal{S}_3},X_{2,\mathcal{S}_3})d_{K^*}(X_{1,\mathcal{S}_4},X_{2,\mathcal{S}_4})\}$$

$$\le C \sum_{t,s,u,v=1}^{p} \sum_{|\mathcal{S}_1|=t,|\mathcal{S}_2|=s,|\mathcal{S}_3|=u,|\mathcal{S}_4|=v} \gamma_1^{t+s+u+v}(1-\gamma_1)^{4p-t-s-u-v}$$

$$\frac{f'(\rho_{X,\mathcal{S}})f'(\rho_{X,\mathcal{S}_2})f'(\rho_{X,\mathcal{S}_3})f'(\rho_{X,\mathcal{S}_4})}{\gamma_{X,\mathcal{S}_1}\gamma_{X,\mathcal{S}_2}\gamma_{X,\mathcal{S}_3}\gamma_{X,\mathcal{S}_4}} \frac{E\{X_{1,\mathcal{S}_1}^\top X_{2,\mathcal{S}_1}X_{1,\mathcal{S}_2}^\top X_{2,\mathcal{S}_2}X_{1,\mathcal{S}_3}^\top X_{2,\mathcal{S}_3}X_{1,\mathcal{S}_4}^\top X_{2,\mathcal{S}_4}\}}{\tau_{X,\mathcal{S}_1}\tau_{X,\mathcal{S}_2}\tau_{X,\mathcal{S}_3}\tau_{X,\mathcal{S}_4}}$$

$$\le C \sum_{t,s,u,v=1}^{p} \sum_{|\mathcal{S}_1|=t,|\mathcal{S}_2|=s,|\mathcal{S}_3|=u,|\mathcal{S}_4|=v} \gamma_1^{t+s+u+v}(1-\gamma_1)^{4p-t-s-u-v}\frac{\max_{i=1,\dots,4}[E\{(X_{1,\mathcal{S}_i}^\top X_{2,\mathcal{S}_i})^4\}]}{|\mathcal{S}_1||\mathcal{S}_2||\mathcal{S}_3||\mathcal{S}_4|}.$$

It follows from the moments of the quadratic form of Gaussian vectors (Proposition A.1 of Chen et al. (2010)) and the bouned spectrum of Σ that

$$E\{(X_{1,\mathcal{S}_i}^\top X_{2,\mathcal{S}_i})^4\} = O[\{\mathrm{tr}(\Sigma_{X,\mathcal{S}_i}^2)\}^2] = O(|\mathcal{S}_i|^2).$$

Similar to the derivation in (S2.5), we obtain

$$\begin{aligned} E\{U(X_1,X_2)^4\} \le & C \sum_{t,s,u,v=1}^{p} \sum_{|\mathcal{S}_1|=t,|\mathcal{S}_2|=s,|\mathcal{S}_3|=u,|\mathcal{S}_4|=v} \gamma_1^{t+s+u+v}(1-\gamma_1)^{4p-t-s-u-v} \\ & \frac{\max\{|\mathcal{S}_1|^2,|\mathcal{S}_2|^2,|\mathcal{S}_3|^2,|\mathcal{S}_4|^2\}}{|\mathcal{S}_1||\mathcal{S}_2||\mathcal{S}_3||\mathcal{S}_4|} \\ = & O\left(\frac{1}{p^2\gamma_1^2}\right). \end{aligned} \tag{S2.7}$$

Combining (S1.2) and (S2.7) leads to

$$\frac{E\{U(X_1,X_2)\}^4}{n^{1/2}[E\{U(X_1,X_2)^2\}]^2} = O\left(\frac{p^4\gamma_1^6}{n^{1/2}}\right)$$

Similar result applies to $E\{V(Y_1,Y_2)\}^4/(n^{1/2}[E\{V(Y_1,Y_2)^2\}]^2)$. Thus (S2.6) can also be implied by $p^4\gamma_1^6/n^{1/2}=o(1)$ and $q^4\gamma_2^6/n^{1/2}=o(1)$.

S3. Proof of Proposition 1

Proof. The class of distance-based kernels defined in (2.4) with positive definite function $f(x)$ in $\mathbb{R}$ is characterized by Schoenberg's theorem (Theorem 1 of Schoenberg (1938) or Section 10 of Stewart (1976)), that is

$$K(X_1,X_2)=f\left(\frac{\|X_1-X_2\|}{\gamma_X}\right)=\int_0^\infty \cos(t\|X_1-X_2\|)d\Lambda(t),\quad \forall X_1,X_2\in\mathbb{R}^p,$$

where Λ is a finite non-negative Borel measure on $[0,\infty)$. For $\alpha\sim N(0_p,I_p)$ and $Z\sim N(0,1)$, it follows from given X_1, X_2, $\alpha^\top(X_1-X_2)\overset{d}{=}\|X_1-X_2\|Z$ that

$$\begin{aligned}
\widetilde{K}(X_1,X_2)&=E\left\{K\left(\alpha^{\mathrm{T}}X_1,\alpha^{\mathrm{T}}X_2\right)\mid X_1,X_2\right\}\\
&=E\left[\int_0^\infty \cos\{t\|X_1-X_2\|Z\}d\Lambda(t)\right]\\
&=\int_0^\infty\sum_{j=0}^\infty\frac{(-1)^j}{(2j)!}(t\|X_1-X_2\|)^{2j}E(Z^{2j})d\Lambda(t)\\
&=\int_0^\infty\sum_{j=0}^\infty\frac{(-1)^j}{(2j)!}(t\|X_1-X_2\|)^{2j}(2j-1)!!d\Lambda(t)\\
&=\int_0^\infty\sum_{j=0}^\infty\frac{(-1)^j}{2^j j!}(t\|X_1-X_2\|)^{2j}d\Lambda(t)\\
&=\int_0^\infty \exp\{-(t\|X_1-X_2\|)^2/2\}d\Lambda(t) \qquad \text{(S3.8)}
\end{aligned}$$

where the third equality follows from Fubini's theorem. For fixed $t > 0$, $\exp\{-(t\|X_1 - X_2\|)^2/2\}$ is the Gaussian kernel which is characteristic. From Proposition 5 of Sriperumbudur et al. (2011), the characteristic distance-based kernel is equivalent to that of integrally strictly positive definite property, that is for every finite signed Radon measure ω on $\mathbb{R}^p$,

$$\int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \exp\{-(t\|x-y\|)^2/2\} d\omega(x) d\omega(y) > 0. \tag{S3.9}$$

Hence from (S3.8), (S3.9) and Fubini's theorem, we have

$$\int_{\mathbb{R}^p}\int_{\mathbb{R}^p} \widetilde{K}(x,y) d\omega(x) d\omega(y) > 0.$$

Thus using Proposition 5 of Sriperumbudur et al. (2011) again, $\widetilde{K}(X_1, X_2)$ is characteristic. This concludes the proof. □

S4. Proof of Proposition 2

From the representation of KPIC(X, Y) in (4.13), it is clear to see that KPIC$(X, Y) \geq 0$ using the fact that HSIC$(X_{1,\mathcal{S}_1}, Y_{1,\mathcal{S}_2}) \geq 0$. Suppose X and Y are independent. From Proposition 1, HSIC$(X_{1,\mathcal{S}_1}, Y_{1,\mathcal{S}_2})$ completely measures the dependence of $X_{1,\mathcal{S}_1}$ and $Y_{1,\mathcal{S}_2}$. Then we obtain HSIC$(X_{1,\mathcal{S}_1}, Y_{1,\mathcal{S}_2}) = 0$ for every $\mathcal{S}_1 \subset \{1, 2, \ldots, p\}$ and $\mathcal{S}_2 \subset \{1, 2, \ldots, q\}$. Thus we have KPIC$(X, Y) = 0$ from the fact that KPIC is the weighted sum of HSIC between the subvectors of X and Y.

On the other hand, from the representation in (4.13), KPIC$(X, Y) = 0$ implies HSIC$(X, Y) = 0$. The characteirstic property of HSIC shows that X and Y are independent. This completes the proof.

S5. Proof of Theorem 1

Recall that $U_{n,k}$ in (3.7) is a generalized U-statistic of with degree (4,1). For $\{z_i = (X_i, Y_i), i = 1, \ldots, n\}$, we define

$$h^{j,0}(z_1, \ldots, z_i) = E\{h(z_1, z_2, z_3, z_4; \alpha_1, \beta_1) \mid z_1, \ldots, z_j\},$$

$$h^{j,1}(z_1, \ldots, z_i, \alpha_1, \beta_1) = E\{h(z_1, z_2, z_3, z_4; \alpha_1, \beta_1) \mid z_1, \ldots, z_j, \alpha_1, \beta_1\},$$

where $j = 1, 2, 3, 4$. By direct calculation,

$$\begin{aligned} h^{1,0}(z_1) =& \frac{1}{2}\Big\{E\left(K_{12}^{\alpha_1} L_{12}^{\beta_1} \mid z_1\right) + E\left(K_{23}^{\alpha_1} L_{23}^{\beta_1}\right) \\ &+ E\left(K_{12}^{\alpha_1} L_{34}^{\beta_1} \mid z_1\right) + E\left(K_{34}^{\alpha_1} L_{12}^{\beta_1} \mid z_1\right) \\ &- E\left(K_{12}^{\alpha_1} L_{13}^{\beta_1} \mid z_1\right) - E\left(K_{12}^{\alpha_1} L_{23}^{\beta_1} \mid z_1\right) \\ &- E\left(K_{23}^{\alpha_1} L_{12}^{\beta_1} \mid z_1\right) - E\left(K_{23}^{\alpha_1} L_{24}^{\beta_1}\right)\Big\} \end{aligned}$$

$$\begin{aligned} h^{0,1}(\alpha_1, \beta_1) =& E\left(K_{12}^{\alpha_1} L_{12}^{\beta_1} \mid \alpha_1, \beta_1\right) + E\left(K_{12}^{\alpha_1} L_{34}^{\beta_1} \mid \alpha_1, \beta_1\right) \\ &- 2E\left(K_{12}^{\alpha_1} L_{13}^{\beta_1} \mid \alpha_1, \beta_1\right) \\ =& E\left\{d_K^{\alpha_1}(X_1, X_2)\, d_L^{\beta_1}(Y_1, Y_2) \mid \alpha_1, \beta_1\right\}, \end{aligned}$$

$$h^{1,1}(z_1,\alpha_1,\beta_1) = \frac{1}{2}\Big\{E(d_K^{\alpha_1}(X_1,X_2)\,d_L^{\beta_1}(Y_1,Y_2)\mid z_1,\alpha_1,\beta_1) + E(d_K^{\alpha_1}(X_3,X_4)\,d_L^{\beta_1}(Y_3,Y_4)\mid \alpha_1,\beta_1)\Big\}.$$

Through similar calculations, it can be verified that under the independence of X and Y,

$$h^{1,0}(z_1) = 0, \quad h^{0,1}(\alpha_1,\beta_1) = 0, \quad h^{1,1}(z_1,\alpha_1,\beta_1) = 0 \tag{S5.10}$$

$$h^{2,0}(z_1,z_2) = \frac{1}{6}U(X_1,X_2)\,V(Y_1,Y_2), \tag{S5.11}$$

$$h^{2,1}(z_1,z_2,\alpha_1,\beta_1) = \frac{1}{6}d_K^{\alpha_1}(X_1,X_2)\,d_L^{\beta_1}(Y_1,Y_2). \tag{S5.12}$$

Thus by the Hoeffding decomposition (e.g. Chapter 2 of Lee (1990)) and (S5.10)-(S5.12), when X and Y are independent, we have

$$U_{n,k} = W_n + R_{n,k}, \tag{S5.13}$$

where $W_n = \binom{n}{2}^{-1}\sum_{i<j}U(X_i,X_j)V(Y_i,Y_j)$. According to the expression of variance of generalized U-statistic (e.g. Chapter 2 of Lee (1990)) and Jensen inequality, we obtain

$$\text{var}(R_{n,k}) \le C(n^{-2}k^{-1} + n^{-3})E\{h(z_1,z_2,z_3,z_4;\alpha,\beta)\}^2 \tag{S5.14}$$

for some constant $C > 0$. By Lemma 2, it holds that

$$\text{var}(R_{n,k}) \le C(n^{-1} + k^{-1})E\{d_K^{\alpha}(X_1,X_2)^2 d_L^{\beta}(Y_1,Y_2)^2\}. \tag{S5.15}$$

Denote $\psi(z_i, z_j) = U(X_i, X_j)V(Y_i, Y_j)$ the kernel of W_n. Denote by $\varphi(z_i, z_j)$ the normalized kernel defined via

$$\varphi(z_i, z_j) = \sqrt{\frac{n(n-1)}{2}} \frac{U(X_i, X_j)V(Y_i, Y_j)}{\sigma_1}.$$

From (S5.13) we can deduce that

$$\frac{U_{n,k}}{\{\mathrm{var}(W_n)\}^{1/2}} = \frac{W_n}{\{\mathrm{var}(W_n)\}^{1/2}} + \frac{R_{n,k}}{\{\mathrm{var}(W_n)\}^{1/2}} = J_n + L_{n,k}, \qquad \text{(S5.16)}$$

where $J_n = \sum_{i<j}^{n} \varphi(z_i, z_j)$ and $L_{n,k} = R_{n,k}/\{\mathrm{var}(W_n)\}^{1/2}$. From the definition of Wasserstein distance and (S5.15), we obtain

$$\begin{aligned} d_{\mathcal{W}}(J_n + L_{n,k}, Z) &\le d_{\mathcal{W}}(J_n, Z) + \{\mathrm{var}(L_{n,k})\}^{1/2} \\ &\le d_{\mathcal{W}}(J_n, Z) + C(n^{-1} + k^{-1})^{1/2}\sigma_1^{-1}[E\{d_K^{\alpha}(X_1, X_2)^2 d_L^{\beta}(Y_1, Y_2)^2\}]^{1/2}. \end{aligned} \qquad \text{(S5.17)}$$

For the degenerate U-statistic J_n of degree 2, the explicit bound on the normal approximation can be derived using Stein's method of exchangeable pair. By Theorem 3.3 of Döbler and Peccati (2019), we have

$$\begin{aligned} d_{\mathcal{W}}(J_n, Z) \le C \Big\{ & n^{-1/2} + \sigma_1^{-2}(E[E\{\psi(z_1, z_2)\psi(z_1, z_3) \mid z_2, z_3\}]^2)^{1/2} \\ & + n^{-1/2}\sigma_1^{-2}[E\{\psi(z_1, z_2)^4\}]^{1/2} \Big\} \\ = C \Big(& n^{-1/2} + \sigma_1^{-2}[E\{g_K(X_1, X_2, X_3, X_4) g_L(Y_1, Y_2, Y_3, Y_4)\}]^{1/2} \\ & + n^{-1/2}\sigma_1^{-2}[E\{U(X_1, X_2)^4 V(Y_1, Y_2)^4\}]^{1/2} \Big), \end{aligned}$$

where C in each step represents a sufficiently large positive constant, and the equality follows from the independence of X and Y and the fact that

$$
\begin{aligned}
& E[E\{U(X_1,X_2)U(X_1,X_3) \mid X_2,X_3\}]^2 \\
&= E[E\{U(X_1,X_2)U(X_1,X_3)U(X_2,X_4)U(X_3,X_4) \mid X_2,X_3\}] \\
&= E\{U(X_1,X_2)U(X_1,X_3)U(X_2,X_4)U(X_3,X_4)\} \\
&= E\{g_K(X_1,X_2,X_3,X_4)\},
\end{aligned}
$$

similar result applies to $E[E\{V(Y_1,Y_2)V(Y_1,Y_3) \mid Y_2,Y_3\}]^2$. Combining (S5.17) and the above derivations, we complete the proof.

S6. Proof of Theorem 2

To prove Theorem 2, it suffices to show $E(\widehat{\sigma}_1^2) = \sigma_1^2\{1+o(1)\}$ and $\mathrm{var}(\widehat{\sigma}_1^2) = o(\sigma_1^4)$. Denote $\tilde{A}_{ij} = E(A_{ijr} \mid X_i, X_j)$ and $\tilde{B}_{ij} = E(B_{ijr} \mid Y_i, Y_j)$. Under the independence of X and Y, we have

$$
\begin{aligned}
E(\widehat{\sigma}_1^2) &= \frac{k-1}{k}\frac{1}{n(n-3)}\sum_{i\neq j} E(\tilde{A}_{ij}^2)E(\tilde{B}_{ij}^2) + \frac{1}{k}\frac{1}{n(n-3)}\sum_{i\neq j} E(A_{ijs}^2)E(B_{ijs}^2) \\
&= \sigma_1^2(1+o(1)) + E\left\{d_K^{\alpha}(X_1,X_2)\, d_L^{\beta}(Y_1,Y_2)\right\}^2\{1+o(1)\} \\
&= \sigma_1^2\{1+o(1)\},
\end{aligned}
$$

where the last equality follows from Condition (A1). On the other hand, by direct calculation, we have

$$\begin{aligned}\mathrm{var}(\widehat{\sigma}_1^2) &= \mathrm{var}\{E(\widehat{\sigma}_1^2 \mid z_1,\ldots,z_n)\} + E\{\mathrm{var}(\widehat{\sigma}_1^2 \mid z_1,\ldots,z_n)\} \\ &= O(n^{-4})\left\{\sum_{i<j}^n \mathrm{var}(\widetilde{A}_{ij}^2\widetilde{B}_{ij}^2) + \sum_{i<j<\ell}^n \mathrm{cov}(\widetilde{A}_{ij}^2\widetilde{B}_{ij}^2, \widetilde{A}_{i\ell}^2\widetilde{B}_{i\ell}^2)\right. \\ &\qquad \left. + \sum_{i<j,l<q,\{i,j\}\cap\{l,q\}=\phi}^n \mathrm{cov}(\widetilde{A}_{ij}^2\widetilde{B}_{ij}^2, \widetilde{A}_{\ell q}^2\widetilde{B}_{\ell q}^2)\right\}.\end{aligned}$$

Denote $\widetilde{K}_{ij} = E(K_{ij}^{\alpha_r} \mid X_i, X_j)$, $\bar{K}_{ij} = \widetilde{K}_{ij} - E(\widetilde{K}_{il} \mid X_i) - E(\widetilde{K}_{kj} \mid X_j) + E(\widetilde{K}_{kl})$ and $\widetilde{L}_{ij} = E(L_{ij}^{\beta_r} \mid Y_i, Y_j)$, $\bar{L}_{ij} = \widetilde{L}_{ij} - E(\widetilde{L}_{il} \mid Y_i) - E(\widetilde{L}_{kj} \mid Y_j) + E(\widetilde{L}_{kl})$. We can write $\widetilde{A}_{ij}$ as four uncorrelated summands

$$\begin{aligned}\widetilde{A}_{ij} &= \widetilde{K}_{ij} - \frac{1}{n-2}\sum_{s=1}^n \widetilde{K}_{sj} - \frac{1}{n-2}\sum_{t=1}^n \widetilde{K}_{it} + \frac{1}{(n-1)(n-2)}\sum_{s,t=1}^n \widetilde{K}_{st} \\ &= \frac{n-3}{n-1}\widetilde{K}_{ij} - \frac{n-3}{(n-1)(n-2)}\sum_{l\notin\{i,j\}} \widetilde{K}_{il} - \frac{n-3}{(n-1)(n-2)}\sum_{k\notin\{i,j\}} \widetilde{K}_{kj} \\ &\quad + \frac{1}{(n-1)(n-2)}\sum_{k,l\notin\{i,j\},k\neq l} \widetilde{K}_{kl} \\ &= \frac{n-3}{n-1}\bar{K}_{ij} - \frac{n-3}{(n-1)(n-2)}\sum_{l\notin\{i,j\}} \bar{K}_{il} - \frac{n-3}{(n-1)(n-2)}\sum_{k\notin\{i,j\}} \bar{K}_{kj} \\ &\quad + \frac{2}{(n-1)(n-2)}\sum_{k,l\notin\{i,j\},k<l} \bar{K}_{kl}.\end{aligned}$$

Similar result applies to $\widetilde{B}_{ij}$ can obtain

$$\begin{aligned}\widetilde{B}_{ij} &= \frac{n-3}{n-1}\bar{L}_{ij} - \frac{n-3}{(n-1)(n-2)}\sum_{l\notin\{i,j\}} \bar{L}_{il} - \frac{n-3}{(n-1)(n-2)}\sum_{k\notin\{i,j\}} \bar{L}_{kj} \\ &\quad + \frac{2}{(n-1)(n-2)}\sum_{k,l\notin\{i,j\},k<l} \bar{L}_{kl},\end{aligned}$$

Following the proof of Lemma D.3 in the supplement of Chakraborty and Zhang (2021), under Conditions (A2)–(A3), $\mathrm{var}(\widehat{\sigma}_U^2) = o(\sigma_U^4)$. This completes the proof.

S7. Proof of Theorem 3

It is known that $U_{n,k} - \mathrm{KPIC}(X,Y)$ can be decomposed as a weighted sum of degenerate U-statistics according to Hoeffding decomposition (see e.g., Lee (1990)), that is $U_{n,k} - \mathrm{KPIC}(X,Y) = \bar{W}_n + \bar{R}_{n,k}$ where $\bar{W}_n = \binom{n}{2}^{-1}\sum_{i<j}\{U(X_i,X_j)V(Y_i,Y_j) - \mathrm{KPIC}(X,Y)\}$ is the centered version of W_n, and $\bar{R}_{n,k}$ is the remainder term. By the variance expression of $U_{n,k} - \mathrm{KPIC}(X,Y)$ (e.g. Chapter 2 of Lee (1990)) and Jensen inequality, the variance of $\bar{R}_{n,k}$ is controlled by

$$\begin{aligned}\mathrm{var}(\bar{R}_{n,k}) \le C\Bigg(&\frac{1}{n}\mathrm{var}[E\{G(z_1,z_2,\alpha_1,\beta_1)\mid z_1\}] + \frac{1}{nk}\mathrm{var}[E\{G(z_1,z_2,\alpha_1,\beta_1)\mid z_1,\alpha_1,\beta_1\}]\\ &+\frac{1}{k}\mathrm{var}[E\{G(z_1,z_2,\alpha_1,\beta_1)\mid \alpha_1,\beta_1\}] + (\frac{1}{n^2k}+\frac{1}{n^3})\,\mathrm{var}\{h(z_1,z_2,z_3,z_4;\alpha_1,\beta_1)^2\}\Bigg).\end{aligned} \tag{S7.18}$$

By Lemma 1, we obtain

$$\begin{aligned}\mathrm{var}\{h(z_1,z_2,z_3,z_4;\alpha_1,\beta_1)^2\} \le C\Big[&E\left\{d_K^{\alpha}(X_1,X_2)\,d_L^{\beta}(Y_1,Y_2)\right\}^2 + E\left\{d_K^{\alpha}(X_1,X_2)d_L^{\beta}(Y_1,Y_3)\right\}^2\\ &+E\left\{d_K^{\alpha}(X_1,X_2)\right\}^2 E\{d_L^{\beta}(Y_1,Y_3)\}^2\Big].\end{aligned} \tag{S7.19}$$

Combining (S7.18), (S7.19), Condition (A1) and Conditions (A4)–(A5), we can deduce that $\bar{R}_{n,k} = o_p(\sigma_1)$. Following the proof of Theorem 1, under Conditions (A2)–(A3), we can deduce that

$$(U_{n,k} - \mathrm{KPIC}(X,Y))/\sigma_1 \xrightarrow{d} N(0,1).$$

This completes the proof.

S8. Proof of Theorem 4

Proof. The proof of Theorem 4 mainly depends on Lemmas 3–4, which are related to the bounds of KPIC and σ_1^2. It follows from Lemma 3–4, $\gamma_1 = \gamma_2 = 1 \wedge (c_n n^{1/2} p^{-1})$ for $c_n \to 0$ that

$$\frac{n\,\mathrm{KPIC}}{\sigma_1} \geq \frac{Cn}{p^2\gamma_1\gamma_2} \to \infty.$$

This concludes the proof.

□

S9. Proof of Lemmas

Lemma 1. *For any random vectors X and Y, we have*

$$h\left(z_1, z_2, z_3, z_4; \alpha, \beta\right)$$
$$= \frac{1}{24} \sum_{(t,u,v,w)}^{(1,2,3,4)} \left\{ d_K^\alpha\left(X_t, X_u\right) d_L^\beta\left(Y_t, Y_u\right) + d_K^\alpha\left(X_t, X_u\right) d_L^\beta\left(Y_v, Y_w\right) - 2 d_K^\alpha\left(X_t, X_u\right) d_L^\beta\left(Y_t, Y_v\right) \right\}$$

Proof. Recall that $h\left(z_1, z_2, z_3, z_4; \alpha, \beta\right)$ is the kernel function based on the projected data $\Big\{\left(\alpha_r^{\mathrm{T}} X_i, \beta_r^{\mathrm{T}} Y_i\right), i=1, \ldots, n; r=1, \ldots, k\Big\}$, namely

$$h\left(z_1, z_2, z_3, z_4; \alpha, \beta\right)=\frac{1}{24} \sum_{(t, u, v, w)}^{(1,2,3,4)}\Big(K_{tu}^{\alpha} L_{tu}^{\beta}+K_{tu}^{\alpha} L_{vw}^{\beta}-2 K_{tu}^{\alpha} L_{tv}^{\beta}\Big). \quad \text{(S9.20)}$$

Define

$$a(X_i, X_j)=K_{ij}^{\alpha}-E(K_{ij}^{\alpha} \mid \alpha) \quad a(X_i)=E\left\{a(X_i, X_j) \mid \alpha, X_i\right\}$$

$$b(Y_i, Y_j)=L_{ij}^{\beta}-E(L_{ij}^{\beta} \mid \beta) \quad b(Y_i)=E\left\{b(Y_i, Y_j) \mid \beta, Y_i\right\},$$

then we can decompose the corresponding terms in (S9.20)

$$\begin{aligned} K_{tu}^{\alpha} L_{tu}^{\beta}= & a(X_t, X_u) b(Y_t, Y_u)+a(X_t, X_u) E(L_{tu}^{\beta} \mid \beta) \\ & +b(Y_t, Y_u) E\left(K_{tu}^{\alpha} \mid \alpha\right)+E\left(K_{tu}^{\alpha} \mid \alpha\right) E(L_{tu}^{\beta} \mid \beta) \end{aligned}$$

$$\begin{aligned} K_{tu}^{\alpha} L_{vw}^{\beta}= & a(X_t, X_u) b(Y_v, Y_w)+a(X_t, X_u) E\left(L_{vw}^{\beta} \mid \beta\right) \\ & +b(Y_v, Y_w) E\left(K_{tu}^{\alpha} \mid \alpha\right)+E\left(K_{tu}^{\alpha} \mid \alpha\right) E\left(L_{vw}^{\beta} \mid \beta\right) \end{aligned}$$

$$\begin{aligned} K_{tu}^{\alpha} L_{tv}^{\beta}= & a(X_t, X_u) b(Y_t, Y_v)+a(X_t, X_u) E(L_{tv}^{\beta} \mid \beta) \\ & +b(Y_t, Y_v) E(K_{tu}^{\alpha} \mid \alpha)+E\left(K_{tu}^{\alpha} \mid \alpha\right) E(L_{tv}^{\beta} \mid \beta). \end{aligned}$$

We can observe that

$$\sum_{(t, u, v, w)}^{(1,2,3,4)}\Big\{a(X_t, X_u) E(L_{tu}^{\beta} \mid \beta)+a(X_t, X_u) E(L_{vw}^{\beta} \mid \beta)-2 a(X_t, X_u) E(L_{tv}^{\beta} \mid \beta)\Big\}=0.$$

This is also true for other terms, thus (S9.20) can be decomposed as

$$\begin{aligned} & h\left(z_1, z_2, z_3, z_4; \alpha, \beta\right) \\ & =\frac{1}{24} \sum_{(t, u, v, w)}^{(1,2,3,4)}\left\{a(X_t, X_u) b(Y_t, Y_u)+a(X_t, X_u) b(Y_v, Y_w)-2 a(X_t, X_u) b(Y_t, Y_v)\right\}. \end{aligned} \quad \text{(S9.21)}$$

Since $a(X_i, X_j) = d_K^\alpha(X_i, X_j) + a(X_i) + a(X_j)$, $b(Y_i, Y_j) = d_L^\beta(Y_i, Y_j) + b(Y_i) + b(Y_j)$, it can be calculated that

$$\begin{aligned} a(X_t, X_u)b(Y_t, Y_u) =& d_K^\alpha(X_t, X_u)\, d_L^\beta(Y_t, Y_u) + d_K^\alpha(X_t, X_u)\, b(Y_t) + d_K^\alpha(X_t, X_u)\, b(Y_u) \\ &+ a(X_t) d_L^\beta(Y_t, Y_u) + a(X_t)b(Y_t) + a(X_t)b(Y_u) \\ &+ a(X_u) d_L^\beta(Y_t, Y_u) + a(X_u)b(Y_t) + a(X_u)b(Y_u) \end{aligned}$$

$$\begin{aligned} a(X_t, X_u)b(Y_v, Y_w) =& d_K^\alpha(X_t, X_u)\, d_L^\beta(Y_v, Y_w) + d_K^\alpha(X_t, X_u)\, b(Y_v) + d_K^\alpha(X_t, X_u)\, b(Y_w) \\ &+ a(X_t) d_L^\beta(Y_v, Y_w) + a(X_t)b(Y_v) + a(X_t)b(Y_w) \\ &+ a(X_u) d_L^\beta(Y_v, Y_w) + a(X_u)b(Y_v) + a(X_u)b(Y_w) \end{aligned}$$

$$\begin{aligned} a(X_t, X_u)b(Y_t, Y_v) =& d_K^\alpha(X_t, X_u)\, d_L^\beta(Y_t, Y_v) + d_K^\alpha(X_t, X_u)\, b(Y_t) + d_K^\alpha(X_t, X_u)\, b(Y_v) \\ &+ a(X_t) d_L^\beta(Y_t, Y_v) + a(X_t)b(Y_t) + a(X_t)b(Y_v) \\ &+ a(X_u) d_L^\beta(Y_t, Y_v) + a(X_u)b(Y_t) + a(X_u)b(Y_v). \end{aligned}$$

We can observe that

$$\begin{aligned} \sum_{(t,u,v,w)}^{(1,2,3,4)} \{& d_K^\alpha(X_t, X_u)\, b(Y_t) + d_K^\alpha(X_t, X_u)\, b(Y_u) + d_K^\alpha(X_t, X_u)\, b(Y_v) + d_K^\alpha(X_t, X_u)\, b(Y_w) \\ &- 2d_K^\alpha(X_t, X_u)\, b(Y_t) - 2d_K^\alpha(X_t, X_u)\, b(Y_v)\} = 0. \end{aligned}$$

Similar observation can also be implied other terms. Then we can show that (S9.21) can be written as

$$\begin{aligned} &h(z_1, z_2, z_3, z_4; \alpha, \beta) \\ &= \frac{1}{24} \sum_{(t,u,v,w)}^{(1,2,3,4)} \left\{ d_K^\alpha(X_t, X_u)\, d_L^\beta(Y_t, Y_u) + d_K^\alpha(X_t, X_u)\, d_L^\beta(Y_v, Y_w) - 2d_K^\alpha(X_t, X_u)\, d_L^\beta(Y_t, Y_v) \right\}. \end{aligned} \tag{S9.22}$$

This completes the proof. □

Lemma 2. *Under the independence of X and Y, we have*

$$E\left\{h\left(z_1,z_2,z_3,z_4;\alpha,\beta\right)\right\}^2=\frac{1}{2}E\left\{d_K^{\alpha}\left(X_1,X_2\right)d_L^{\beta}\left(Y_1,Y_2\right)\right\}^2$$

Proof. Follow Lemma 1, we can deduce that

$$\begin{aligned}
&E\left(h\left(z_1,z_2,z_3,z_4;\alpha,\beta\right)\right)^2\\
=&\frac{1}{576}\sum_{(t,u,v,w)}^{(1,2,3,4)}\sum_{(i,j,q,r)}^{(1,2,3,4)}\left\{d_K^{\alpha}\left(X_t,X_u\right)d_L^{\beta}\left(Y_t,Y_u\right)+d_K^{\alpha}\left(X_t,X_u\right)d_L^{\beta}\left(Y_v,Y_w\right)-2d_K^{\alpha}\left(X_t,X_u\right)d_L^{\beta}\left(Y_t,Y_v\right)\right\}\\
&\left\{d_K^{\alpha}\left(X_i,X_j\right)d_L^{\beta}\left(Y_i,Y_j\right)+d_K^{\alpha}\left(X_i,X_j\right)d_L^{\beta}\left(Y_q,Y_r\right)-2d_K^{\alpha}\left(X_i,X_j\right)d_L^{\beta}\left(Y_i,Y_q\right)\right\}.
\end{aligned}$$

Since $d_K^{\alpha}\left(X_i,X_j\right)$ and $d_L^{\beta}\left(Y_i,Y_j\right)$ statisfy the double-centered property, under the independence of X and Y, it holds that

$$\begin{aligned}
E\left(h\left(z_1,z_2,z_3,z_4;\alpha,\beta\right)\right)^2=&\frac{1}{576}(48+48+192)E\left\{d_K^{\alpha}\left(X_1,X_2\right)d_L^{\beta}\left(Y_1,Y_2\right)\right\}^2\\
=&\frac{1}{2}E\left\{d_K^{\alpha}\left(X_1,X_2\right)d_L^{\beta}\left(Y_1,Y_2\right)\right\}^2.
\end{aligned}$$

Thus we complete the proof. □

Lemma 3. *Under the assumptions in Theorem 3, we have* $\mathrm{KPIC}\geq Cp^{-2}\gamma_1^{-1}\gamma_2^{-1}$ *for some positive constant C.*

Proof. In the remaining proof, we only need to use the bounded derivatives of the kernels. Without losing generality, we assume that $K(X_1,X_2)$ and $L(Y_1,Y_2)$ are Gaussian kernels. Denote $\tau_{\alpha,X}^2=E\|\alpha^{\top}X_1\|^2=O(\gamma_1 p)$ and $\tau_{\beta,Y}^2=E\|\beta^{\top}Y_1\|^2=O(\gamma_2 p)$. Let $\gamma_{\alpha,X}$ and $\gamma_{\beta,Y}$ denote the corresponding bandwidths satisfying $\gamma^2_{\alpha,X}/\tau^2_{\alpha,X}$ and $\gamma^2_{\beta,Y}/\tau^2_{\beta,Y}$ are bounded away from 0 and ∞. Let $F(x) = \exp(-x/2)$. Taking the Taylor expansion of $F(x)$ around $\tau^2_{\alpha,X}/\gamma^2_{\alpha,X}$, we obtain

$$\begin{aligned}
&K(\alpha^\top X_1, \alpha^\top X_2)\\
&= b_0 + b_1 \frac{\tau^2_{\alpha,X}}{\gamma^2_{\alpha,X}} \left\{ \frac{(\alpha^\top X_1 - \alpha^\top X_2)^2}{\tau^2_{\alpha,X}} - 1 \right\} + b_2 \frac{\tau^4_{\alpha,X}}{\gamma^4_{\alpha,X}} \left\{ \frac{(\alpha^\top X_1 - \alpha^\top X_2)^2}{\tau^2_{\alpha,X}} - 1 \right\}^2\\
&\quad + b_3 \frac{\tau^6_{\alpha,X}}{\gamma^6_{\alpha,X}} \left\{ \frac{(\alpha^\top X_1 - \alpha^\top X_2)^2}{\tau^2_{\alpha,X}} - 1 \right\}^3,
\end{aligned} \tag{S9.23}$$

where $b_0 = F(\tau^2_{\alpha,X}/\gamma^2_{\alpha,X})$, $b_1 = F^{(1)}(\tau^2_{\alpha,X}/\gamma^2_{\alpha,X})$, $b_2 = F^{(2)}(\tau^2_{\alpha,X}/\gamma^2_{\alpha,X})/2$ and $b_3 = F^{(3)}(\xi)/6$ for some $\xi > 0$. Similar result applies to $L(\beta^\top Y_1, \beta^\top Y_2)$ can obtain $c_0 = F(\tau^2_{\beta,Y}/\gamma^2_{\beta,Y})$, $c_1 = F^{(1)}(\tau^2_{\beta,Y}/\gamma^2_{\beta,Y})$, $c_2 = F^{(2)}(\tau^2_{\beta,Y}/\gamma^2_{\beta,Y})$ and $c_3 = F^{(3)}(\zeta)$ for some $\zeta > 0$. Define

$$W_{ij} = \frac{(\alpha^\top X_1 - \alpha^\top X_2)^2}{\tau^2_{\alpha,X}} - 1, V_{ij} = \frac{(\beta^\top Y_1 - \beta^\top Y_2)^2}{\tau^2_{\beta,Y}} - 1, \rho_{\alpha,X} = \frac{\tau^2_{\alpha,X}}{\gamma^2_{\alpha,X}}, \rho_{\beta,Y} = \frac{\tau^2_{\beta,Y}}{\gamma^2_{\beta,Y}}.$$

By invoking the bounded derivatives of $F(x)$ we have $C_0 \le b_i, c_i \le C_1$ for some positive constants C_0, C_1 and for each $0 \le i \le 3$. Combining the above expansions and Lemma 10 of Gao et al. (2021) yields that $\mathrm{KPIC}(X_1, Y_1) =$

$I_1 + I_2 + I_3 + I_4 + I_5$, where

$I_1 = b_1 c_1 \rho_{\alpha,X} \rho_{\beta,Y} \{E(W_{12}V_{12}) - 2E(W_{12}V_{13})\}$,

$I_2 = b_1 c_2 \rho_{\alpha,X} \rho^2_{\beta,Y} \{E(W_{12}V^2_{12}) - 2E(W_{12}V^2_{13})\} + b_2 c_1 \rho^2_{\alpha,X} \rho_{\beta,Y} \{E(W^2_{12}V_{12}) - 2E(W^2_{12}V_{13})\}$,

$I_3 = b_1 c_3 \rho_{\alpha,X} \rho^3_{\beta,Y} \{E(W_{12}V^3_{12}) - 2E(W_{12}V^3_{13})\} + b_3 c_1 \rho^3_{\alpha,X} \rho_{\beta,Y} \{E(W^3_{12}V_{12}) - 2E(W^3_{12}V_{13})\}$,

$I_4 = b_2 c_2 \rho^2_{\alpha,X} \rho^2_{\beta,Y} \{E(W^2_{12}V^2_{12}) - 2E(W^2_{12}V^2_{13}) + E(W^2_{12})E(V^2_{12})\}$,

$$I_5 = O\left\{\rho^2_{\alpha,X}\rho^3_{\beta,Y}\left(E\left|W_{12}\right|^5\right)^{2/5}\left(E\left|V_{12}\right|^5\right)^{3/5} + \rho^3_{\alpha,X}\rho^2_{\beta,Y}\left(E\left|W_{12}\right|^5\right)^{3/5}\left(E\left|V_{12}\right|^5\right)^{2/5}\right.$$
$$\left.+\rho^3_{\alpha,X}\rho^3_{\beta,Y}\left(E\left|W_{12}\right|^6\right)^{1/2}\left(E\left|V_{12}\right|^6\right)^{1/2}\right\}.$$

Assume the expectations of the components of X are zero. Denote by $\tilde{Y}_1 = Y_1 - E(Y_1)$ and $\tilde{Y}_2 = Y_2 - E(Y_2)$ the centered random variables and define

$$\alpha_1(X_1) = (\alpha^\top X_1)^2 - E(\alpha^\top X_1)^2, \alpha_2(X_1, X_2) = \alpha^\top X_1 \alpha^\top X_2,$$

$$\beta_1(Y_1) = (\beta^\top Y_1)^2 - E(\beta^\top Y_1)^2, \beta_2(Y_1, Y_2) = \beta^\top \tilde{Y}_1 \beta^\top \tilde{Y}_2.$$

Observe that

$$W_{12} = \tau^{-2}_{\alpha,X}\{\alpha_1(X_1)+\alpha_1(X_2)-2\alpha_2(X_1,X_2)\}, \quad V_{12} = \tau^{-2}_{\beta,Y}\{\beta_1(Y_1)+\beta_1(Y_2)-2\beta_2(Y_1,Y_2)\},$$

then we have

$$E(W_{12}V_{12}) = \tau^{-2}_{\alpha,X}\tau^{-2}_{\beta,Y}[2E\{\alpha_1(X_1)\beta_1(Y_1)\} + 4E\{\alpha_2(X_1,X_2)\beta_2(Y_1,Y_2)\}],$$

similarly, $E(W_{12}V_{13}) = \tau^{-2}_{\alpha,X}\tau^{-2}_{\beta,Y}E\{\alpha_1(X_1)\beta_1(Y_1)\}$. Thus

$$I_1 = 4b_1c_1\gamma^{-2}_{\alpha,X}\gamma^{-2}_{\beta,Y}E\{\alpha_2(X_1,X_2)\beta_2(Y_1,Y_2)\} = 4b_1c_1\gamma^{-2}_{\alpha,X}\gamma^{-2}_{\beta,Y}\gamma_1\gamma_2\sum_{i=1}^{p}\sum_{j=1}^{p}\{\mathrm{cov}(X_{1,i},Y_{1,j})\}^2.$$

Under the symmetry assumptions, we can deduce that $\mathrm{cov}(X_{1,i}, Y_{1,j}) = 0$ for $1 \leq i \leq p$, $1 \leq j \leq p$. Thus $I_1 = 0$.

For term I_2, by direct calculation, we have

$$\begin{aligned} I_2 = b_1 c_2 \gamma_{\alpha,X}^{-2} \gamma_{\beta,Y}^{-4} [2E\left\{\alpha_2\left(X_1, X_2\right) \beta_2^2\left(Y_1, Y_2\right)\right\} + E\left\{\alpha_2\left(X_1, X_2\right) \beta_1\left(Y_1\right) \beta_1\left(Y_2\right)\right\} \\ - 4E\left\{\alpha_2\left(X_1, X_2\right) \beta_1\left(Y_1\right) \beta_2\left(Y_1, Y_2\right)\right\}] \\ + b_2 c_1 \gamma_{\alpha,X}^{-4} \gamma_{\beta,Y}^{-2} [2E\left\{\beta_2\left(Y_1, Y_2\right) \alpha_2^2\left(X_1, X_2\right)\right\} + E\left\{\beta_2\left(Y_1, Y_2\right) \alpha_1\left(X_1\right) \alpha_1\left(X_2\right)\right\} \\ - 4E\left\{\beta_2\left(Y_1, Y_2\right) \alpha_1\left(X_2\right) \alpha_2\left(X_1, X_2\right)\right\}]. \end{aligned}$$

Since X_1 has symmetry distribution and $Y_{1,j} = g_j(X_{1,j})$ with symmetry function $g_j(x)$ for each $1 \leq j \leq p$, it follows that

$$\begin{aligned} E\left\{\alpha_2\left(X_1, X_2\right) \beta_2^2\left(Y_1, Y_2\right)\right\} &= E\left\{\left(X_1^\top \alpha\alpha^\top X_2\right) (\tilde{Y}_1^\top \beta\beta^\top \tilde{Y}_2)^2\right\} \\ &= 3\gamma_1\gamma_2 E\left\{\left(X_1^T X_2\right) (\tilde{Y}_1^T \tilde{Y}_2)^2\right\} \\ &= 3\gamma_1\gamma_2 E\left\{\left(-X_1^T X_2\right) (\tilde{Y}_1^T \tilde{Y}_2)^2\right\} \\ &= 0. \end{aligned}$$

Similarly, we can show that

$$\begin{aligned} &E\left\{\alpha_2\left(X_1, X_2\right) \beta_1\left(Y_1\right) \beta_1\left(Y_2\right)\right\} = 0, \\ &E\left\{\alpha_2\left(X_1, X_2\right) \beta_1\left(Y_1\right) \beta_2\left(Y_1, Y_2\right)\right\} = 0, \\ &E\left\{\beta_2\left(Y_1, Y_2\right) \alpha_1\left(X_2\right) \alpha_2\left(X_1, X_2\right)\right\} = 0. \end{aligned}$$

Besides, it holds that

$$E\left\{\beta_2\left(Y_1,Y_2\right)\alpha_2^2\left(X_1,X_2\right)\right\}=3\gamma_1\gamma_2\sum_{i,j}^{p}\sum_{k=1}^{p}\left\{E\left(X_{1,i}X_{1,j}\widetilde{Y}_{1,k}\right)\right\}^2\geq 0,$$

$$E\left\{\beta_2\left(Y_1,Y_2\right)\alpha_1(X_1)\alpha_1(X_2)\right\}=3\gamma_1\gamma_2\sum_{i=1}^{p}\left[\sum_{j=1}^{p}E\left\{\tilde{Y}_{1,i}\left(X_{1,j}^2-EX_{1,j}^2\right)\right\}\right]^2\geq 0.$$

Thus $I_2\geq 0$.

By similar but tedious calculations, under symmetry assumptions, we have

$$I_3=b_3c_1\gamma_{\alpha,X}^{-6}\gamma_{\beta,Y}^{-2}\left[-3E\left\{\beta_2\left(Y_1,Y_2\right)\alpha_1(X_1)\alpha_1^2\left(X_2\right)\right\}\right.$$
$$\left.-12E\left\{\beta_2\left(Y_1,Y_2\right)\alpha_1(X_1)\alpha_2^2\left(X_1,X_2\right)\right\}\right],$$

and

$$\begin{aligned}I_4=&b_2c_2\gamma_{\alpha,X}^{-4}\gamma_{\beta,Y}^{-4}\left[4E\left\{\alpha_2^2\left(X_1,X_2\right)\beta_2^2\left(Y_1,Y_2\right)\right\}+\left[E\left\{\alpha_1(X_1)\beta_1(Y_2)\right\}\right]^2\right.\\&+8E\left\{\alpha_2^2\left(X_1,X_2\right)\beta_2^2\left(Y_1,Y_3\right)\right\}+4E\left\{\alpha_2^2\left(X_1,X_2\right)\right\}E\left\{\beta_2^2\left(Y_1,Y_2\right)\right\}\\&+2E\left\{\alpha_2^2\left(X_1,X_2\right)\beta_1(Y_1)\beta_1(Y_2)\right\}-8E\left\{\alpha_2^2\left(X_1,X_2\right)\beta_1(Y_1)\beta_2\left(Y_1,Y_2\right)\right\}\\&+2E\left\{\alpha_1(X_1)\alpha_1(X_2)\beta_2^2\left(Y_1,Y_2\right)\right\}-4E\left\{\alpha_1(X_1)\alpha_1(X_2)\beta_1(Y_1)\beta_2\left(Y_1,Y_2\right)\right\}\\&\left.-8E\left\{\alpha_2^2\left(X_1,X_2\right)\beta_1\left(Y_3\right)\beta_2\left(Y_1,Y_3\right)\right\}\right].\end{aligned}$$

For term I_3, denote $\mathcal{D}(i)=\{(j,k,l):\max(|j-i|,|k-i|,|l-i|)\leq 3m+1\}$.

By the m-dependent structure,

$$\begin{aligned}&E\left\{\beta_2\left(Y_1,Y_2\right)\alpha_1(X_1)\alpha_1^2\left(X_2\right)\right\}\\=&15\gamma_1\gamma_2\sum_{i=1}^{p}\sum_{(j,k,l)\in\mathcal{D}(i)}E\left\{Y_{1,i}\left(X_{1,j}^2-EX_{1,j}^2\right)\right\}E\left\{Y_{1,i}\left(X_{1,k}^2-EX_{1,k}^2\right)\left(X_{1,l}^2-EX_{1,l}^2\right)\right\}\\=&O(\gamma_1\gamma_2d_1^8m^3p)\end{aligned}$$

and

$$
\begin{aligned}
&E\left\{\beta_2\left(Y_1, Y_2\right) \alpha_1(X_1) \alpha_2^2\left(X_1, X_2\right)\right\} \\
=&15 \gamma_1 \gamma_2 \sum_{i=1}^p \sum_{(j,k,l) \in \mathcal{D}(i)} E\left(Y_{1,i} X_{1,k} X_{1,l}\right) E\left\{Y_{1,i}\left(X_{1,j}^2-E X_{1,j}^2\right) X_{1,k} X_{1,l}\right\} \\
=&O(\gamma_1 \gamma_2 d_1^8 m^3 p).
\end{aligned}
$$

Thus $|I_3| = O(d_1^8 d_2^{-10} m^3 \gamma_1^{-2} p^{-3})$. For term I_4, it can be seen that $\left[E\left\{\alpha_1(X_1)\beta_1(Y_2)\right\}\right]^2 \geq 0$ and $E\left\{\alpha_2^2\left(X_1, X_2\right) \beta_2^2\left(Y_1, Y_3\right)\right\} \geq 0$. Moreover,

$$
\begin{aligned}
E\left\{\alpha_2^2\left(X_1, X_2\right) \beta_2^2\left(Y_1, Y_2\right)\right\} &= 9 \gamma_1 \gamma_2 \sum_{i,j,k,l=1}^p E\left\{X_{1,i} X_{1,j} Y_{1,k} Y_{1,l}\right\}^2 \\
&\geq 9 \gamma_1 \gamma_2 \sum_{|i-k|>m} \left\{E\left(X_{1,i}^2\right) E\left(Y_{1,k}^2\right)\right\}^2 \\
&\geq 9 d_2^8 \gamma_1 \gamma_2 p(p-2m),
\end{aligned}
$$

and

$$
\begin{aligned}
E\left\{\alpha_2^2\left(X_1, X_2\right)\right\} E\left\{\beta_2^2\left(Y_1, Y_2\right)\right\} &= 9 \gamma_1 \gamma_2 \sum_{i,j,k,l=1}^p \left\{E\left(X_{1,i} X_{1,j}\right)\right\}^2 \left\{E(\widetilde{Y}_{1,k} \widetilde{Y}_{1,l})\right\}^2 \\
&\geq 9 d_2 \gamma_1 \gamma_2^8 p^2.
\end{aligned}
$$

In the same fashion, we can deduce that

$$
\begin{aligned}
E\left\{\alpha_2^2\left(X_1, X_2\right) \beta_1(Y_1) \beta_1(Y_2)\right\} &= O(\gamma_1 \gamma_2 d_1^8 m^3 p), \\
E\left\{\alpha_2^2\left(X_1, X_2\right) \beta_1(Y_1) \beta_2\left(Y_1, Y_2\right)\right\} &= O(\gamma_1 \gamma_2 d_1^8 m^3 p), \\
E\left\{\alpha_1(X_1) \alpha_1(X_2) \beta_2^2\left(Y_1, Y_2\right)\right\} &= O(\gamma_1 \gamma_2 d_1^8 m^3 p), \\
E\left\{\alpha_1(X_1) \alpha_1(X_2) \beta_1(Y_1) \beta_2\left(Y_1, Y_2\right)\right\} &= O(\gamma_1 \gamma_2 d_1^8 m^3 p), \\
E\left\{\alpha_2^2\left(X_1, X_2\right) \beta_1\left(Y_3\right) \beta_2\left(Y_1, Y_3\right)\right\} &= O(\gamma_1 \gamma_2 d_1^8 m^3 p).
\end{aligned}
$$

As a consequence, there exists a constant C such that $I_4 \geq Cp^{-2}\gamma_1^{-1}\gamma_2^{-1}$. By similar statements and the m-dependent structure, we can obtain

$$I_5 = O\left(p^{-4}\gamma_1^{-2}\gamma_2^{-2}\right).$$

As a consequence, combining all the bounds above and the fact that $\gamma_i p \to \infty$ leads to $\mathrm{KPIC}(X_1, Y_1) \geq Cp^{-2}\gamma_1^{-1}\gamma_2^{-1}$ for some positive constant C. Thus we complete the proof.

□

Lemma 4. *Under the assumptions in Theorem 3, then $\sigma_1^2 \leq C$ for some positive constant C.*

Proof. Recall that $\sigma_1^2 = E\{U(X_1, X_2)^2\}E\{V(Y_1, Y_2)^2\}$. It follows from the definition of $U(X_1, X_2)$ that $E\{U(X_1, X_2)^2\}$ is equal to a weighted sum of $\mathrm{HSIC}(X_{1,\mathcal{S}_1}, X_{1,\mathcal{S}_2})$ with projective kernels

$$E\left\{K\left(\theta_{\mathcal{S}_1}^\top X_{1,\mathcal{S}_1}, \theta_{\mathcal{S}_1}^\top X_{2,\mathcal{S}_1}\right) \mid X_1, X_2\right\} = (1 + \|X_{1,\mathcal{S}_1} - X_{2,\mathcal{S}_1}\|^2/\gamma_{X,\mathcal{S}_1}^2)^{-1/2},$$

$$E\left\{K\left(\eta_{\mathcal{S}_2}^\top X_{1,\mathcal{S}_2}, \eta_{\mathcal{S}_2}^\top X_{2,\mathcal{S}_2}\right) \mid X_1, X_2\right\} = (1 + \|X_{1,\mathcal{S}_2} - X_{2,\mathcal{S}_2}\|^2/\gamma_{Y,\mathcal{S}_2}^2)^{-1/2}.$$

Suppose $X_{1,\mathcal{S}_1} \in \mathbb{R}^t$ and $X_{1,\mathcal{S}_2} \in \mathbb{R}^s$. With a slight abuse of notation, define $\tau_{X,\mathcal{S}_1}^2 = E(\|X_{1,\mathcal{S}_1} - X_{2,\mathcal{S}_1}\|^2)$, $\tau_{X,\mathcal{S}_2}^2 = E(\|X_{1,\mathcal{S}_2} - X_{2,\mathcal{S}_2}\|^2)$ and

$$W_{ij} = \frac{\|X_{1,\mathcal{S}_1} - X_{2,\mathcal{S}_1}\|^2}{\tau_{X,\mathcal{S}_1}^2} - 1, V_{ij} = \frac{\|X_{1,\mathcal{S}_2} - X_{2,\mathcal{S}_2}\|^2}{\tau_{X,\mathcal{S}_2}^2} - 1, \rho_{\mathcal{S}_1} = \frac{\tau_{X,\mathcal{S}_1}^2}{\gamma_{X,\mathcal{S}_1}^2}, \rho_{\mathcal{S}_2} = \frac{\tau_{X,\mathcal{S}_2}^2}{\gamma_{X,\mathcal{S}_2}^2}.$$

Similar to the proof of Lemma 3, we have $\rho_{\mathcal{S}_1} \xrightarrow{P} 1$ and $\rho_{\mathcal{S}_2} \xrightarrow{P} 1$. Using the Taylor expansion of $F(x) = (1+x)^{-1/2}$ around $\tau^2_{X,\mathcal{S}_1}/\gamma_{X,\mathcal{S}_1}$,

$$\left(1 + \frac{\|X_{1,\mathcal{S}_1} - X_{2,\mathcal{S}_1}\|^2}{\gamma_{X,\mathcal{S}_1}}\right)^{-1/2}$$
$$= b_0 + b_1 \frac{\tau^2_{X,\mathcal{S}_1}}{\gamma_{X,\mathcal{S}_1}} \left(\frac{\|X_{1,\mathcal{S}_1} - X_{2,\mathcal{S}_1}\|^2}{\tau^2_{X,\mathcal{S}_1}} - 1\right) + b_2 \frac{\tau^4_{X,\mathcal{S}_1}}{\gamma^4_{X,\mathcal{S}_1}} \left(\frac{\|X_{1,\mathcal{S}_1} - X_{2,\mathcal{S}_1}\|^2}{\tau^2_{X,\mathcal{S}_1}} - 1\right)^2,$$

where $b_0 = F(\tau^2_{X,\mathcal{S}_1}/\gamma_{X,\mathcal{S}_1})$, $b_1 = F^{(1)}(\tau^2_{X,\mathcal{S}_1}/\gamma_{X,\mathcal{S}_1})$, $b_2 = F^{(2)}(\xi)/2$ for some $\xi > 0$. Similar result applies to $(1 + \|X_{1,\mathcal{S}_2} - X_{2,\mathcal{S}_2}\|^2/\gamma^2_{\beta,Y})^{-1/2}$ can obtain $c_0 = F(\tau^2_{X,\mathcal{S}_2}/\gamma^2_{X,\mathcal{S}_2})$, $c_1 = F^{(1)}(\tau^2_{X,\mathcal{S}_2}/\gamma^2_{X,\mathcal{S}_2})$, $c_2 = F^{(2)}(\zeta)$ for some $\zeta > 0$. Following the proof of Lemma 3, we can obtain $\mathrm{HSIC}(X_{1,\mathcal{S}_1}, X_{2,\mathcal{S}_2}) = J_1 + J_2 + J_3$ where

$J_1 = b_1 c_1 \rho_{\mathcal{S}_1} \rho_{\mathcal{S}_2} \{E(W_{12}V_{12}) - 2E(W_{12}V_{13})\}$,

$J_2 = b_1 c_2 \rho_{\mathcal{S}_1} \rho^2_{\mathcal{S}_2} \{E(W_{12}V^2_{12}) - 2E(W_{12}V^2_{13})\} + b_2 c_1 \rho^2_{\mathcal{S}_1} \rho_{\mathcal{S}_2} \{E(W^2_{12}V_{12}) - 2E(W^2_{12}V_{13})\}$,

$J_3 = b_2 c_2 \rho^2_{\mathcal{S}_1} \rho^2_{\mathcal{S}_2} \{E(W^2_{12}V^2_{12}) - 2E(W^2_{12}V^2_{13}) + E(W^2_{12})E(V^2_{12})\}$.

Observe that

$$W_{12} = \tau^{-2}_{X,\mathcal{S}_1} \{\alpha_1(X_1) + \alpha_1(X_2) - 2\alpha_2(X_1, X_2)\},$$

$$V_{12} = \tau^{-2}_{X,\mathcal{S}_2} \{\alpha_1(X_{1,\mathcal{S}_2}) + \alpha_1(X_{2,\mathcal{S}_2}) - 2\alpha_2(X_{1,\mathcal{S}_2}, X_{2,\mathcal{S}_2})\}.$$

Then there exists a constant K such that

$$E(W_{12}V_{12}) - 2E(W_{12}V_{13}) = 4\tau^{-2}_{X,\mathcal{S}_1}\tau^{-2}_{X,\mathcal{S}_2} \sum_{i=1}^{t} \sum_{j=1}^{s} \{E(X_{1,i}X_{1,j})\}^2 \le K$$

Thus $J_1 \le C$ for some sufficiently large constant. By similar but tedious calculations, we can obtain $J_2 = O(1)$ and $J_3 = O(1)$. Combining the above bounds leads to

$$
\begin{aligned}
E\{U(X_1, X_2)^2\} &\leq C \sum_{t=1}^{p} \sum_{s=1}^{p} \binom{p}{t} \binom{p}{s} \gamma_1^t \gamma_1^s (1-\gamma_1)^{p-t} (1-\gamma_1)^{p-s} \\
&= C \left\{ \sum_{t=1}^{p} \binom{p}{t} \gamma_1^t (1-\gamma_1)^{p-t} \right\}^2 \\
&= C.
\end{aligned}
$$

where C in each step represents a sufficiently large constant. Similar statement applies to $E\{V(Y_1, Y_2)^2\}$. This completes the proof.

□

S10. Additional simulation results

S10.1 Number of random projections

We first examine how sensitive the proposed test is to the choices of projection numbers k. We mainly focus on the following two scenarios:

Example 1. Let $\Sigma_p = (0.5^{|i-j|}) \in \mathbb{R}^{p \times p}$ and $r, \epsilon \in (0, 1]$. Generate i.i.d. samples from the following models for $i = 1, \ldots, n$.

(i) $X_i = (X_{i,1}, \ldots, X_{i,p})^\top \sim N(0_p, \Sigma_p)$, $Y_i = (Y_{i,1}, \ldots, Y_{i,p})^\top$, where $Y_{i,j} = X_{i,j}^2$ for $1 \leq j \leq rp$, and $(Y_{i,rp+1}, \ldots, Y_{i,p}) \sim N\left(0_{(1-r)p}, \Sigma_{(1-r)p}\right)$;

(ii) $X_i = (X_{i,1}, \ldots, X_{i,p})^\top \sim N(0_p, \Sigma_p)$, $Z_i = (Z_{i,1}, \ldots, Z_{i,p})$, $\{Z_{i,j}\}_{j=1}^p$ are i.i.d. standard Cauchy random variables, $Y_i = (Y_{i,1}, \ldots, Y_{i,p})^\top$, where $Y_{i,j} = \epsilon X_{i,j}^2 + (1-\epsilon) Z_{i,j}^2$ for $j = 1, \ldots, p$.

In the above scenarios, the parameters r and ϵ can represent the strength of the dependence between X and Y, where r represents the ratio of related components and ϵ represents the dependence level of components, respectively. We set $r = 0$, 0.3, 0.7, 1 and $\epsilon = 0$, 0.3, 0.7, 1, where $r = 0$ and $\epsilon = 0$ indicate that X and Y are independent, $r = 1$ and $\epsilon = 1$ indicate that X and Y are component-wisely dependent. Figures 1–2 summarize the rejection rates with the increasing number of random projections k from 500 to 10000 by the increment of 500. We set $(n, p) = (100, 300)$. It can be seen that the proposed test has good performance in size control, and it is not sensitive to the choice of projection numbers k. As for the power performance in Example 5.1 (i), the proposed tests performs better when the number of projections increases, especially in cases of strong component-wise dependence. When k is greater than 8000, the power performance of the tests gradually becomes stable in the sense that the increase of power will not exceed 0.1. On the other hand, similar performance of the proposed tests appears in Example 5.1 (ii), except that the increase of power is not significant when the strength of dependence is relatively weak (e.g., $\epsilon = 0.3$). The computational complexity of the proposed test will increase as the projection number increases. Nevertheless, the computing time of the proposed test can be reduced by parallelly executing the repetitive computations of projected statistic $\mathrm{HSIC}_n(\alpha_i^\top X, \beta_i Y)$ on machines with multiple cores. In view of the computation efficiency, we suggest to choose a relatively larger value $k = 8000$ throughout.

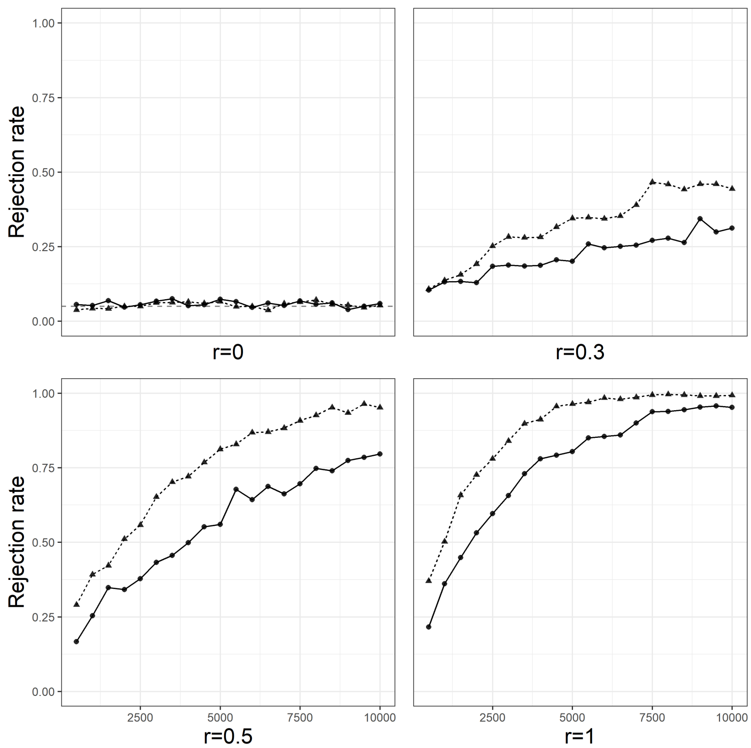

Figure 1: The rejection rates of $\mathrm{KP_C}$ (solid) and $\mathrm{KP_L}$ (dashed) in Example 1 (i) for $(n, p) = (100, 300)$ and different values of k.

S10.2 Normal approximation accuracy

In this subsection, we present the kernel density estimates of the standardized test statistics under the cases of Example 5.2 in the main text and

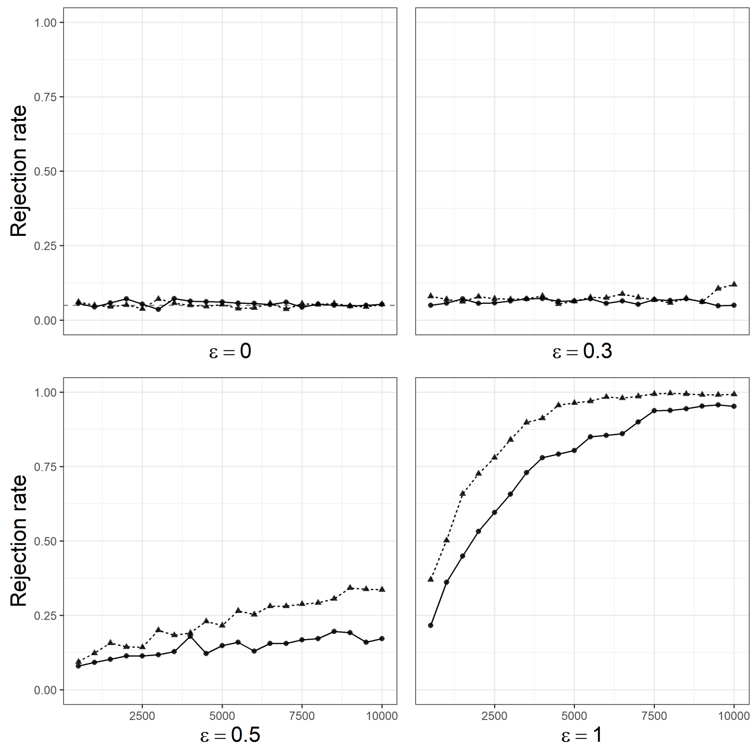

Figure 2: The rejection rates of $\mathrm{KP_C}$ (solid) and $\mathrm{KP_L}$ (dashed) in Example 1 (ii) for $(n, p) = (100, 300)$ and different values of k.

compared them with the standard normal distribution. Recall from Section 5.1 that

Example 5.2. Let $\Sigma = (1-c)I_p + c1_p1_p^\top \in \mathbb{R}^{p\times p}$ with $c = 0.3$ be a equicorrelation matrix, which has diagonal elements 1 and off-diagonal elements c. Generate i.i.d. samples from the following models for $i = 1, \ldots, n$.

(i) $X_i = (X_{i,1}, \ldots, X_{i,p})^\top \sim N(0_p, I_p), Y_i = (Y_{i,1}, \ldots, Y_{i,p})^\top \sim N(0_p, I_p)$;

(ii) $X_i = (X_{i,1}, \ldots, X_{i,p})^\top \sim N(0_p, \Sigma), Y_i = (Y_{i,1}, \ldots, Y_{i,p})^\top \sim N(0_p, \Sigma)$;

(iii) $X_i = (X_{i,1}, \dots, X_{i,p})^\top$, $\{X_{i,j}\}_{j=1}^p$ are i.i.d. standardized χ^2 random variables with degree of freedom 1, $Y_i = (Y_{i,1}, \dots, Y_{i,p})^\top$, $\{Y_{i,j}\}_{j=1}^p$ are i.i.d. standardized χ^2 random variables with degree of freedom 1;

(iv) $X_i = (X_{i,1}, \dots, X_{i,p})^\top$, $\{X_{i,j}\}_{j=1}^p$ are i.i.d. standard Cauchy random variables, $Y_i = (Y_{i,1}, \dots, Y_{i,p})^\top$, $\{Y_{i,j}\}_{j=1}^p$ are i.i.d. standard Cauchy random variables.

Figure 3–6 show that the null distribution is quite close to standard normal distribution when the dimensions and sample size increase for Example 5.2 (i), (iii) and (iv), which confirms the asymptotic normality of the standardized statistic under H_0 given in Corollary 1. However, there is some right skewness for Example 5.2 (ii). This is because Conditions (A2)–(A3) that ensure the asymptotic normality of $U_{n,k}$ may exclude some situations such as the spiked model. See more detailed analysis in Example 5.2 of the main text.

S10.3 Power performance

In this subsection, we conduct additional simulations to assess the power performance of different tests.

Example 2. Let $\Sigma_p = (0.5^{|i-j|}) \in \mathbb{R}^{p \times p}$ and $r, \epsilon \in (0, 1]$. Generate i.i.d. samples from the following models for $i = 1, \dots, n$.

(i) $X_i = (X_{i,1}, \ldots, X_{i,p})^\top \sim N(0_p, \Sigma_p)$, $Y_i = (Y_{i,1}, \ldots, Y_{i,p})^\top$, where $Y_{i,j} = X_{i,j}^2$ for $1 \le j \le rp$, and $\{Y_{i,j}\}_{j=rp+1}^p$ are i.i.d. standard Cauchy random variables;

(ii) $X_i = (X_{i,1}, \ldots, X_{i,p})^\top \sim N(0_p, \Sigma_p)$, $Z_i = (Z_{i,1}, \ldots, Z_{i,p})$, $\{Z_{i,j}\}_{j=1}^p$ are i.i.d. standard Cauchy random variables, $Y_i = (Y_{i,1}, \ldots, Y_{i,p})^\top$, where $Y_{i,j} = \epsilon X_{i,j}^2 + (1-\epsilon) Z_{i,j}^2$ for $j = 1, \ldots, p$.

Table 1: Power comparison from Example 2 (i)

r	n	p	dC	hC$_\mathrm{G}$	CZ$_\mathrm{G}$	KP$_\mathrm{C}$	hC$_\mathrm{L}$	CZ$_\mathrm{L}$	KP$_\mathrm{L}$	BSG	HHG
0.3	50	50	0.058	0.056	0.113	0.456	0.063	0.114	0.847	0.066	0.045
	50	100	0.063	0.048	0.128	0.299	0.061	0.084	0.568	0.069	0.051
	100	100	0.056	0.063	0.250	0.626	0.058	0.139	0.909	0.053	0.045
	100	300	0.068	0.049	0.239	0.212	0.059	0.109	0.359	0.051	0.052
0.5	50	50	0.052	0.068	0.293	0.859	0.076	0.218	0.996	0.099	0.047
	50	100	0.064	0.045	0.281	0.558	0.057	0.183	0.891	0.058	0.047
	100	100	0.068	0.058	0.583	0.957	0.060	0.367	0.999	0.070	0.049
	100	300	0.059	0.054	0.580	0.407	0.061	0.207	0.666	0.063	0.048
0.7	50	50	0.058	0.070	0.628	0.983	0.080	0.573	1.000	0.222	0.088
	50	100	0.055	0.059	0.577	0.822	0.060	0.424	0.993	0.119	0.061
	100	100	0.086	0.059	0.898	1.000	0.070	0.830	1.000	0.173	0.070
	100	300	0.058	0.051	0.874	0.665	0.055	0.555	0.895	0.071	0.060
1	50	50	0.335	0.166	1.000	1.000	0.712	1.000	1.000	1.000	1.000
	50	100	0.186	0.127	1.000	0.991	0.403	1.000	1.000	1.000	1.000
	100	100	0.388	0.169	1.000	1.000	0.839	1.000	1.000	1.000	1.000
	100	300	0.135	0.080	1.000	0.929	0.294	1.000	0.998	1.000	1.000

Tables 1–2 chart the empirical power of Example 2. As mentioned in Example 1, the parameters r and ϵ can represent the strength of the non-linear dependence between X and Y. We set $r = 0.3$, 0.5, 0.7, 1 and $\epsilon = 0.3$, 0.5, 0.7, 1, where $r = 1$ and $\epsilon = 1$ indicate that X and Y are component-wisely dependent. In particular, when $r < 1$, the dependence between X

Table 2: Power comparison from Example 2 (ii)

ϵ	n	p	dC	$\mathrm{hC_G}$	$\mathrm{CZ_G}$	$\mathrm{KP_C}$	$\mathrm{hC_L}$	$\mathrm{CZ_L}$	$\mathrm{KP_L}$	BSG	HHG
0.3	50	50	0.050	0.063	0.052	0.164	0.067	0.042	0.330	0.049	0.054
	50	100	0.067	0.057	0.071	0.116	0.054	0.052	0.194	0.060	0.061
	100	100	0.052	0.050	0.057	0.074	0.061	0.059	0.144	0.048	0.042
	100	300	0.049	0.058	0.037	0.060	0.059	0.054	0.073	0.047	0.047
0.5	50	50	0.070	0.052	0.048	0.478	0.048	0.067	0.773	0.047	0.048
	50	100	0.065	0.066	0.051	0.241	0.078	0.061	0.494	0.047	0.050
	100	100	0.054	0.062	0.063	0.171	0.056	0.052	0.404	0.056	0.052
	100	300	0.067	0.044	0.055	0.088	0.049	0.063	0.146	0.049	0.054
0.7	50	50	0.070	0.053	0.048	0.478	0.048	0.067	0.773	0.047	0.048
	50	100	0.065	0.066	0.051	0.241	0.078	0.061	0.494	0.047	0.050
	100	100	0.062	0.063	0.066	0.550	0.060	0.062	0.858	0.037	0.034
	100	300	0.076	0.056	0.057	0.163	0.055	0.074	0.310	0.064	0.056
1	50	50	0.335	0.166	1.000	1.000	0.712	1.000	1.000	1.000	1.000
	50	100	0.186	0.127	1.000	0.991	0.403	1.000	1.000	1.000	1.000
	100	100	0.388	0.169	1.000	1.000	0.839	1.000	1.000	1.000	1.000
	100	300	0.135	0.080	1.000	0.929	0.294	1.000	0.998	1.000	1.000

and Y in Example 2 (i) is beyond the coordinate-wise dependence assumed in Theorem 4. The proposed tests have higher power than other tests in most cases, especially in scenarios with weak dependence. It means that the proposed test with special selected parameters may have a wider range of applications, going beyond the scenario considered in Theorem 4. Similar to Example 5.3, the distance correlation test and the HSIC tests have poor performance in all cases since the dependence between X and Y is pure non-linear. The graph-based tests exhibit very low powers in detecting the weak dependence. While the group-wise HSIC tests have relatively good performance in Example 2 (i), these tests have substantial power loss when the dependence is weak in Example 2 (ii). It means that the tests of Chakraborty and Zhang (2019) may not be well-performing when the magnitude of coordinate-wise signals is weak.

Example 3. Let $\Sigma = (1-c)I_p + c1_p1_p^\top \in \mathbb{R}^{p\times p}$ with $c \in (0,1)$ be a equicorrelation matrix, which has diagonal elements 1 and off-diagonal elements c. Let $\epsilon \in (0,1)$. Generate i.i.d. samples from the following models for $i = 1, \ldots, n$.

(i) $X_i = (X_{i,1}, \ldots, X_{i,p})^\top \sim N(0_p, \Sigma)$, $Z_i = (Z_{i,1}, \ldots, Z_{i,p})$, $\{Z_{i,j}\}_{j=1}^p$ are i.i.d. standard Cauchy random variables, $Y_i = (Y_{i,1}, \ldots, Y_{i,p})^\top$, where $Y_{i,j} = \epsilon X_{i,j}^2 + (1-\epsilon)Z_{i,j}^2$ for $j = 1, \ldots, p$.

In the above scenario, the parameter c represents the strength of dependence between the components of X, while ϵ represents the strength of the dependence between X and Y. Tables 3–5 summarizes the empirical powers of Example 3 (i) with $c = 0.3$, 0.5, 0.7 and $\epsilon = 0.3$, 0.7, 1. From Table 5 which the magnitude of coordinate-wise dependence between X and Y is strong, we can see that all tests except the dC test have good power performance in all cases. The power of the dC test can still approach one as the sample size n, the dimensionality p and the dependence strength c increase. This interesting phenomenon indicates that the dependence between components can strengthen the coordinate-wise dependence between X and Y, which is also observed by some recent researcher (e.g., Zhu et al.

(2020), Gao et al. (2021)). On the other hand, it can be seen from Tables 3–4 that the proposed tests have superior performance in detecting the weak dependence between X and Y, while the other tests exhibit very low powers in these cases.

Table 3: Power comparison from Example 3 (i) with $\epsilon = 0.3$

c	n	p	dC	hC_G	CZ_G	KP_C	hC_L	CZ_L	KP_L	BSG	HHG
0.3	50	50	0.061	0.063	0.045	0.079	0.064	0.057	0.133	0.057	0.047
	50	100	0.052	0.065	0.042	0.100	0.066	0.052	0.140	0.051	0.052
	100	100	0.052	0.057	0.052	0.102	0.053	0.049	0.181	0.067	0.039
	100	300	0.056	0.060	0.055	0.086	0.065	0.052	0.146	0.054	0.048
0.5	50	50	0.070	0.073	0.063	0.137	0.071	0.077	0.235	0.054	0.052
	50	100	0.048	0.057	0.047	0.157	0.064	0.052	0.291	0.052	0.053
	100	100	0.063	0.080	0.067	0.285	0.076	0.075	0.582	0.048	0.056
	100	300	0.076	0.063	0.050	0.328	0.064	0.069	0.685	0.044	0.047
0.7	50	50	0.075	0.057	0.064	0.279	0.063	0.065	0.550	0.054	0.046
	50	100	0.072	0.079	0.055	0.383	0.073	0.073	0.704	0.056	0.051
	100	100	0.061	0.065	0.058	0.678	0.063	0.077	0.957	0.058	0.050
	100	300	0.066	0.068	0.053	0.832	0.066	0.064	0.997	0.056	0.049

Table 4: Power comparison from Example 3 (i) with $\epsilon = 0.7$

c	n	p	dC	hC_G	CZ_G	KP_C	hC_L	CZ_L	KP_L	BSG	HHG
0.3	50	50	0.067	0.078	0.056	0.457	0.066	0.064	0.708	0.049	0.051
	50	100	0.072	0.064	0.057	0.362	0.069	0.061	0.565	0.056	0.053
	100	100	0.081	0.058	0.076	0.730	0.059	0.077	0.927	0.043	0.054
	100	300	0.051	0.061	0.053	0.575	0.057	0.049	0.773	0.049	0.045
0.5	50	50	0.063	0.060	0.054	0.766	0.071	0.061	0.924	0.052	0.061
	50	100	0.066	0.072	0.062	0.767	0.071	0.069	0.905	0.054	0.051
	100	100	0.055	0.051	0.044	0.989	0.062	0.058	1.000	0.055	0.055
	100	300	0.061	0.069	0.061	0.998	0.064	0.070	1.000	0.071	0.047
0.7	50	50	0.072	0.079	0.072	0.970	0.074	0.082	0.998	0.043	0.067
	50	100	0.061	0.062	0.062	0.984	0.060	0.065	0.999	0.047	0.061
	100	100	0.061	0.081	0.052	1.000	0.071	0.073	1.000	0.043	0.056
	100	300	0.063	0.073	0.060	1.000	0.067	0.063	1.000	0.053	0.049

Table 5: Power comparison from Example 3 (i) with $\epsilon = 1$

c	n	p	dC	$\mathrm{hC_G}$	$\mathrm{CZ_G}$	$\mathrm{KP_C}$	$\mathrm{hC_L}$	$\mathrm{CZ_L}$	$\mathrm{KP_L}$	BSG	HHG
0.3	50	50	0.730	0.924	0.997	0.999	0.989	1.000	1.000	0.998	1.000
	50	100	0.735	0.947	0.991	0.971	0.996	1.000	1.000	0.996	1.000
	100	100	0.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	300	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	50	50	0.986	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	100	0.991	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
0.7	50	50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	50	100	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	100	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	100	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

References

Chakraborty, S. and X. Zhang (2019). Distance metrics for measuring joint dependence with application to causal inference. *Journal of the American Statistical Association 114*(528), 1638–1650.

Chakraborty, S. and X. Zhang (2021). A new framework for distance and kernel-based metrics in high dimensions. *Electronic Journal of Statistics 15*(2), 5455–5522.

Chen, S. X. and Y.-L. Qin (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics 38*(2), 808–835.

Chen, S. X., L.-X. Zhang, and P.-S. Zhong (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association 105*(490), 810–819.

Döbler, C. and G. Peccati (2019). Quantitative clts for symmetric u-statistics using contractions. *Electronic Journal of Probability 24*, 1–43.

Gao, L., Y. Fan, J. Lv, and Q.-M. Shao (2021). Asymptotic distributions of high-dimensional distance correlation inference. *The Annals of Statistics 49*(4), 1999–2020.

Han, Q. and Y. Shen (2021). Generalized kernel distance covariance in high dimensions: non-null clts and power universality. *arXiv preprint arXiv:2106.07725*.

Hu, S., X. Wang, W. Yang, and X. Wang (2014). A note on the inverse moment for the non negative random variables. *Communications in Statistics-Theory and Methods 43*(8), 1750–1757.

Lee, A. J. (1990). *U-statistics: Theory and Practice*. CRC Press.

Reed, M. (1972). Methods of modern mathematical physics i. *Functional analysis*.

Schoenberg, I. J. (1938). Metric spaces and completely monotone functions. *Annals of Mathematics*, 811–841.

Sriperumbudur, B. K., K. Fukumizu, and G. R. Lanckriet (2011). Universality, characteristic kernels and rkhs embedding of measures. *Journal of Machine Learning Research 12*(7), 2389–2410.

Stewart, J. (1976). Positive definite functions and generalizations, an historical survey. *The Rocky Mountain Journal of Mathematics 6*(3), 409–434.

Zhong, P.-S. and S. X. Chen (2011). Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association 106*(493), 260–274.

Zhu, C., X. Zhang, S. Yao, and X. Shao (2020). Distance-based and rkhs-based dependence

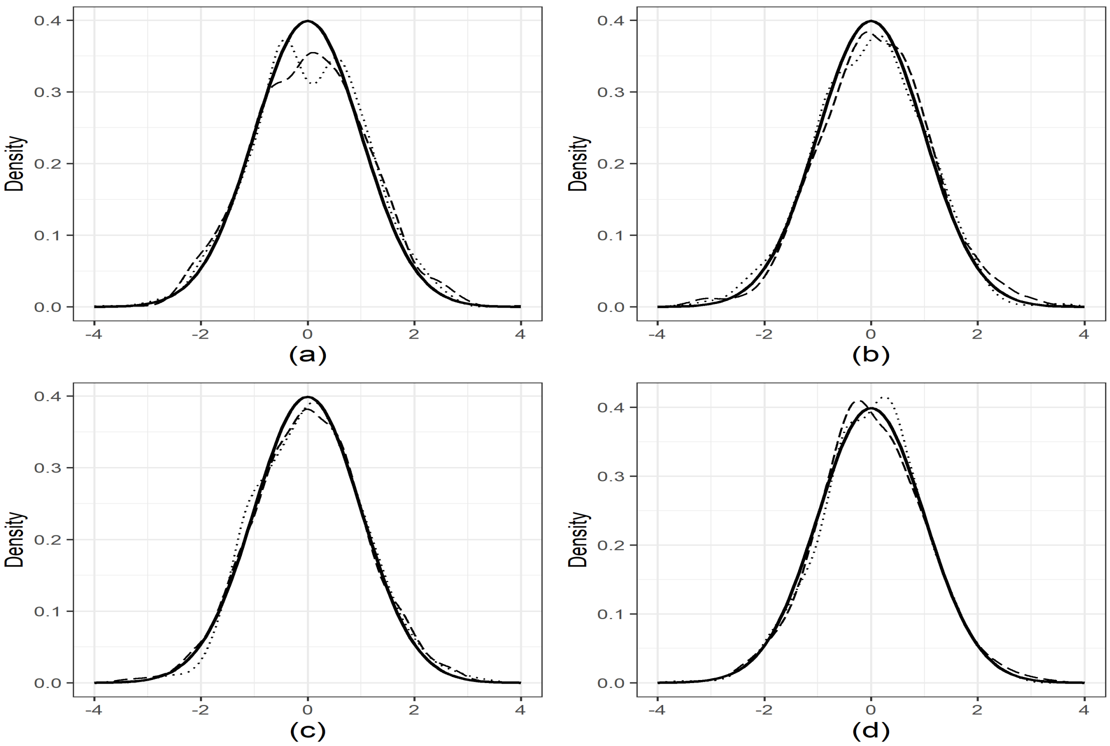

Figure 3: The kernel density plot of KP_C (dashed) and KP_L (dotted) in Example 5.2 (i): (a) $(n, p) = (50, 50)$; (b) $(n, p) = (50, 100)$; (c) $(n, p) = (100, 100)$; (d) $(n, p) = (100, 300)$. Solid lines correspond to the density of standard normal distribution.

metrics in high dimension. *The Annals of Statistics 48*(6), 3366–3394.

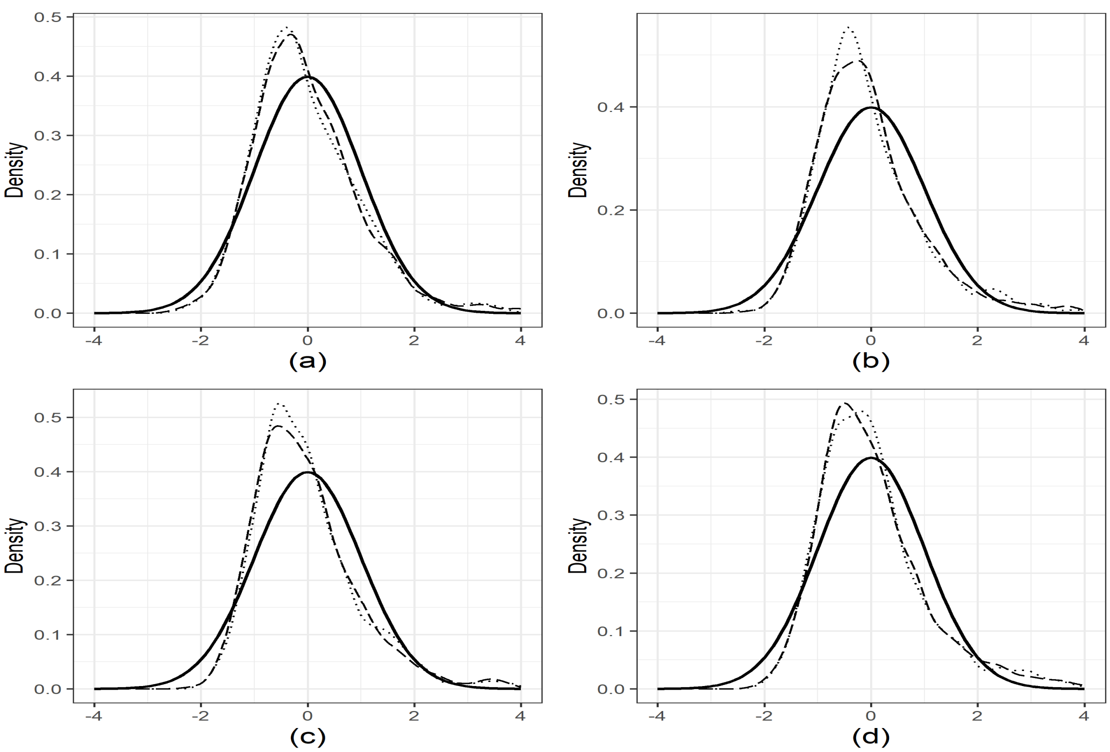

Figure 4: The kernel density plot of $\mathrm{KP_C}$ (dashed) and $\mathrm{KP_L}$ (dotted) in Example 5.2 (ii): (a) $(n, p) = (50, 50)$; (b) $(n, p) = (50, 100)$; (c) $(n, p) = (100, 100)$; (d) $(n, p) = (100, 300)$. Solid lines correspond to the density of standard normal distribution.

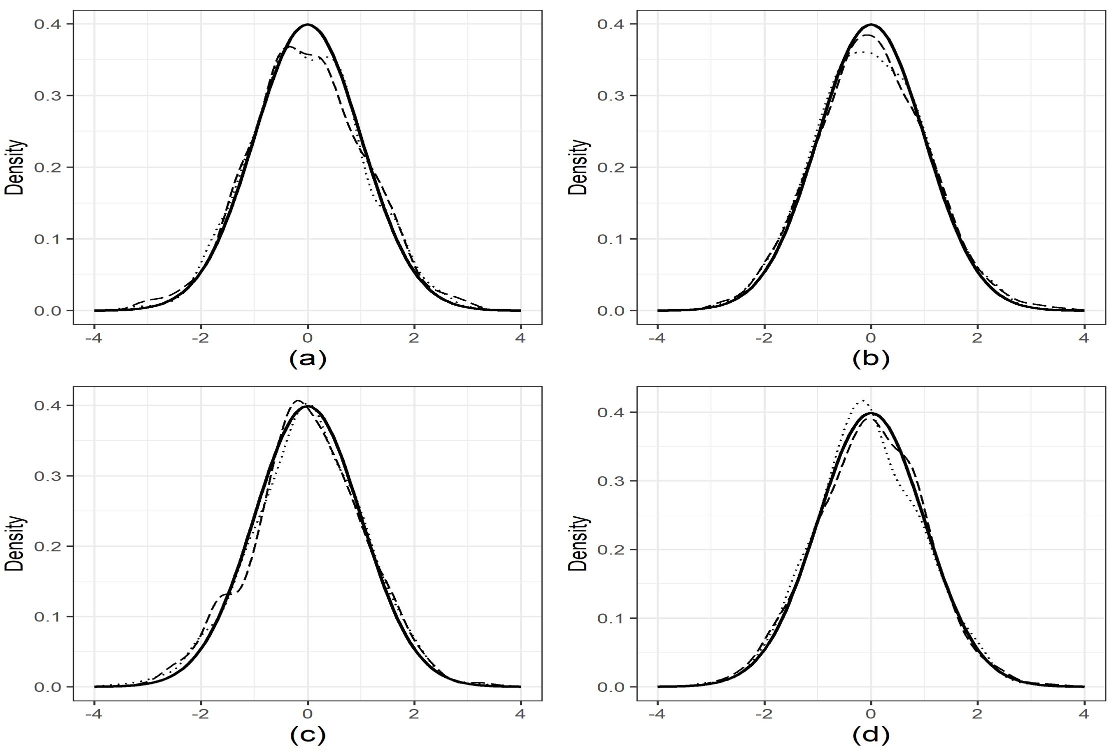

Figure 5: The kernel density plot of KP_C (dashed) and KP_L (dotted) in Example 5.2 (iii): (a) $(n, p) = (50, 50)$; (b) $(n, p) = (50, 100)$; (c) $(n, p) = (100, 100)$; (d) $(n, p) = (100, 300)$. Solid lines correspond to the density of standard normal distribution.

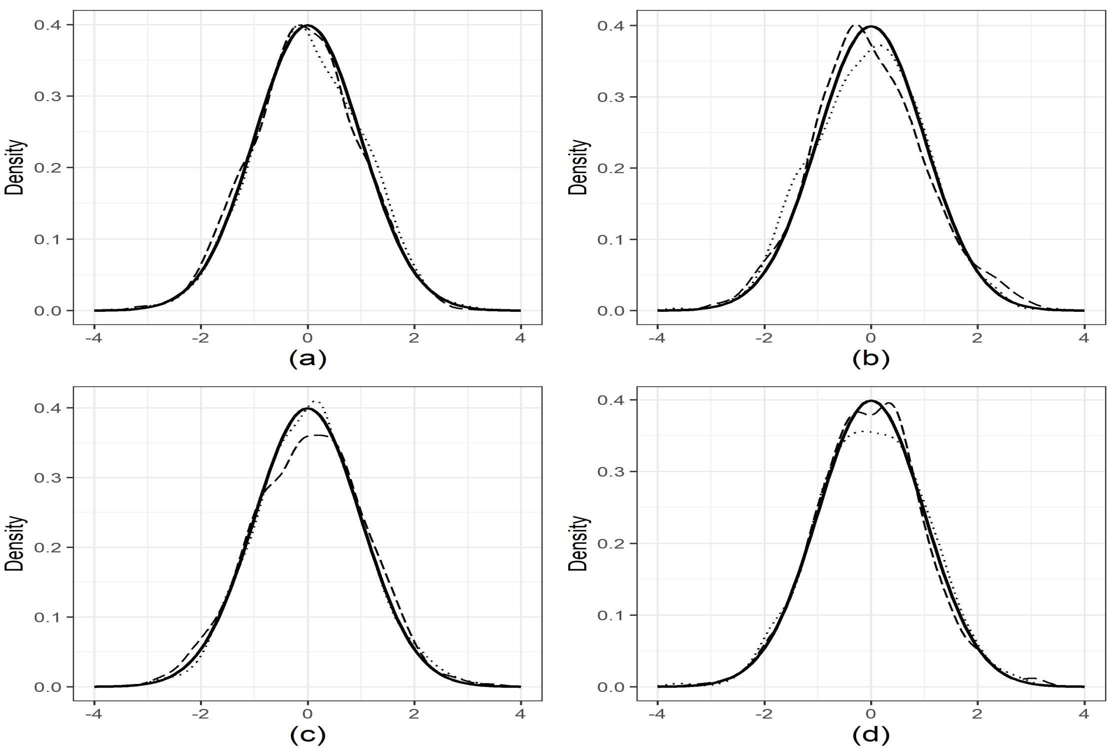

Figure 6: The kernel density plot of $\mathrm{KP_C}$ (dashed) and $\mathrm{KP_L}$ (dotted) in Example 5.2 (iv): (a) $(n, p) = (50, 50)$; (b) $(n, p) = (50, 100)$; (c) $(n, p) = (100, 100)$; (d) $(n, p) = (100, 300)$. Solid lines correspond to the density of standard normal distribution.