Statistica Sinica Preprint No: SS-2022-0366

Title	A Kernel Independence Test Using Projection-Based Measure in High-Dimension
Manuscript ID	SS-2022-0366
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0366
Complete List of Authors	Yuexin Chen and Wangli Xu
Corresponding Authors	Wangli Xu
E-mails	wlxu@ruc.edu.cn

Notice: Accepted version subject to English editing.

Statistica Sinica

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

Yuexin Chen and Wangli Xu

Renmin University of China

Abstract: Testing the independence between two high-dimensional random vectors is a fundamental and challenging problem in statistics. Most existing tests based on distance and kernel may fail to detect the non-linear dependence in the high-dimensional regime. To tackle this obstacle, this paper proposes a kernel independence test for assessing the independence between two random vectors based on a class of Gaussian projections relying on tuning parameters. The proposed test can be generally implemented for a wide class of distance-based kernels and completely characterizes dependence in the low-dimensional regime. Besides, the test captures pure non-linear dependence in the high-dimensional regime. Theoretically, we develop central limit theorem and associated rate of convergence for the proposed statistic under some mild regularity conditions and the null hypothesis. Moreover, we derive the asymptotic power of the proposed test enabling us to select suitable parameters for a special alternative, to achieve superior power in the high-dimensional regime. The choices of tuning parameters ensure that the proposed test has comparable power with the original kernel-based test in the moderately high-dimensional regime. Numerical experiments also demonstrate the satisfactory empirical performance of the proposed test in various scenarios.

Key words and phrases: High-dimension, independence test, kernel independence measure, random projections, U-statistics.

1. Introduction

Testing the independence of a pair of potentially high-dimensional random vectors has gained importance due to the increasing attention from big data applications (see e.g., Kong et al. (2015), Chakraborty and Zhang (2019)). Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be two random vectors with probability measures P_X and P_Y, respectively. Given independent samples $\{(X_i, Y_i), i = 1, \ldots, n\}$ from P_{XY}, the hypothesis of interest is

$$\mathrm{H}_0 : P_{XY} = P_X P_Y \quad \mathrm{v.s.} \quad \mathrm{H}_1 : P_{XY} \neq P_X P_Y. \tag{1.1}$$

There exists a wide spectrum of dependency measures and tests. Notable examples include Pearson correlation (Pearson (1895)), rank correlation coefficients (Kendall (1938), Spearman (1961)), coefficients based on cumulative distribution function (Hoeffding (1994), Blum et al. (1961), Zhu et al. (2017), Pan et al. (2020)), measures based on signs and empirical characteristic functions (Sinha and Wieand (1977), Gieser and Randles (1997)), sign covariances (Bergsma and Dassios (2014), Weihs et al. (2018)), graph-based tests (Heller et al. (2012), Heller et al. (2013), Biswas et al. (2016)), measures based on distance and kernel (Bach and Jordan (2002), Székely et al. (2007), Gretton et al. (2007)). See also Pfister et al. (2018), Chakraborty and Zhang (2019) and Roy et al. (2020) for some recent developments on testing the joint dependence among more than two random vectors.

Among the plentiful independence tests, the distance and kernel-based tests have gained growing popularity in recent years. These tests utilize some dependency correlations which possess the characteristic property that a zero correlation is equivalent to independence. Székely et al. (2007) introduced a dependence metric named distance covaraiance, which has gained lots of attention due to its ability to quantify non-linear dependence and the flexibility to be applicable to two random vectors in arbitrary dimensions. Gretton et al. (2007) proposed a kernel independence measure named Hilbert-Schmidt Independence Criterion using the Hilbert-Schmidt norm of the cross-covariance opertor. This criterion, denoted by HSIC, requires the use of characteristic kernels to guarantee that the zero value implies the independence between random vectors. Sejdinovic et al. (2013) provided a unifying framework establishing the equivalence between the distance correlation and HSIC. For every negative type metric, there exists a positive-definite kernel such that the quantities are equal.

In the high-dimensional setting, the asymptotic behavior of the sample distance correlation and sample HSIC was recently studied in the literature (e.g., Zhu et al. (2020), Gao et al. (2021) and Han and Shen (2021)), where they established the asymptotic normality for these test statistics under the null and alternative hypotheses. However, there have been some recent works to gain insight on the limitation of the kernel and distance-based tests for high-dimensional setting (see Székely and Rizzo (2013); Ramdas et al. (2015); Chakraborty and Zhang (2021)). For example, Chakraborty and Zhang (2021) showed that the distance and kernel-based tests can only detect component-wise linear dependency and fail to detect non-linear dependency in the regime of fast growing dimensionality, e.g., $\min\{p, q\}/n^2 \to \infty$.

Various attempts have been made to improve the behaviors of the distance and kernel-based tests. Székely and Rizzo (2013) extended the distance correlation and further proposed a t-test for (1.1) under the setting that the dimensions p and q grow while sample size n is fixed. Leung and Drton (2018) proposed using sum of pairwise rank correlations to test for mutual independence of high-dimensional vectors. Recently, some researchers considered the tests based on the low-dimensional structures of X and Y. Zhu et al. (2020) suggested test of independence by aggregating the pairwise squared sample HSIC and studied its asymptotic behavior in the high-dimensional setup. However, the marginally aggregated statistic can detect only the component-wise dependency and thus may be less powerful when X and Y have more complex dependency. Subsequently, a generalization of the marginally aggregated method of Zhu et al. (2020) was presented in Chakraborty and Zhang (2021). They considered a new dependence metric based on grouping the components of two high-dimensional random vectors separately, and further showed that their statistic is able to detect the non-linear dependencies between the different groups of X and Y. However, the theoretical framework in Chakraborty and Zhang (2021) does not encompass the applicable methods to partition the random vectors in the group-wise statistic. Additionally, their test mainly accounts for group-wise dependency at the risk that it may suffer from power loss if the true dependence in data is more than group-wise dependence.

Identifying the optimal grouping structures including the grouping dimensions and components is pivotal to the group-wise methods. However, it is a non-trivial task when the prior knowledge of the true data structure is unavailable (see Chakraborty and Zhang (2021)). Therefore, employing randomness in the selection of low-dimensional structures is highly suggested. In this paper, we propose a new independence criterion incorporating

HSIC with random projections to characterize the dependence between two random vectors. We introduce a class of Gaussian projections relying on tuning parameters to randomly sparsify X and Y and project them into one-dimensional spaces. The main contributions can be summarized as follows.

- Under the proposed class of projections, the proposed independence criterion inherits the desirable characteristic property that completely characterizes dependence for low-dimensional setting. In particular, the proposed criterion boils down to HSIC when the tuning parameters are set to 1. Furthermore, the proposed criterion is generally applicable to the kernels with positive definite functions without requiring the use of characteristic kernels.

- We propose an unbiased U-statistic type estimator of the proposed criterion. Moreover, we establish the explicit rate of convergence to normal distribution and further obtain the central limit theorem of the proposed statistic under the null hypothesis. Thus the test based on the proposed statistic can be conveniently implemented by using standard normal critical values.

- We also derive the asymptotic power of the proposed test. The signal-to-noise ratio related to the power function enables the principled selection of tuning parameters to obtain the power consistency of the proposed test under a special alternative. This demonstrates the capability of the proposed test in detecting the pure non-linear dependency in the high-dimensional regime, as opposed to merely measuring component-wise linear dependence by the HSIC.

The rest of this paper is organized as follow. In Section 2, we review the kernel test based on HSIC. Section 3 introduces the new independence criterion and further constructs a natural unbiased U-statistic type estimator of the proposed criterion. In Section 4, we investigate the asymptotic null distribution and power study of the proposed statistic. The numerical simulation results and real data analysis will be presented in Section 5.

Notation. For a, $b \in \mathbb{R}$, $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. Let 0_p be the origin of $\mathbb{R}^p$. For $b \in \mathbb{R}^p$ and $U \in \mathbb{R}^{p\times p}$, denote by $\|b\|$ and $\|U\|_F$ the Euclidean norm of b and the matrix Frobenious norm of U. Let I_p denote the $p \times p$ identity matrix, respectively. Given two real-valued random variables U, V, we write

$$d_{\mathcal{W}}(U, V) = \sup_{h\in \mathrm{Lip}(1)} |E(h(U)) - E(h(V))|,$$

where Lip(1) is the class of all 1-Lipschitz mappings $h : \mathbb{R} \to \mathbb{R}$, to indicate the Wasserstein distance between the distributions of U and V.

2. Overview of HSIC

Gretton et al. (2007) proposed the HSIC to measure the dependence between X and Y. It embeds the joint distribution and the product of the marginal distributions into a reproducing kernel Hilbert space (RKHS) and measures their squared distance. Following the notation in Gretton et al. (2007), let $\mathcal{F}$ be a RKHS on $\mathbb{R}^p$ with the positive definite kernel $K(\cdot,\cdot)$ that satisfies $K(x,x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{F}}$. Likewise, let $\mathcal{G}$ be a RKHS on $\mathbb{R}^q$ with the positive definite kernel $L(\cdot,\cdot)$ that satisfies $L(y,y') = \langle \psi(y), \psi(y') \rangle_{\mathcal{G}}$. The HSIC is defined as

$$\begin{aligned}\text{HSIC}(X,Y) = {} & E\{K(X_1,X_2)L(Y_1,Y_2)\} + E\{K(X_1,X_2)\}E\{L(Y_1,Y_2)\} \\ & - 2E\{K(X_1,X_2)L(Y_1,Y_3)\},\end{aligned} \tag{2.2}$$

where $(X_1,Y_1),(X_2,Y_2),(X_3,Y_3)$ are independent identically distributed samples. Moreover, one can show that $\text{HSIC}(X,Y) = E\{d_K(X_1,X_2)d_L(Y_1,Y_2)\}$ with the double-centered distance

$$\begin{aligned}d_K(X_1,X_2) = {} & K(X_1,X_2) - E\{K(X_1,X_2) \mid X_1\} - E\{K(X_1,X_2) \mid X_2\} \\ & + E\{K(X_1,X_2)\},\end{aligned} \tag{2.3}$$

and $d_L(Y_1,Y_2)$ can be defined similarly.

Key to the kernel measure of independence is the notion of characteristic kernels, which provide sufficiently rich RKHSs to characterize dependence between two random vectors (see, e.g., Fukumizu et al. (2008), Hofmann et al. (2008)). Sriperumbudur et al. (2011) proved that some widely used kernels such as Laplacian kernel, Gaussian kernel and rational quadratic kernel are characteristic. Generally, we assume that the kernels can be represented compactly as

$$K(X_1, X_2) = f\left(\frac{\|X_1 - X_2\|}{\gamma_X}\right), L(Y_1, Y_2) = g\left(\frac{\|Y_1 - Y_2\|}{\gamma_Y}\right), \tag{2.4}$$

where $f(x)$, $g(x)$ are continuously differentiable, real-valued functions, γ_X and γ_Y are the bandwidth parameters. For technical convenience, this paper focuses on deterministic choices of γ_X and γ_Y satisfying that $\gamma_X^2/E\|X_1 - X_2\|^2$ and $\gamma_Y/E\|Y_1 - Y_2\|^2$ are bounded away from 0 and ∞.

We also assume $f(x)$ and $g(x)$ have bounded derivatives, which are commonly assumed in the analysis of kernel measures (see e.g., Han and Shen (2021) and Yan and Zhang (2021)). The function $f : \mathbb{R} \to \mathbb{R}$ is called positive definite in $\mathbb{R}$ if

$$\sum_{i,j=1}^{n} f(\zeta_i - \zeta_j)\xi_i\xi_j \geq 0$$

holds for $n \in \mathbb{N}$, every choice of $\zeta_1, \ldots, \zeta_n \in \mathbb{R}$ and $\xi_1, \ldots, \xi_n \in \mathbb{R}$. Furthermore, $f(x)$ is said to be of positive type if for all $m, n \in \mathbb{N}$, every choice of $x_1, \ldots, x_n \in \mathbb{R}^m$ and $\xi_1, \ldots, \xi_n \in \mathbb{R}$, it holds that

$$\sum_{i,j=1}^{n} f(\|x_i - x_j\|)\xi_i\xi_j \geq 0.$$

Similar definitions apply to $g(x)$. In general, a distance-based kernel $K(X_1, X_2) = f(\|X_1 - X_2\|/\gamma_X)$ is positive definite if and only $f(x)$ is of positive type. We emphasis that positive definite function in $\mathbb{R}$ is not necessarily to be of positive type. For example, $f(x) = \cos(x)$ is positive definite in $\mathbb{R}$ (Stewart (1976)), but not of positive type since $\cos(\sqrt{x})$ does not satisfy the completely monotone condition (see e.g., Theorem 7.14 of Wendland (2004)). Therefore, the class of distance-based kernels with positive definite functions in $\mathbb{R}$ is not necessarily positive definite.

The class of distance-based kernels of the form in (2.4) contains the aforementioned kernels. For example, characteristic Laplacian kernel can be defined by choosing $f(y) = \exp(-y)$, i.e., $K(X_1, X_2) = \exp(-\|X_1 - X_2\|/\gamma_X)$. Non-characteristic cosine kernel can be defined by choosing $f(y) = \cos(x)$, i.e., $K(X_1, X_2) = \cos(\|X_1 - X_2\|/\gamma_X)$. In practice, the bandwidth parameters γ_X and γ_Y are heuristically chosen as the median distance between the sample observations.

While the HSIC can be used as a measure of dependence when using characteristic kernels, Chakraborty and Zhang (2021) showed that in the high-dimensional regime, HSIC can be asymptotically represented as

$$\text{HSIC}(X, Y) = \frac{1}{4\rho_X \rho_Y} \sum_{i=1}^{p} \sum_{j=1}^{q} \text{cov}^2(X_{1,i}, Y_{1,j}) + \mathcal{R},$$

where $\rho_X^2 = E\{K(X_1, X_2)^2\}$, $\rho_Y^2 = E\{L(Y_1, Y_2)^2\}$, and $\mathcal{R} = o(1)$. Thus

HSIC can only measure the linear dependency when dimensions grow high. Recently, Chakraborty and Zhang (2021) proposed independence test using the sum of group-wise squared sample HSIC, and further showed that the sample estimator asymptotically quantifies group-wise non-linear dependence between two high-dimensional vectors. In Example 2 of Supplementary Material, we give an example illustrating that the group-wise type test can indeed be less powerful when the dependence between X and Y is weak.

3. The Proposed Test Statistic

3.1 The proposed independence criterion

Modern research seeks to construct test statistics based on projection onto lower dimensional subspace (see e.g., Zhu et al. (2017), Wang and Xu (2018), Kim et al. (2020)). Testing the independence between X and Y is equivalent to testing the independence between all one-dimensional projections of X and Y. To cope with the high-dimensionality issue, we propose to randomly sparsify X and Y and project them into one-dimensional spaces.

Heuristically, a pair of projection directions (θ, η) is of interest if it captures sufficient dependence between X and Y in the sense that the projected vectors $(\theta^\top X, \eta^\top Y)$ are strongly correlated relative to other projections.

However, the prior information of (X, Y) on these interesting directions is generally unknown or unavailable in the real world applications. Therefore, the projection distribution should cover as many directions as possible. As shown in Section 2 of Fang et al. (2018), spherically symmetric projection distributions possess good properties. In this view, spherically symmetric projections should be chosen. It is well-known that the standard normal distribution is spherically distributed (see Section 4 of Fang et al. (2018)).

In this paper, we consider p-dimensional projection α with i.i.d. component-wise mixture distribution having a point mass at 0 with probability $1-\gamma_1$ and a $N(0,1)$ distribution with probability γ_1, where the parameter $\gamma_1 \in (0,1]$. The projection β is independent of α and has a similar component-wise distribution with parameter $\gamma_2 \in (0,1]$. Let μ be the probability measure of α on $\mathbb{R}^p$ and υ be the probability measure of β on $\mathbb{R}^q$. To assess the independence of X and Y, we define Kernel Projection Independence Criterion, denoted by KPIC(X, Y) as

$$\mathrm{KPIC}\,(X, Y) = \int_{\mathbb{R}^p} \int_{\mathbb{R}^q} \mathrm{HSIC}\left(\alpha^\top X, \beta^\top Y\right) d\mu(\alpha) d\upsilon(\beta).$$

Define the projection kernels $\widetilde{K}\left(X_1, X_2\right) = E\left\{K\left(\alpha^\top X_1, \alpha^\top X_2\right) \mid X_1, X_2\right\}$ and $\widetilde{L}\left(Y_1, Y_2\right) = E\left\{L\left(\beta^\top Y_1, \beta^\top Y_2\right) \mid Y_1, Y_2\right\}$. Using the definition of HSIC in (2.2), the proposed criterion can be represented as

$$\begin{aligned} \mathrm{KPIC}\,(X,Y) = {}& E\left\{\widetilde{K}\,(X_1,X_2)\,\widetilde{L}\,(Y_1,Y_2)\right\} + E\left\{\widetilde{K}\,(X_1,X_2)\right\}E\left\{\widetilde{L}\,(Y_1,Y_2)\right\} \\ & - 2E\left\{\widetilde{K}\,(X_1,X_2)\,\widetilde{L}\,(Y_1,Y_3)\right\}. \end{aligned} \tag{3.5}$$

By the construction of projection distribution, the parameter γ_1 represents the ratio of non-zero entries of the projection. In the completely dense case ($\gamma_1 = 1$), the following proposition indicates that the projection kernels coincide with the characteristic kernels required in HSIC.

Proposition 1. *Suppose that* $\gamma_1 = 1$ *and* $K(X_1, X_2)$ *is a distance-based kernel defined in* (2.4) *with positive definite function* $f(x)$ *in* $\mathbb{R}$*, then the projection kernel* $\widetilde{K}(X_1, X_2)$ *is characteristic.*

As we can see, the proposed criterion with dense projections is actually HSIC using characteristic kernels. Furthermore, the following proposition establishes the characteristic property of KPIC for any $\gamma_1, \gamma_2 \in (0, 1]$, which means that (3.5) equal to zero if and only if X and Y are independent.

Proposition 2. *Suppose* $K(X_1, X_2)$ *and* $L(Y_1, Y_2)$ *are distance-based kernels defined in* (2.4) *with positive definite functions* $f(x)$ *and* $g(x)$ *in* $\mathbb{R}$*, then for* $\gamma_1, \gamma_2 \in (0, 1]$*,* $\mathrm{KPIC}\,(X, Y)$ *is nonnegative and has the characteristic property*

$$\mathrm{KPIC}\,(X,Y) = 0 \quad \textit{if and only if} \quad P_{XY} = P_X P_Y.$$

In the low-dimensional setting, Proposition 2 states that the proposed criterion completely characterizes independence between X and Y. Recently, Banerjee and Ghosh (2022) proposed a similar projection-based kernel independence criterion for quantifying the independence of random functions modelled as elements of a separable Hilbert space. The results presented in Banerjee and Ghosh (2022) can also be applied to the random vectors. From this view, they proposed using the joint probability measure P_{XY} as the projection distribution and obtained the explicit expression of their independence criterion under some condition on the support set of P_{XY}. While their criterion has a wider range of applications to random functions, the theoretical results for the random vectors in the high-dimensional regime ($p, q \to \infty$) remain largely unknown. In contrast to their work, our main goal is to tackle the high-dimensional issue of HSIC (see Section 2) and thus the randomly sparsified Gaussian projections are introduced. For simplicity and when there is no ambiguity, we may write $\mathrm{KPIC}\,(X, Y)$ and $\mathrm{HSIC}(X, Y)$ as KPIC and HSIC, respectively.

Remark 1. A distinctive feature of KPIC is that it only requires kernels with positive definite function in $\mathbb{R}$. As commented in Section 2, this class of kernel is not necessarily positive definite or characteristic. This implies that the proposed criterion has wider application prospect than the original kernel independence measures that requires a restricted class of characteristic kernels (see Sriperumbudur et al. (2010)). The key to this advantage is the adoption of Gaussian distribution in the proposed projection. If $f(x) = \cos(x)$ in (2.4) and $\alpha \sim N(0_p, I_p)$, then the projection kernel is

$$\widetilde{K}\left(X_1, X_2\right) = E\left[\cos\left\{\frac{|\alpha^\top(X_1 - X_2)|}{\gamma_X}\right\} \mid X_1, X_2\right] = \exp\left(-\frac{\|X_1 - X_2\|^2}{2\gamma_X^2}\right).$$

After the Gaussian projection, the non-characteristic cosine kernel (see Theorem 9 in Sriperumbudur et al. (2010)) can derive the characteristic Gaussian kernel.

3.2 The proposed statistic

Suppose that $\{z_i = (X_i, Y_i), i = 1, 2, \ldots, n\}$ and $\{(\alpha_r, \beta_r), r = 1, 2, \ldots, k\}$ are the independent copies of (X, Y) and (α, β), respectively. Define $K_{ij}^{\alpha_s} = K(\alpha_s^\top X_i, \alpha_s^\top X_j)$ and $L_{ij}^{\beta_s} = L(\beta_s^\top Y_i, \beta_s^\top Y_j)$. A natural estimator for (3.5) can be derived using generalised two-sample U-statistic of degree (4,1)

$$U_{n,k} = \left\{\binom{n}{4} k\right\}^{-1} \sum_{i<j<\ell<r}^{n} \sum_{s=1}^{k} h\left(z_i, z_j, z_\ell, z_r; \alpha_s, \beta_s\right), \tag{3.6}$$

where

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

The summation $\sum_{(t,u,v,w)}^{(1,2,3,4)}$ is over the set of all 4-permutations from $\{1, 2, 3, 4\}$. Leveraging the theory of U-statistics (see e.g. Korolyuk and Borovskich

(2013)), it is clear that $U_{n,k}$ is an unbiased estimator of KPIC. Moreover, we can apply the asymptotic theory for degenerate two-sample U-statistic (see page 158 of Korolyuk and Borovskich (2013)). When p, q are fixed, under the null hypothesis, we have $nU_{n,k} \xrightarrow{d} \sum_{i=1}^{\infty} \lambda_i(Z_i^2-1)$, where $\{Z_i, i \geq 1\}$ are i.i.d. standard normal random variables and $\{\lambda_i, i \geq 1\}$ are the eigenvalues of some operator. In general, this asymptotic null distribution is intractable and only allows for low-dimensional setting. Thus it is of great significance to obtain the asymptotic distribution of the proposed statistic in high-dimensional regime.

Remark 2. The idea of kernel projection can be further extended to the independence testing among multiple random vectors. Pfister et al. (2018) proposed dHSIC as an extension of HSIC to measure the joint independence among $d \geq 2$ random vectors. For $j \in \{1, \dots, d\}$, let $p_j \in \mathbb{N}$ and $X^{(j)}$ be a $\mathbb{R}^{p_j}$-valued random vector, and let $K^{(j)}(\cdot, \cdot)$ be a positive definite distance-based kernel on $\mathbb{R}^{p_j}$. The dHSIC can be defined similarly as

$$
\begin{aligned}
\text{dHSIC}(X^{(1)}, \dots, X^{(d)}) =& E\left\{\prod_{j=1}^{d} K^{(j)}(X_1^{(j)}, X_2^{(j)})\right\} + E\left\{\prod_{j=1}^{d} K^{(j)}(X_{2j-1}^{(j)}, X_{2j}^{(j)})\right\} \\
& - 2E\left\{\prod_{j=1}^{d} K^{(j)}(X_1^{(j)}, X_{j+1}^{(j)})\right\}.
\end{aligned} \tag{3.7}
$$

For $j \in \{1, \dots, d\}$, let $\alpha^{(j)}$ and $\mu^{(j)}$ be the p_j-dimensional projection and the corresponding probability measure respectively. Using the definition of dHSIC in (3.7), the d-variate KPIC can be defined as

$$\begin{aligned}\text{dKPIC} =&E\left\{\prod_{j=1}^{d} K^{(j)}(\alpha^{(j)\top}X_1^{(j)}, \alpha^{(j)\top}X_2^{(j)})\right\} + E\left\{\prod_{j=1}^{d} K^{(j)}(\alpha^{(j)\top}X_{2j-1}^{(j)}, \alpha^{(j)\top}X_{2j}^{(j)})\right\}\\ &- 2E\left\{\prod_{j=1}^{d} K^{(j)}(\alpha^{(j)\top}X_1^{(j)}, \alpha^{(j)\top}X_{j+1}^{(j)})\right\}.\end{aligned}$$

Similarly to Pfister et al. (2018), for a set of random samples $\{(X_i^{(1)}, \ldots, X_i^{(d)}), i = 1, 2, \ldots, n\}$ and $\{(\alpha_r^{(1)}, \ldots, \alpha_r^{(d)}), r = 1, 2, \ldots, k\}$, the sample dKPIC can be defined using V-statistic:

$$\begin{aligned}T_{n,k} =&\frac{1}{k}\sum_{s=1}^{k}\left[\frac{1}{n^2}\sum_{M_{2,n}}\prod_{j=1}^{d} K^{(j)}(\alpha_s^{(j)\top}X_{i_1}^{(j)}, \alpha_s^{(j)\top}X_{i_2}^{(j)}) + \frac{1}{n^{2d}}\sum_{M_{2d,n}}\prod_{j=1}^{d} K^{(j)}(\alpha_s^{(j)\top}X_{i_{2j-1}}^{(j)}, \alpha_s^{(j)\top}X_{i_{2j}}^{(j)})\right.\\ &\left.-\frac{2}{n^{d+1}}\sum_{M_{d+1,n}}\prod_{j=1}^{d} K^{(j)}(\alpha_s^{(j)\top}X_{i_1}^{(j)}, \alpha_s^{(j)\top}X_{i_{j+1}}^{(j)})\right],\end{aligned}$$

where $M_{q,n} = \{1, \ldots, n\}^q$ is the q-fold Cartesian product of the set $\{1, \ldots, n\}$ and $(i_1, \ldots, i_q) \in M_{q,n}$ for $n \geq 2q$.

4. Asymptotic Properties

4.1 The null distribution

Similar to (2.3), the double-centered distances based on the projection data are defined as

$$d_K^{\alpha}(X_1, X_2) = K_{12}^{\alpha} + E(K_{12}^{\alpha} \mid \alpha) - E(K_{12}^{\alpha} \mid \alpha, X_1) - E(K_{12}^{\alpha} \mid \alpha, X_2),$$

$$d_L^{\beta}(Y_1, Y_2) = L_{12}^{\beta} + E(L_{12}^{\beta} \mid \beta) - E(L_{12}^{\beta} \mid \beta, Y_1) - E(L_{12}^{\beta} \mid \beta, Y_2).$$

Denote $U(X_1, X_2) = E\{d_K^\alpha(X_1, X_2) \mid X_1, X_2\}$, $V(Y_1, Y_2) = E\{d_L^\beta(Y_1, Y_2) \mid Y_1, Y_2\}$ and $\sigma_1^2 = E\{U(X_1, X_2)\}^2 E\{V(Y_1, Y_2)\}^2$. We define the following quantities

$$g_K(X_1, X_2, X_3, X_4) = U(X_1, X_2) U(X_1, X_3) U(X_2, X_4) U(X_3, X_4),$$
$$g_L(Y_1, Y_2, Y_3, Y_4) = V(Y_1, Y_2) V(Y_1, Y_3) V(Y_2, Y_4) V(Y_3, Y_4).$$

By the Hoeffding decomposition (see e.g., Korolyuk and Borovskich (2013)), the proposed statistic (3.6) can be decomposed as $U_{n,k} = W_n + R_{n,k}$, where $W_n = \binom{n}{2}^{-1} \sum_{i<j} U(X_i, X_j) V(Y_i, Y_j)$ and $R_{n,k}$ is the remainder term. The following theorem establishes explicitly the rate of convergence for the normal approximation of the proposed statistic $U_{n,k}$.

Theorem 1. *Suppose that $E\{K(X_1, X_2)\}^4 < \infty$ and $E\{L(Y_1, Y_2)\}^4 < \infty$, then under H_0,*

$$\begin{aligned} d_{\mathcal{W}}\left\{\sqrt{\frac{n(n-1)}{2}}\frac{U_{n,k}}{\sigma_1}, Z\right\} \leq C \Big(& n^{-1/2} + \sigma_1^{-2}\left[E\left\{g_K(X_1, X_2, X_3, X_4) g_L(Y_1, Y_2, Y_3, Y_4)\right\}\right]^{1/2} \\ & + n^{-1/2}\sigma_1^{-2}\left[E\left\{U(X_1, X_2)^4 V(Y_1, Y_2)^4\right\}\right]^{1/2} \\ & + (n^{-1} + k^{-1})^{1/2}\sigma_1^{-1}[E\{d_K^\alpha(X_1, X_2)^2 d_L^\beta(Y_1, Y_2)^2\}]^{1/2}\Big), \end{aligned}$$

where $Z \sim N(0, 1)$ and C is some positive constant.

Theorem 1 quantifies the accuracy of the normal approximation and reveals how the rate of convergence depends on the sample size n, projection number k and dimensions p, q. The proof of Theorem 1 is built on Hoeffding decomposition and Stein's method. In fact, via the Hoeffding decomposition we obtain the dominating part of $U_{n,k}$ and the Wasserstein distance between the rescaled statistic and normal distribution can be derived using Stein's method.

To derive the asymptotic null normality of $U_{n,k}$, we make the following conditions on the moments of the corresponding quantities.

(A1) $E\{d_K^{\alpha}(X_1, X_2)^2 d_L^{\beta}(Y_1, Y_2)^2\} = o\{(k \wedge n)\sigma_1^2\}$.

(A2) $E\{g_K(X_1, X_2, X_3, X_4) g_L(Y_1, Y_2, Y_3, Y_4)\} = o(\sigma_1^4)$.

(A3) $E\left\{U(X_1, X_2)^4 V(Y_1, Y_2)^4\right\} = o(n\sigma_1^4)$.

Combining the above regularity conditions and the non-asymptotic bound obtained in Theorem 1 immediately yields the central limit theorem.

Corollary 1. *Suppose Conditions (A1)–(A3) hold, then* $\{n(n-1)/2\}^{1/2} U_{n,k}/\sigma_1$ *is asymptotically standard normal under* H_0 *as* n, $k \to \infty$ *and* $\min\{p, q\} \to \infty$.

Condition (A1) guarantees that the remainder term $R_{n,k}$ in the Hoeffding decomposition of $U_{n,k}$ is asymptotically negligible, while Conditions (A2)–(A3) ensure the asymptotic normality of the leading term W_n. Although Corollary 1 can be applied to the general case, the calculation of the moments involved for the general joint distribution of (X, Y) can be challenging. To gain more intuition on the Conditions (A1)–(A3), we provide a illustration in Sections S1–S2 which states that these conditions can be made more explicit.

Remark 3. Under the null hypothesis, these conditions are in fact fairly mild in the Gaussian design. For example, when the element-wise fourth moments and the eigenvalues of the covariance matrix are lower and upper bounded uniformly, Condition (A1) is satisfied when $p^2q^2\gamma_1^3\gamma_2^3/(k \wedge n) = o(1)$. Moreover, Condition (A2) can be satisfied when $p^3q^3\gamma_1^5\gamma_2^5 = o(1)$, and Condition (A3) holds true when $p^4q^4\gamma_1^6\gamma_2^6/n = o(1)$. See more details in the Sections S1–S2 of the Supplementary Material.

Remark 4. It seems from Remark 3 that γ_1 and γ_2, corresponding to the ratios of non-zero entries of the projections, can be chosen to approach zero arbitrarily fast. However, these parameters cannot approach zero at a faster rate than p^{-1}, which would result in complete loss of information since the average number of non-zero entries $\gamma_1 p \to 0$. In this paper, we assume $\gamma_1 p \to \infty$ and $\gamma_2 q \to \infty$ to avoid these extreme cases.

In order to formulate a testing procedure based on Corollary 1, we propose estimating σ_1^2 by $\widehat{\sigma}_1^2 = \{n(n-3)k^2\}^{-1}\sum_{i\neq j}^{n}\sum_{s,t}^{k} A_{ijs}A_{ijt}B_{ijs}B_{ijt}$, where A_{ijr}, B_{ijr} are the double-centered distances defined as

$$A_{ijr} = K_{ij}^{\alpha_r} - \frac{1}{n-2}\sum_{s=1}^{n} K_{sj}^{\alpha_r} - \frac{1}{n-2}\sum_{t=1}^{n} K_{it}^{\alpha_r} + \frac{1}{(n-1)(n-2)}\sum_{s,t=1}^{n} K_{st}^{\alpha_r},$$
$$B_{ijr} = L_{ij}^{\beta_r} - \frac{1}{n-2}\sum_{s=1}^{n} L_{sj}^{\beta_r} - \frac{1}{n-2}\sum_{t=1}^{n} L_{it}^{\beta_r} + \frac{1}{(n-1)(n-2)}\sum_{s,t=1}^{n} L_{st}^{\beta_r}. \tag{4.8}$$

The ratio consistency of $\hat{\sigma}_1^2$ is shown in the next theorem.

Theorem 2. *Under the conditions of Corollary 1 and* H_0*,* $\hat{\sigma}_1^2/\sigma_1^2$ *converges in probability to 1, as* n*,* $k \to \infty$ *and* $\min\{p, q\} \to \infty$*.*

For $\alpha \in (0, 1)$, denote by ξ_α the $1 - \alpha$ quantile of the standard normal distribution. Combining Corollary 1 and Theorem 2, the size-α test procedure can be defined as

$$\phi_{n,\alpha} = \mathbb{I}\left\{\sqrt{\frac{n(n-1)}{2}}\frac{U_{n,k}}{\hat{\sigma}_1} > \xi_\alpha\right\}, \tag{4.9}$$

which rejects the null hypothesis if $\{n(n-1)/2\}^{1/2}U_{n,k}/\hat{\sigma}_1 > \xi_\alpha$.

4.2 Power analysis

In this subsection, we focus on the asymptotic power performance for the proposed test. Let $G(z_1, z_2, \alpha_1, \beta_1) = d_K^{\alpha_1}(X_1, X_2)d_L^{\beta_1}(Y_1, Y_2)$. To derive limiting distribution of the proposed statistic (3.6) under H_1, we assume the following regularity conditions.

(A4) $E\left[E\left\{G(z_1, z_2, \alpha_1, \beta_1) \mid z_1\right\}\right]^2 = o(n^{-1}\sigma_1^2)$,

$$E\left[E\left\{G(z_1, z_2, \alpha_1, \beta_1) \mid z_1, \alpha_1, \beta_1\right\}\right]^2 = o(k^{-1}\sigma_1^2),$$

$$E\left[E\left\{G(z_1, z_2, \alpha_1, \beta_1) \mid \alpha_1, \beta_1\right\}\right]^2 = o(kn^{-2}\sigma_1^2).$$

(A5) $E\left\{d_K^\alpha(X_1, X_2)^2 d_L^\beta(Y_1, Y_3)^2\right\} = o\{(k \wedge n)\sigma_1^2\}$,

$$E\left\{d_K^\alpha(X_1, X_2)\right\}^2 E\{d_L^\beta(Y_1, Y_3)\}^2 = o\{(k \wedge n)\sigma_1^2\}.$$

Conditions (A4)–(A5) characterize the local alternative in an abstract way. Under the null that X and Y are independent, Condition (A4) is automatically satisfied and Condition (A5) is equivalent to Condition (A1). For the local alternative, these conditions require the alternative to be not too far away from the null. The following theorem establishes the asymptotic normality of $U_{n,k}$ under H_1.

Theorem 3. *Suppose Conditions (A1)–(A5) hold, then* $\{n(n-1)/2\}^{1/2}(U_{n,k} - \mathrm{KPIC})/\sigma_1$ *is asymptotically standard normal as* n, $k \to \infty$ *and* $\min\{p, q\} \to \infty$.

Theorem 3 indicates that the asymptotic power of the proposed test is given by

$$\beta_n = \Phi\left\{-\xi_\alpha + \sqrt{\frac{n(n-1)}{2}}\frac{\mathrm{KPIC}}{\sigma_1}\right\}, \tag{4.10}$$

which is related to the signal-to-noise ratio

$$\sqrt{\frac{n(n-1)}{2}}\frac{\mathrm{KPIC}}{\sigma_1}. \tag{4.11}$$

Now we utilize the signal-to-noise ratio in (4.11) to choose the parameters γ_1 and γ_2. To ease further illustration, we make the following notations. Let $\mathcal{S}$ be a subset of $(1, 2, \ldots, p)$ and $|\mathcal{S}|$ be the size of $\mathcal{S}$. Denote $X_{1,\mathcal{S}}$ the subvector of X_1 which the coordinates belong to $\mathcal{S}$. For example, when $\mathcal{S} = \{2, 3, 4\}$, we have $X_{1,\mathcal{S}} = (X_{1,2}, X_{1,3}, X_{1,4})$. For $|\mathcal{S}_1|$-dimensional and $|\mathcal{S}_2|$-dimensional standard normal vectors θ and η respectively, let

$$K^*(X_{1,\mathcal{S}_1}, X_{2,\mathcal{S}_1}) = E\left\{K\left(\theta_{\mathcal{S}_1}^\top X_{1,\mathcal{S}_1}, \theta_{\mathcal{S}_1}^\top X_{2,\mathcal{S}_1}\right) \mid X_1, X_2\right\},$$

$$L^*(Y_{1,\mathcal{S}_2}, Y_{2,\mathcal{S}_2}) = E\left\{L\left(\eta_{\mathcal{S}_2}^\top Y_{1,\mathcal{S}_2}, \eta_{\mathcal{S}_2}^\top Y_{2,\mathcal{S}_2}\right) \mid Y_1, Y_2\right\}. \qquad (4.12)$$

Under the projections α and β with the parameters γ_1 and γ_2 respectively, the proposed criterion in (3.5) can be equivalently expressed as

$$\mathrm{KPIC} = \sum_{t=1}^{p}\sum_{s=1}^{q} h_1(t,\gamma_1)h_2(s,\gamma_2) \sum_{|\mathcal{S}_1|=t}\sum_{|\mathcal{S}_2|=s} \mathrm{HSIC}(X_{1,\mathcal{S}_1}, Y_{1,\mathcal{S}_2}), \qquad (4.13)$$

where $h_1(t,\gamma_1) = \gamma_1^t(1-\gamma_1)^{p-t}$, $h_{2,s} = \gamma_2^s(1-\gamma_2)^{q-s}$ and the summations $\sum_{|\mathcal{S}_1|=t}$ and $\sum_{|\mathcal{S}_2|=s}$ are over all t-subsets and s-subsets of $(1, 2, \ldots, p)$ and $(1, 2, \ldots, q)$, respectively. From Proposition 1, $\mathrm{HSIC}(X_{1,\mathcal{S}_1}, Y_{1,\mathcal{S}_2})$ based on projection kernels $K^*(X_{1,\mathcal{S}_1}, X_{2,\mathcal{S}_1})$ and $L^*(Y_{1,\mathcal{S}_2}, Y_{2,\mathcal{S}_2})$ can measure the dependence between $X_{\mathcal{S}_1}$ and $Y_{\mathcal{S}_2}$. The expression in (4.13) indicates that the proposed criterion is essentially a weighted sum of HSIC between the subvectors of X and Y, which can be considered to contain low-dimensional structures. Thus it is natural to expect that the proposed test has competitive power performance when encountering high-dimensionality.

In the following theorem, we show that the signal strength (4.11) can be enhanced with specific choice of tuning parameters.

Theorem 4. *Assume that* $X_1 = (X_{1,1}, \ldots, X_{1,p})$ *has a symmetric distribution and* $\{X_{1,i}, 1 \le i \le p\}$ *are* m*-dependent for some fixed positive integer* m*. Let* $Y_1 = (Y_{1,1}, \ldots, Y_{1,p})$ *be given by* $Y_{1,i} = g_i(X_{1,i})$ *for each* $1 \le i \le p$*, where* $\{g_i(x), 1 \le i \le p\}$ *are symmetric functions satisfying* $g_i(x) = g_i(-x)$ *for* $x \in \mathbb{R}$*. Assume further that* $E(X_{1,i}^{12}) + E(Y_{1,i}^{12}) \le d_1^{12}$*,* $\mathrm{var}(X_{1,i}) \ge d_2^2$ *and* $\mathrm{var}(Y_{1,i}) \ge d_2^2$ *for some positive constants* d_1, d_2*. Suppose* $\gamma_1 = \gamma_2 = 1 \wedge (c_n n^{1/2} p^{-1})$ *where the positive sequence* c_n *satisties* $c_n \to 0$ *and* $c_n n^{1/2} \to \infty$*. Then*

$$\sqrt{\frac{n(n-1)}{2}} \frac{\mathrm{KPIC}}{\sigma_1} \to \infty \tag{4.14}$$

as $n \to \infty$ *and* $\min\{p, q\} \to \infty$*.*

The power consistency of the proposed test $\phi_{n,\alpha}$ can be established combining (4.14) and the power function (4.10), that is $\beta_n \to 1$ as $n \to \infty$, $\min\{p, q\} \to \infty$. Recently, Han and Shen (2021) worked with the Gaussian assumption $(X, Y) \sim N(0_{p+q}, \Sigma)$ and established the power consistency for the sample HSIC under the bounded spectrum condition on Σ in the high-dimensional regime of $n \to \infty$ and $\min\{p, q\} \to \infty$. Let Σ_X, Σ_Y and Σ_{XY}

be the sub-blocks of the covariance matrix $\Sigma = [\Sigma_X, \Sigma_{XY}; \Sigma_{YX}, \Sigma_Y]$. They obtained the power consistency for the sample HSIC using the signal-to-noise ratio $n\|\Sigma_{XY}\|_F^2/\|\Sigma_X\|_F\|\Sigma_Y\|_F$ without requiring the m-dependence structure assumed in Theorem 4. However, they only studied the power performance of the HSIC test in the Gaussian case that $\Sigma_{XY} \neq 0$, while the HSIC is designed to detect the complete dependence including non-linear dependence between X and Y without the Gaussian assumption. On the other hand, under the general multivariate model, Zhu et al. (2020) showed that the HSIC test has trivial limiting power when X and Y are non-linearly dependent but component-wisely uncorrelated.

Under the symmetry alternatives and the m-dependence structure in Theorem 4, there is no linear dependence between X and Y by noting that $\mathrm{cov}(X_{1,i}, Y_{1,j}) = 0$ for each $1 \leq i, j \leq p$. This clearly highlights the merit of the proposed test in detecting the pure non-linear dependence under high-dimensional setting, going beyond the scope of HSIC which has been shown merely to capture the component-wise linear dependence when encountering high-dimensional problem.

To gain some insights into the parameters $\gamma_1 = \gamma_2 = 1 \wedge (c_n n^{1/2} p^{-1})$, it can be seen that when $p = o(c_n n^{1/2})$, the choices of γ_1 and γ_2 should boil down to the completely dense case, i.e., $\gamma_1 = \gamma_2 = 1$. Then from Proposition

1, the signal-to-noise ratio in (4.11) coincides with the counterpart of HSIC in the dense case. This is consistent with the results in Gao et al. (2021), which states that the distance correlation achieves asymptotic power one when $p = o(n^{1/2})$ under the symmetry alternatives in Theorem 4. By the equivalence between distance correlation and HSIC (see Sejdinovic et al. (2013)), the proposed test can serve as an adaptable test since it has comparable power performance with HSIC in the moderately high-dimensional regime.

5. Numerical Results

In this section, we examine the simulation performance of the proposed test. Here, we consider two types of kernels, non-characteristic cosine kernel and characteristic Laplacian kernel. Denote these two tests as $\mathrm{KP_C}$ and $\mathrm{KP_L}$. In view of Theorem 4 and Remark 4, the parameters are chosen as $\gamma_1 = \gamma_2 = 1 \wedge [n^{1/2}/\{p \log(n)\}]$ where we set $c_n = \{\log(n)\}^{-1}$. Following Gretton et al. (2009), we choose the bandwidth parameter of kernel heuristically as the median distance between the sample observations. However, the theoretical results in this paper do not accommodate cases with data-dependent kernels, which also remains an open problem in the field of kernel literature (e.g., Ramdas et al. (2015), Garreau et al. (2017)). Nevertheless, the simulation results show that the median heuristic choice of bandwidth works well in most cases.

In view of the computation efficiency, we suggest to choose a relatively larger value $k = 8000$ throughout. To save space, results for the depictions of the sensitivity on the choice of k are reported in the Supplementary Material, Section S10.1. The significance level is set as $\alpha = 0.05$. Throughout these simulations, we set (n, p) to be (50,50), (50,100), (100,100), (100,300). All the numerical results will repeat 1000 times to report the empirical sizes and powers.

5.1 Comparison with other tests

In this subsection, we consider several numerical examples to compare the proposed test with some existing tests, including the distance correlation test of Székely et al. (2007) (dC), the HSIC tests of Gretton et al. (2007) using characteristic Gaussian kernel and Laplacian kernel ($\mathrm{hC_G}$ and $\mathrm{hC_L}$), the group-wise HSIC tests of Chakraborty and Zhang (2021) using characteristic Gaussian kernel and Laplacian kernel ($\mathrm{CZ_G}$ and $\mathrm{CZ_L}$), the graph-based test of Biswas et al. (2016) (BSG) and the test based on ranks of distances of Heller et al. (2013) (HHG). We first consider some simulated examples to compare the aforementioned tests for testing the independence between two random vectors in high-dimensions.

Example 5.1. 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) $X_i = (X_{i,1}, \ldots, X_{i,p})^\top$, $\{X_{i,j}\}_{j=1}^p$ are i.i.d. standardized χ^2 random variables with degree of freedom 1, $Y_i = (Y_{i,1}, \ldots, Y_{i,p})^\top$, $\{Y_{i,j}\}_{j=1}^p$ are i.i.d. standardized χ^2 random variables with degree of freedom 1;

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

Table 1 summarizes the empirical sizes of Example 5.1. The graph-based tests of Biswas et al. (2016) and Heller et al. (2013) can control the Type I error rate well in all cases. The distance correlation test of Székely et al. (2007), the HSIC tests of Gretton et al. (2007) and the group-wise HSIC tests of Chakraborty and Zhang (2021) tend to have inflated or conservative empirical sizes for Example 5.1 (ii) and (iv). The proposed tests have slightly inflated empirical sizes for Example 5.1 (ii). We provide in Figure 4–6 of the Supplementary Material the kernel density estimates of the

Table 1: Size comparison from Example 5.1

	n	p	dC	hC_G	CZ_G	KP_C	hC_L	CZ_L	KP_L	BSG	HHG
(i)	50	50	0.058	0.058	0.057	0.052	0.059	0.056	0.055	0.059	0.051
	50	100	0.050	0.051	0.049	0.073	0.048	0.048	0.060	0.043	0.048
	100	100	0.047	0.049	0.045	0.055	0.052	0.045	0.043	0.057	0.043
	100	300	0.053	0.053	0.061	0.053	0.051	0.058	0.056	0.059	0.052
(ii)	50	50	0.068	0.069	0.071	0.068	0.066	0.066	0.076	0.054	0.050
	50	100	0.071	0.076	0.071	0.070	0.071	0.072	0.070	0.056	0.044
	100	100	0.073	0.081	0.076	0.086	0.079	0.079	0.080	0.052	0.050
	100	300	0.063	0.060	0.060	0.064	0.058	0.058	0.059	0.041	0.038
(iii)	50	50	0.057	0.063	0.058	0.059	0.056	0.055	0.061	0.058	0.046
	50	100	0.067	0.063	0.052	0.048	0.064	0.051	0.058	0.053	0.044
	100	100	0.052	0.052	0.061	0.059	0.058	0.062	0.047	0.038	0.051
	100	300	0.058	0.057	0.036	0.045	0.062	0.039	0.051	0.051	0.040
(iv)	50	50	0.034	0.070	0.045	0.050	0.067	0.050	0.061	0.062	0.045
	50	100	0.047	0.069	0.047	0.059	0.067	0.062	0.061	0.046	0.049
	100	100	0.031	0.055	0.040	0.055	0.051	0.041	0.044	0.052	0.042
	100	300	0.027	0.060	0.039	0.056	0.068	0.050	0.057	0.069	0.046

standardized test statistics under the cases of Example 5.1 and compared them with the standard normal distribution. It shows that the null distribution is quite close to standard normal distribution when the dimensions and sample size increase for Example 5.1 (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 5.1 (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. For example, as analysed in Section S2 of the Supplementary Material, Condition (A2) holds true when $\sum_{i=1}^{\infty}\lambda_i^4/(\sum_{i=1}^{\infty}\lambda_i^2)^2=o(1)$ where $\{\lambda_i\}_{i=1}^{\infty}$ are the eigenvalues of an operator, and it can be viewed as a generalization of Condition (3.6) of Chen and Qin (2010) which is formulated as $\mathrm{tr}(\Sigma^4)/\{\mathrm{tr}(\Sigma^2)\}^2=o(1)$ where Σ is the covariance matrix of (X,Y). Recent studies have shown that when when there exists unbounded eigenvalues, these types of assumptions can be violated (e.g., Ma et al. (2015), Wang and Xu (2022)). For example, the equicorrelation matrix considered in Example 5.1 (ii) has unbounded eigenvalue $(1-c)+cp$. Next we compare the empirical power of the previous tests in the following examples.

Example 5.2. Generate i.i.d. samples from the following models for $i=1,\ldots,n$.

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

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

(iii) $X_i=(X_{i,1},\ldots,X_{i,p})^\top$, $\{X_{i,j}\}_{j=1}^p$ are i.i.d. standard Cauchy random variables, $Y_i=(Y_{i,1},\ldots,Y_{i,p})^\top$ where $Y_{i,j}=|X_{i,j}|^{-1}$ for $j=1,\ldots,p$;

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

Table 2 summarizes the empirical powers of Example 5.2. Notice that

Table 2: Power comparison from Example 5.2

	n	p	dC	$\mathrm{hC_G}$	$\mathrm{CZ_G}$	$\mathrm{KP_C}$	$\mathrm{hC_L}$	$\mathrm{CZ_L}$	$\mathrm{KP_L}$	BSG	HHG
(i)	50	50	0.054	0.053	0.386	0.970	0.062	0.259	0.961	0.032	0.083
	50	100	0.056	0.050	0.378	0.697	0.053	0.191	0.699	0.034	0.073
	100	100	0.051	0.069	0.737	0.979	0.060	0.427	0.984	0.028	0.096
	100	300	0.068	0.067	0.727	0.431	0.071	0.255	0.437	0.029	0.066
(ii)	50	50	0.094	0.094	1.000	1.000	0.133	1.000	1.000	0.307	0.052
	50	100	0.066	0.066	1.000	0.975	0.084	1.000	0.975	0.198	0.050
	100	100	0.077	0.077	1.000	1.000	0.121	1.000	1.000	0.370	0.067
	100	300	0.060	0.061	1.000	0.859	0.065	1.000	0.857	0.185	0.063
(iii)	50	50	0.045	0.067	0.151	0.997	0.057	0.095	0.994	0.047	0.036
	50	100	0.039	0.079	0.147	0.799	0.078	0.078	0.816	0.045	0.057
	100	100	0.030	0.061	0.256	0.995	0.062	0.089	0.997	0.055	0.047
	100	300	0.041	0.073	0.232	0.482	0.068	0.071	0.475	0.062	0.061
(iv)	50	50	0.339	0.442	0.983	1.000	0.749	0.996	1.000	0.998	1.000
	50	100	0.208	0.229	0.978	1.000	0.419	0.932	1.000	0.991	0.984
	100	100	0.318	0.462	0.979	1.000	0.845	0.998	1.000	1.000	1.000
	100	300	0.150	0.164	0.981	0.984	0.302	0.969	0.982	0.998	0.999

in Example 5.2, the m-dependence condition of X and the coordinate-wise non-linear dependence between X and Y assumed in Theorem 4 are satisfied. From Table 2, we can observe that the performance of distance correlation test and the HSIC tests deteriorate quickly in the high-dimensional regime. This is in line with the aforementioned issue that HSIC and distance correlation can only catch linear dependence in high-dimension. The graph-based tests appear to be very ineffective except in Example 5.2 (iv). The group-wise HSIC tests have relatively good performance in Example 5.2 (i), (ii) and (iv). In contrast, the proposed tests are much more powerful in capturing the coordinate-wise non-linear dependence in the high-dimensional regime. Furthermore, as commented in Section 3.1, KPIC only requires kernels with positive definite function in $\mathbb{R}$ (e.g., cosine kernel), as opposed to the requirement of characteristic kernels for the existing kernel-based tests. It can be seen that the proposed test based on cosine kernel performs reasonably well in these cases. This suggests that the proposed test has wider application prospects compared to the other kernel-based tests. See Remark 1 for more details.

In the Supplementary Material, Section S10.3, we present the additional simulated examples that does not meet the conditions in Theorem 4, such as the non-coordinate-wise dependence between X and Y. It shows that the proposed tests still have higher power than other tests in most cases.

5.2 Real data analysis

We consider the independence testing problem on Earthquakes data. The data set is orginally from the Northern California Earthquake Data Center and has classes of positive and negative major earthquake events. It can be download from UCR Time Series Classification Archive (Dau et al. (2019)) and has been analyzed by Zhu et al. (2020) and Chakraborty and Zhang (2021) in a similar way. Each data point is of length 512, which is an averaged reading of 512 surrounding areas for one hour. Let the reading vector $Z_i = (Z_{i,1}, Z_{i,2}, \ldots, Z_{i,512})$ for $i = 1, \ldots, 461$. According to the data description, a major earthquake event is defined as any reading of over 5 on the Richter scale. A positive case is defined as a major event which is not preceded by another major event for at least 512 hours, while the negative cases are instances where there is a reading below 4 and is preceded by at least 20 non-zero readings in the previous 512 hours. There are 93 positive cases and 368 negative cases. We are interested in the correlation of earthquake readings between two regions which are composed of different surrounding areas, and proceed by testing the independence between two random vectors of the selected areas from the reading vectors. For the two regions, we select p different areas starting from the center of Z_i and moving forward and backward respectively, that is, we set $X_i = (Z_{i,256-p+1}, \ldots, Z_{i,256})$ and $Y_i = (Z_{i,257}, \ldots, Z_{i,256+p})$ for $i = 1, \ldots, 461$. We consider different cases for $(n, p) = (50, 50)$, (50,100), (100,100), (100,256). We will randomly sample n rows from the full dataset $\{(X_i, Y_i)\}_{i=1}^{461}$ without replacement.

Table 3 shows the rejection rates corresponding to the different tests. X and Y are expected to be dependent due to the serial nature of Z_i, which means that components arranged in sequence correspond to the readings from near areas. The graph-based tests have relatively poor power performance especially when the sample size is small. The tests based on the proposed criterion and the group-wise tests have higher power as compared to the tests based on distance correlation and HSIC. These results indicate that the dependence among the earthquake readings of the surrounding areas is non-linear and thus cannot be fully detected by traditional distance and kernel-based metrics in the high-dimensional regime.

Table 3: Rejection rates of tests for Earthquakes data

n	p	dC	$\mathrm{hC_G}$	$\mathrm{CZ_G}$	$\mathrm{KP_C}$	$\mathrm{hC_L}$	$\mathrm{CZ_L}$	$\mathrm{KP_L}$	BSG	HHG
50	50	0.070	0.071	0.988	1.000	0.079	0.991	1.000	0.051	0.077
50	100	0.078	0.080	1.000	1.000	0.088	1.000	1.000	0.010	0.343
100	100	0.137	0.135	1.000	1.000	0.167	1.000	1.000	0.003	0.665
100	256	0.243	0.242	1.000	1.000	0.455	1.000	1.000	0.000	1.000

Supplementary Materials

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

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No 11971478) and MOE Project of Key Research Institute of Humanities and Social Sciences (22JJD910001).

References

Bach, F. R. and M. I. Jordan (2002). Kernel independent component analysis. *Journal of Machine Learning Research 3*(Jul), 1–48.

Banerjee, B. and A. K. Ghosh (2022). Test of independence for hilbertian random variables. *Stat 11*(1), e474.

Bergsma, W. and A. Dassios (2014). A consistent test of independence based on a sign covariance related to kendalls tau. *Bernoulli 20*(2), 1006–1028.

Biswas, M., S. Sarkar, and A. K. Ghosh (2016). On some exact distribution-free tests of independence between two random vectors of arbitrary dimensions. *Journal of Statistical Planning and Inference 175*, 78–86.

Blum, J. R., J. Kiefer, and M. Rosenblatt (1961). *Distribution free tests of independence based on the sample distribution function.* Sandia Corporation.

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.

Dau, H. A., A. Bagnall, K. Kamgar, C.-C. M. Yeh, Y. Zhu, S. Gharghabi, C. A. Ratanamahatana, and E. Keogh (2019). The ucr time series archive. *IEEE/CAA Journal of Automatica Sinica 6*(6), 1293–1305.

Fang, K.-T., S. Kotz, and K. W. Ng (2018). *Symmetric multivariate and related distributions*. Chapman and Hall/CRC.

Fukumizu, K., A. Gretton, B. Schölkopf, and B. K. Sriperumbudur (2008). Characteristic kernels on groups and semigroups. *Advances in Neural Information Processing Systems 21*, 473–480.

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.

Garreau, D., W. Jitkrittum, and M. Kanagawa (2017). Large sample analysis of the median heuristic. *arXiv preprint arXiv:1707.07269*.

Gieser, P. W. and R. H. Randles (1997). A nonparametric test of independence between two vectors. *Journal of the American Statistical Association 92*(438), 561–567.

Gretton, A., K. Fukumizu, Z. Harchaoui, and B. K. Sriperumbudur (2009). A fast, consistent kernel two-sample test. *Advances in Neural Information Processing Systems 22*.

Gretton, A., K. Fukumizu, C. Teo, L. Song, B. Schölkopf, and A. Smola (2007). A kernel statistical test of independence. *Advances in Neural Information Processing Systems 20*.

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

Heller, R., M. Gorfine, and Y. Heller (2012). A class of multivariate distribution-free tests of independence based on graphs. *Journal of Statistical Planning and Inference 142*(12), 3097–3106.

Heller, R., Y. Heller, and M. Gorfine (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika 100*(2), 503–510.

Hoeffding, W. (1994). A non-parametric test of independence. In *The Collected Works of Wassily Hoeffding*, pp. 214–226. Springer.

Hofmann, T., B. Schölkopf, and A. J. Smola (2008). Kernel methods in machine learning. *The Annals of Statistics 36*(3), 1171–1220.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika 30*(1/2), 81–93.

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

Kong, J., S. Wang, and G. Wahba (2015). Using distance covariance for improved variable selection with application to learning genetic risk models. *Statistics in Medicine 34*(10), 1708–1720.

Korolyuk, V. S. and Y. V. Borovskich (2013). *Theory of U-statistics*, Volume 273. Springer Science & Business Media.

Leung, D. and M. Drton (2018). Testing independence in high dimensions with sums of rank correlations. *The Annals of Statistics 46*(1), 280–307.

Ma, Y., W. Lan, and H. Wang (2015). A high dimensional two-sample test under a low dimensional factor structure. *Journal of Multivariate Analysis 140*, 162–170.

Pan, W., X. Wang, H. Zhang, H. Zhu, and J. Zhu (2020). Ball covariance: A generic measure of dependence in banach space. *Journal of the American Statistical Association 115*(529), 307–317.

Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London 58*, 240–243.

Pfister, N., P. Bühlmann, B. Schölkopf, and J. Peters (2018). Kernel-based tests for joint independence. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 80*(1), 5–31.

Ramdas, A., S. J. Reddi, B. Póczos, A. Singh, and L. Wasserman (2015). On the decreasing power of kernel and distance based nonparametric hypothesis tests in high dimensions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 29, pp. 3571–3577.

Roy, A., S. Sarkar, A. K. Ghosh, and A. Goswami (2020). On some consistent tests of mutual independence among several random vectors of arbitrary dimensions. *Statistics and Computing 30*(6), 1707–1723.

Sejdinovic, D., B. Sriperumbudur, A. Gretton, and K. Fukumizu (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics 41*(5), 2263–2291.

Sinha, B. K. and H. Wieand (1977). Multivariate nonparametric tests for independence. *Journal of Multivariate Analysis 7*(4), 572–583.

Spearman, C. (1961). The proof and measurement of association between two things.

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.

Sriperumbudur, B. K., A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. Lanckriet (2010). Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research 11*, 1517–1561.

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

Székely, G. J. and M. L. Rizzo (2013). The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis 117*, 193–213.

Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics 35*(6), 2769–2794.

Wang, R. and W. Xu (2022). An approximate randomization test for high-dimensional two-sample behrens-fisher problem under arbitrary covariances. *Biometrika 109*, 1117–1132.

Wang, R. and X. Xu (2018). On two-sample mean tests under spiked covariances. *Journal of*

Multivariate Analysis 167, 225–249.

Weihs, L., M. Drton, and N. Meinshausen (2018). Symmetric rank covariances: a generalized framework for nonparametric measures of dependence. *Biometrika 105*(3), 547–562.

Wendland, H. (2004). *Scattered data approximation*, Volume 17. Cambridge university press.

Yan, J. and X. Zhang (2021). Kernel two-sample tests in high dimension: Interplay between moment discrepancy and dimension-and-sample orders. *arXiv preprint arXiv:2201.00073*.

Zhu, C., X. Zhang, S. Yao, and X. Shao (2020). Distance-based and rkhs-based dependence metrics in high dimension. *The Annals of Statistics 48*(6), 3366–3394.

Zhu, L., K. Xu, R. Li, and W. Zhong (2017). Projection correlation between two random vectors. *Biometrika 104*(4), 829.

Yuexin Chen

Center for Applied Statistics and School of Statistics, Renmin University of China.

E-mail: chenyuexin0107@ruc.edu.cn

Wangli Xu (corresponding author)

Center for Applied Statistics and School of Statistics, Renmin University of China.

E-mail: wlxu@ruc.edu.cn