*The Annals of Statistics*
2020, Vol. 48, No. 6, 3366–3394
https://doi.org/10.1214/19-AOS1934


# DISTANCE-BASED AND RKHS-BASED DEPENDENCE METRICS IN HIGH DIMENSION

BY CHANGBO ZHU[1,*], XIANYANG ZHANG[2], SHUN YAO[3] AND XIAOFENG SHAO[1,†]

[1]*Department of Statistics, University of Illinois at Urbana-Champaign,* [*]*changbo2@illinois.edu;* [†]*xshao@illinois.edu*

[2]*Department of Statistics, Texas A&M University, zhangxiany@stat.tamu.edu*

[3]*Quantitative Researcher, Point72 Asset Management, cnshunyao@gmail.com*

In this paper, we study distance covariance, Hilbert–Schmidt covariance (aka Hilbert–Schmidt independence criterion [In *Advances in Neural Information Processing Systems* (2008) 585–592]) and related independence tests under the high dimensional scenario. We show that the sample distance/Hilbert–Schmidt covariance between two random vectors can be approximated by the sum of squared componentwise sample cross-covariances up to an asymptotically constant factor, which indicates that the standard distance/Hilbert–Schmidt covariance based test can only capture linear dependence in high dimension. Under the assumption that the components within each high dimensional vector are weakly dependent, the distance correlation based $t$ test developed by Székely and Rizzo (*J. Multivariate Anal.* **117** (2013) 193–213) for independence is shown to have trivial limiting power when the two random vectors are nonlinearly dependent but componentwisely uncorrelated. This new and surprising phenomenon, which seems to be discovered and carefully studied for the first time, is further confirmed in our simulation study. As a remedy, we propose tests based on an aggregation of marginal sample distance/Hilbert–Schmidt covariances and show their superior power behavior against their joint counterparts in simulations. We further extend the distance correlation based $t$ test to those based on Hilbert–Schmidt covariance and marginal distance/Hilbert–Schmidt covariance. A novel unified approach is developed to analyze the studentized sample distance/Hilbert–Schmidt covariance as well as the studentized sample marginal distance covariance under both null and alternative hypothesis. Our theoretical and simulation results shed light on the limitation of distance/Hilbert–Schmidt covariance when used jointly in the high dimensional setting and suggest the aggregation of marginal distance/Hilbert–Schmidt covariance as a useful alternative.

**1. Introduction.** Testing for independence between two random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ is a fundamental problem in statistics. There is a huge literature in the low dimensional context. Here, we mention rank correlation coefficients based tests and nonparametric Cramér–von Mises-type statistics in Hoeffding (1948), Blum, Kiefer and Rosenblatt (1961), de Wet (1980); tests based on signs or empirical characteristic functions, see Sinha and Wieand (1977), Deheuvels (1981), Csörgő (1985), Hettmansperger and Oja (1994), Gieser and Randles (1997), Taskinen, Kankainen and Oja (2003) among others; tests based on recently developed nonlinear dependence metrics that target at nonlinear and nonmonotone dependence include distance covariance [Székely, Rizzo and Bakirov (2007)], Hilbert–Schmidt independence criterion (HSIC) [Gretton et al. (2008)] (aka Hilbert–Schmidt covariance in this work) and sign covariance [Bergsma and Dassios (2014)]. Also see Berrett and Samworth (2019) for some recent work on independence testing via mutual information.

Received April 2019; revised November 2019.
*MSC2020 subject classifications.* Primary 62G10, 60K35; secondary 62G20.
*Key words and phrases.* Distance covariance, high dimensionality, Hilbert–Schmidt independence criterion, independence test, $\mathcal{U}$-statistics.

In the high dimensional setting, the literature is scarce. Székely and Rizzo (2013) extended the distance correlation proposed in Székely, Rizzo and Bakirov (2007) to the problem of testing independence of two random vectors under the setting that the dimensions $p$ and $q$ grow while sample size $n$ is fixed. This setting is known as high dimension, low sample size (HDLSS) in the literature and has been adopted in Hall, Marron and Neeman (2005), Ahn et al. (2007), Jung and Marron (2009) and Wei et al. (2016), etc. A closely related asymptotic framework is the high dimension medium sample size (HDMSS) [Aoshima et al. (2018)], where $n \wedge p \wedge q \to \infty$ with $p$, $q$ growing more rapidly. Among the recent work that is related to independence testing in the high dimensional setting, Pan, Gao and Yang (2014) proposed tests of independence among a large number of high dimensional random vectors using insights from random matrix theory; Yang and Pan (2015) proposed a new statistic based on the sum of regularized sample canonical correlation coefficients of $X$ and $Y$, which is limited to testing for uncorrelatedness due to the use of canonical correlation. Leung and Drton (2018) proposed to test for mutual independence of high dimensional vectors using sum of pairwise rank correlations and sign covariances; Yao, Zhang and Shao (2018) addressed the mutual independence testing problem in the high dimensional context by using sum of pairwise squared sample distance covariances; Zhang, Yao and Shao (2018) proposed a $L^2$ type test for conditional mean/quantile dependence of a univariate response variable given a high dimensional covariate vector based on martingale difference divergence [Shao and Zhang (2014)], which is an extension of distance covariance to quantify (conditional) mean dependence.

Distance covariance/correlation was first introduced in Székely, Rizzo and Bakirov (2007) and has received much attention since then. Owing to its notable ability to quantify any types of dependence including nonmonotone, nonlinear dependence and also the flexibility to be applicable to two random vectors in arbitrary, not necessarily equal dimensions, a lot of research work has been done to extend and apply distance covariance into many modern statistical problems; see, for example, Kong et al. (2012), Li, Zhong and Zhu (2012), Zhou (2012), Lyons (2013), Székely and Rizzo (2014), Dueck et al. (2014), Shao and Zhang (2014), Park, Shao and Yao (2015), Matteson and Tsay (2017), Zhang, Yao and Shao (2018), Edelmann, Richards and Vogel (2017), Yao, Zhang and Shao (2018) among others. In this paper, we shall revisit the test proposed by Székely and Rizzo (2013), which seems to be the only test in the high dimensional setting that captures nonlinear and nonmonotonic dependence. Unlike the positive finding reported in Székely and Rizzo (2013), we obtained some negative results that show the limitation of distance covariance/correlation in the high dimensional context.

Specifically, we show that for two random vectors $X = (x_1, \dots, x_p)^T \in \mathbb{R}^p$ and $Y = (y_1, \dots, y_q)^T \in \mathbb{R}^q$ with finite componentwise second moments, as $p, q \to \infty$ and $n$ can either be fixed or grows to infinity at a slower rate,

$$\text{dCov}_n^2(\mathbf{X}, \mathbf{Y}) \approx \frac{1}{\tau} \sum_{i=1}^{p} \sum_{j=1}^{q} \text{cov}_n^2(\mathcal{X}_i, \mathcal{Y}_j), \tag{1}$$

where $X_k \stackrel{d}{=} X$ and $Y_k \stackrel{d}{=} Y$ are independent samples, $\mathcal{X}_i$ and $\mathcal{Y}_j$ are the componentwise samples, $\mathbf{X} = (X_1, X_2, \dots, X_n)^T = (\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_p)$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T = (\mathcal{Y}_1, \mathcal{Y}_2, \dots, \mathcal{Y}_q)$ denote the sample matrices, $\text{dCov}_n^2(\mathbf{X}, \mathbf{Y})$ is the unbiased sample distance covariance, $\tau$ is a constant quantity depending on the marginal distributions of $X$ and $Y$ as well as $p$ and $q$, $\text{cov}_n^2(\mathcal{X}_i, \mathcal{Y}_j)$ is an unbiased sample estimate of $\text{cov}^2(x_i, y_j)$ to be defined later. To the best of our knowledge, this is the first work in the literature uncovering the connection between sample distance covariance and sample covariance, the latter of which can only measure the linear dependence between two random variables. This approximation suggests that the distance covariance can only measure linear dependence in the high dimensional setting although it is well known to be capable of capturing nonlinear dependence in the fixed dimensional case.

Gretton et al. (2008) proposed Hilbert–Schmidt independence criterion (aka Hilbert–Schmidt covariance in this paper), which can be seen as a generalization of distance covariance by kernelizing the $L^2$ distance as shown by Sejdinovic et al. (2013). Despite the kernelization process, we show that the Hilbert–Schmidt covariance (hCov) enjoys similar approximation property under high dimension low/medium sample size setting, that is,

$$\mathrm{hCov}_n^2(\mathbf{X},\mathbf{Y}) \approx A_p B_q \times \frac{1}{\tau^2}\sum_{i=1}^{p}\sum_{j=1}^{q}\mathrm{cov}_n^2(\mathcal{X}_i,\mathcal{Y}_j), \tag{2}$$

where $\mathrm{hCov}_n^2(\mathbf{X},\mathbf{Y})$ is the unbiased sample Hilbert–Schmidt covariance, $A_p$ and $B_q$ both converge in probability to constants that depend on the pre-chosen kernels. This approximation also suggests that when the dimension is high, the standard Hilbert–Schmidt covariance (hCov) applied to the whole components of the vectors also exhibits the loss of power when $X$ and $Y$ are nonlinearly dependent, but componentwisely uncorrelated or weakly correlated.

As a natural remedy, we propose a distance covariance based marginal test statistic, that is,

$$\mathrm{mdCov}_n^2(\mathbf{X},\mathbf{Y}) = \sqrt{\binom{n}{2}}\sum_{i=1}^{p}\sum_{j=1}^{q}\mathrm{dCov}_n^2(\mathcal{X}_i,\mathcal{Y}_j).$$

This test statistic is an aggregate of the componentwise sample distance covariances and captures the component by component nonlinear dependence. Similarly, the marginal Hilbert–Schmidt covariance (mhCov) is defined as

$$\mathrm{mhCov}_n^2(\mathbf{X},\mathbf{Y}) = \sqrt{\binom{n}{2}}\sum_{i=1}^{p}\sum_{j=1}^{q}\mathrm{hCov}_n^2(\mathcal{X}_i,\mathcal{Y}_j).$$

The distance covariance, Hilbert–Schmidt covariance, marginal distance covariance and marginal Hilbert–Schmidt covariance based tests can be carried out by standard permutation procedures. The superiority of mdCov and mhCov based tests over its joint counterparts in power is demonstrated in the simulation studies. On the other hand, Székely and Rizzo (2013) discussed the distance correlation (dCor) based $t$-test under HDLSS and derived the limiting null distribution of the test statistic under suitable assumptions. We consider the same $t$-test statistic and further extend to Hilbert–Schmidt covariance (hCov), marginal distance covariance (mdCov) and marginal Hilbert–Schmidt covariance (mhCov). To derive the asymptotic distribution of studentized version of dCov, hCov, mdCov and mhCov under both the null of independence (for HDLSS and HDMSS setting) and some specific alternative classes (for HDLSS setting), we develop a novel unified approach. In particular, we define a unified quantity (uCov) based on the bivariate kernel $k$ and show that under HDLSS setting, properly scaled $\mathrm{dCov}_n^2$, $\mathrm{hCov}_n^2$ and $\mathrm{mdCov}_n^2$ are all asymptotically equal to $\mathrm{uCov}_n^2$ up to different choices of kernels, that is,

$$\begin{array}{ll}\left.\begin{array}{l}\mathrm{dCov}_n^2(\mathbf{X},\mathbf{Y}) \approx a \times \mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y})\\ \mathrm{hCov}_n^2(\mathbf{X},\mathbf{Y}) \approx A_p B_q \times \mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y})\end{array}\right\} & \text{when } k(x,y)=|x-y|^2,\\ \left.\mathrm{mdCov}_n^2(\mathbf{X},\mathbf{Y}) = b \times \mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y})\right\} & \text{when } k(x,y)=|x-y|,\end{array} \tag{3}$$

where $a$, $b$ are constants and $A_p$, $B_p$ both converge in probability to constants. Next, we show that

$$\begin{cases} \mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y}) \xrightarrow{d} \dfrac{2}{n(n-3)}\mathbf{c}^T\mathbf{M}\mathbf{d}, & \text{under HDLSS,} \\ C_{n,p,q}\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y}) \xrightarrow{d} N(0,1), & \text{under HDMSS,} \end{cases}$$

where $\mathbf{c}$, $\mathbf{d}$ are jointly Gaussian, $\mathbf{M}$ is a projection matrix and $C_{n,p,q}$ is a normalizing constant. Thus, we can easily apply the above results to dCov, hCov and mdCov-based $t$-test statistics using (3). The unified approach still works for mhCov-based $t$-test if we consider the bandwidth parameters appeared in the kernel distance to be fixed constants. However, we encounter technical difficulties if the bandwidth parameters along each dimension depends on the whole componentwise samples, since this makes the pairwise sample distance correlated with each other and complicates the asymptotic analysis.

We obtain the same limiting null distribution as Székely and Rizzo (2013) and further show that this test statistic has a trivial power against the alternative where $X$ and $Y$ are nonlinearly dependent, but component-wisely uncorrelated. This clearly demonstrates that the distance covariance/correlation based joint independence test (i.e., treating all components of a vector as a whole jointly) fails to capture the nonlinear dependence in high dimension. This phenomenon is new and was not reported in Székely and Rizzo (2013). It shows that there might be some intrinsic difficulties for standard distance covariance to capture the nonlinear dependence when the dimension is high and provide a cautionary note on the use of distance covariance/correlation directly to the whole components of high dimensional data. Besides, we have the following additional contributions relative to Székely and Rizzo (2013): (i) we relax the componentwise i.i.d. assumption used for asymptotic analysis; (ii) the limiting distributions are derived under both the null and certain classes of alternative hypothesis for the HDLSS framework; (iii) our unified approach holds for any bivariate kernel that has continuous second-order derivative in a neighborhood containing 1; (iv) our approach is built on some new technical arguments which reveal some insights on $\mathcal{U}$-centering; (v) the limiting null distribution is also derived under the HDMSS setting.

As pointed out by a referee, there is an awareness of the importance of choosing the right kernel in the machine learning community and it may be clear to many researchers in this area that the standard distance covariance (or Hilbert–Schmidt covariance) is not the right choice when the dimension is moderate or high. It is worth noting that the phenomenon of decreasing power with higher dimension for Hilbert–Schmidt covariance (with Gaussian/Laplacian kernels) based independence test has been observed in Ramdas et al. (2015), but they did not provide a complete theoretical explanation. In this sense, our theory to a large extent settles their conjecture and offers a deeper understanding about the dimension's impact on the behavior of Hilbert–Schmidt covariance. In addition, the marginally aggregated distance/Hilbert–Schmidt covariance can be viewed as a kind of distance/Hilbert–Schmidt covariance with the Euclidean norm/kernel replaced by a new norm/kernel; see Remark 2.2. Therefore, our theory demonstrates the limitation of using Euclidean norm in distance/Hilbert–Schmidt covariance in the high dimensional setting and motivates the question of what is the optimal norm or kernel for independence testing in the high dimension, which is left for future research.

Standard Distance and Hilbert–Schmidt (with Gaussian kernel or Laplacian kernel) covariance have been frequently applied to testing dependence between high dimensional vectors in biological science; see Kroupi et al. (2012, 2014), Hua and Ghosh (2015), Yang (2017), etc. In particular, Kroupi et al. (2014) used Hilbert–Schmidt (with Gaussian or Laplacian kernel) covariance to test the dependence between EEG signals for the perception of pleasant and unpleasant odors across subjects, where the data are collected for 5 subjects, each with 18 trials and dimension 250. Hua and Ghosh (2015) used Hilbert–Schmidt covariance with Gaussian kernel to examine the association between phenotype variable and genotype variable. The Alzheimer's Disease Neuroimaging Initiative (ADNI) data is used in their simulation studies, where phenotype variable has dimension 119 and genotype variable has dimension 141. Finally, Hilbert–Schmidt covariance with Gaussian kernel is used by Yang (2017) to conduct independence test between neural responses and visual features, where the dimensions are of several hundreds. In the machine learning community, the application of Hilbert–Schmidt (with linear, Gaussian or Laplacian kernel) covariance for high dimensional data involves multilabel dimension reduction [Zhang and Zhou (2010), Xu et al. (2016), Mikalsen et al. (2019), etc] and unsupervised feature selection [Bedo (2008)], among others. In all the above mentioned applications, the dimensions of the vectors involved are at hundreds or thousands.

1.1. *Notation.* In this paper, random data samples are denoted as, for each $i = 1, 2, \dots, n$, $X_i \overset{d}{=} X = (x_1, \dots, x_p)^T \in \mathbb{R}^p$, $Y_i \overset{d}{=} Y = (y_1, \dots, y_q)^T \in \mathbb{R}^q$. Next, let $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$ denote the random sample matrices. In addition, the random componentwise samples are denoted as $\mathcal{X}_1, \dots, \mathcal{X}_p$ and $\mathcal{Y}_1, \dots, \mathcal{Y}_q$, which are illustrated in the following table:

$$
\begin{array}{cc|cccc}
 & & \mathcal{X}_1 & \mathcal{X}_2 & \cdots & \mathcal{X}_p \\
 & & \Downarrow & & & \\
X_1^T & \Rightarrow & x_{11} & x_{12} & \cdots & x_{1p} \\
X_2^T & & x_{21} & x_{22} & \cdots & x_{2p} \\
\vdots & & \vdots & \vdots & & \vdots \\
X_n^T & & x_{n1} & x_{n2} & \cdots & x_{np}
\end{array}
\;\Leftarrow \mathbf{X}
\qquad
\mathbf{Y} \Rightarrow
\begin{array}{cccc|cc}
\mathcal{Y}_1 & \mathcal{Y}_2 & \cdots & \mathcal{Y}_q & & \\
 & & & \Downarrow & & \\
y_{11} & y_{12} & \cdots & y_{1q} & & Y_1^T \\
y_{21} & y_{22} & \cdots & y_{2q} & & Y_2^T \\
\vdots & \vdots & & \vdots & & \vdots \\
y_{n1} & y_{n2} & \cdots & y_{nq} & \Leftarrow & Y_n^T
\end{array}
$$

Furthermore, matrices are denoted by upper case boldface letters (e.g., $\mathbf{A}$, $\mathbf{B}$). For any matrix $\mathbf{A} = (a_{st}) \in \mathbb{R}^{n \times n}$, we use $\widetilde{\mathbf{A}} = (\tilde{a}_{st}) \in \mathbb{R}^{n \times n}$ to denote the $\mathcal{U}$-centered version of $\mathbf{A}$, that is,

$$
\tilde{a}_{st} = \begin{cases} a_{st} - \dfrac{1}{n-2} \displaystyle\sum_{v=1}^{n} a_{sv} - \dfrac{1}{n-2} \displaystyle\sum_{u=1}^{n} a_{ut} + \dfrac{1}{(n-1)(n-2)} \displaystyle\sum_{u,v=1}^{n} a_{uv}, & s \neq t, \\ 0, & s = t. \end{cases}
$$

Following Székely and Rizzo (2014), the inner product between two $\mathcal{U}$-centered matrices $\widetilde{\mathbf{A}} = (\tilde{a}_{st}) \in \mathbb{R}^{n \times n}$ and $\widetilde{\mathbf{B}} = (\tilde{b}_{st}) \in \mathbb{R}^{n \times n}$ is defined as

$$
(\widetilde{\mathbf{A}} \cdot \widetilde{\mathbf{B}}) := \frac{1}{n(n-3)} \sum_{s \neq t} \tilde{a}_{st} \tilde{b}_{st}.
$$

Next, we use $\mathbf{1}_n$ to denote the $n$ dimensional column vector whose entries are all equal to 1. Similarly, we use $\mathbf{0}_n$ to denote the $n$ dimensional column vector whose entries are all equal to 0. Finally, we use $|\cdot|$ to denote the $L^2$ norm of a vector, $(X', Y')$ and $(X'', Y'')$ to be independent copies of $(X, Y)$ and $X \perp Y$ to indicate that $X$ and $Y$ are independent.

We utilize the order in probability notations such as stochastic boundedness $O_p$ (big O in probability), convergence in probability $o_p$ (small o in probability) and equivalent order $\asymp_p$, which is defined as follows: for a sequence of random variables $\{Z_s\}_{s \in \mathbb{Z}}$ and a sequence of numbers $\{a_s\}_{s \in \mathbb{Z}}$, $Z_s \asymp_p a_s$ if and only if $Z_s / a_s = O_p(1)$ and $a_s / Z_s = O_p(1)$ as $s \to \infty$. For more details about this notation, please see DasGupta (2008).

## 2. High dimension low sample size.

The analyses in this section are conducted under the HDLSS setting, that is, the sample size $n$ is fixed and the dimensions $p \wedge q \to \infty$.

2.1. *Distance covariance and variants.* In this section, we introduce the following test statistics based on distance covariance (dCov), marginal distance covariance (mdCov), Hilbert–Schmidt covariance (hCov) and marginal Hilbert–Schmidt covariance (mhCov). In addition, their asymptotic behaviors under the HDLSS setting are derived. The following moment conditions will be used throughout the paper.

ASSUMPTION D1. For any $p, q$, the variance and the second moment of any coordinate of $X = (x_1, x_2, \dots, x_p)^T$ and $Y = (y_1, y_2, \dots, y_q)^T$ is uniformly bounded below and above, that is,

$$0 < a \le \inf_i \operatorname{var}(x_i) \le \sup_i \mathrm{E}(x_i^2) \le b < \infty,$$

$$0 < a' \le \inf_j \operatorname{var}(y_j) \le \sup_j \mathrm{E}(y_j^2) \le b' < \infty,$$

for some constants $a, b, a', b'$.

Next, denote $\tau_X^2 = \mathrm{E}|X - X'|^2$, $\tau_Y^2 = \mathrm{E}|Y - Y'|^2$ and $\tau^2 := \tau_X^2 \tau_Y^2 = \mathrm{E}|X - X'|^2 \mathrm{E}|Y - Y'|^2$. Notice that under Assumption D1, it can be easily seen that

$$\tau_X \asymp \sqrt{p}, \qquad \tau_Y \asymp \sqrt{q} \quad \text{and} \quad \tau \asymp \sqrt{pq}.$$

The statistics we study in this work use the pairwise $L^2$ distance between data points. The following proposition presents an expansion formula on the normalized $L^2$ distance when the dimension is high, which plays a key role in our theoretical analysis.

PROPOSITION 2.1. *Under Assumption* D1, *we have*

$$\frac{|X - X'|}{\tau_X} = 1 + \frac{1}{2} L_X(X, X') + R_X(X, X'),$$

*where*

$$L_X(X, X') := \frac{|X - X'|^2 - \tau_X^2}{\tau_X^2},$$

*and $R_X(X, X')$ is the remainder term. If we further assume that as $p \wedge q \to \infty$, $L_X(X, X') = o_p(1)$, then $R_X(X, X') = O_p(L_X(X, X')^2)$. A similar result holds for $Y$.*

In order for the approximations in equations (1) and (2) to work well, it is required that $L_X(X_s, X_t)$ and $L_Y(Y_s, Y_t)$ should decay relatively fast as $p \wedge q \to \infty$. The following assumption specifies the order of $L_X(X_s, X_t)$ and $L_Y(Y_s, Y_t)$.

ASSUMPTION D2. $L_X(X, X') = O_p(a_p)$ and $L_Y(Y, Y') = O_p(b_q)$, where $a_p, b_q$ are sequences of numbers such that

$$a_p = o(1), \qquad b_q = o(1),$$

$$\tau_X^2 a_p^3 = o(1), \qquad \tau_Y^2 b_q^3 = o(1), \qquad \tau a_p^2 b_q = o(1), \qquad \tau a_p b_q^2 = o(1).$$

REMARK 2.1. A sufficient condition for $L_X(X, X') = o_p(1)$ is that $\mathrm{E}[L_X(X, X')^2] = o(1)$. Let $\mathbf{\Sigma}_X = \operatorname{cov}(X)$. By a straightforward calculation, we obtain $|X - X'|^2 = \sum_{j=1}^p (x_j - x_j')^2$, $\mathrm{E}|X - X'|^2 = 2\sum_{j=1}^p \operatorname{var}(x_j) = 2\operatorname{tr}(\mathbf{\Sigma}_X)$, and

$$\mathrm{E}[L_X(X, X')^2] = \frac{\sum_{j,j'=1}^p [\operatorname{cov}(x_j^2, x_{j'}^2) + 2\operatorname{cov}^2(x_j, x_{j'})]}{2\operatorname{tr}^2(\mathbf{\Sigma}_X)}.$$

Therefore, $\mathrm{E}[L_X(X, X')^2] = o(1)$ holds if the componentwise dependence within $X$ is not too strong. To illustrate this point, we consider the general multivariate model,

$$X_{p\times 1} = \mathbf{A}_{p\times s_1} U_{s_1\times 1} + \mu_{p\times 1},$$

where $\mathbf{A}$ is a constant matrix with $s_1 \geq p$, $\mu$ is the mean vector for $X$, and $U = (u_1, \dots, u_{s_1})^T$ has i.i.d. components with mean zero and variance one. Suppose

$$\frac{\mathrm{tr}(\mathbf{A}\mathbf{A}^T\mathbf{A}\mathbf{A}^T)}{\mathrm{tr}^2(\mathbf{A}\mathbf{A}^T)} = \frac{\mathrm{tr}(\Sigma_X^2)}{\mathrm{tr}^2(\Sigma_X)} = O(p^{-1})$$

and $\sup_s \mathrm{E}[u_s^4] < \infty$. Then the multivariate model satisfies Assumption D2 with $a_p = 1/\sqrt{p}$, see Section C.2 of the Supplemetary Material (Zhu et al. (2020)) for more details.

2.1.1. *Distance covariance.* Distance covariance was first introduced by Székely, Rizzo and Bakirov (2007) to measure the dependence between two random vectors of arbitrary dimensions. For two random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, the (squared) distance covariance is defined as

$$\mathrm{dCov}^2(X, Y) = \int_{\mathbb{R}^{p+q}} \frac{|\phi_{X,Y}(t, s) - \phi_X(t)\phi_Y(s)|^2}{c_p c_q |t|^{1+p}|s|^{1+q}}\, dt\, ds,$$

where $c_p = \pi^{(1+p)/2}/\Gamma((1+p)/2)$, $|\cdot|$ is the (complex) Euclidean norm defined as $|x| = \sqrt{\bar{x}^T x}$ for any vector $x$ in the complex vector space ($\bar{x}$ denotes the conjugate of $x$), $\phi_X$ and $\phi_Y$ are the characteristic functions of $X$ and $Y$, respectively, $\phi_{X,Y}$ is the joint characteristic function. According to Theorem 7 of Székely and Rizzo (2009), an alternative definition of distance covariance is given by

$$\begin{aligned}\mathrm{dCov}^2(X, Y) = {} & \mathrm{E}|X - X'||Y - Y'| \\ & + \mathrm{E}|X - X'|E|Y - Y'| - 2\mathrm{E}|X - X'||Y - Y''|,\end{aligned} \tag{4}$$

where $(X', Y')$ and $(X'', Y'')$ are independent copies of $(X, Y)$. It has been shown that $\mathrm{dCov}^2(X, Y) = 0$ if and only if $X$ and $Y$ are independent. Therefore, it is able to measure any type of dependence including nonlinear and nonmonotonic dependence between $X$ and $Y$, whereas the commonly used Pearson correlation can only measure the linear dependence and the rank correlation coefficients (Kendall's $\tau$ and Spearman's $\rho$) can only capture the monotonic dependence.

Notice that in the above setting, $p, q$ are arbitrary positive integers. Therefore, distance covariance is applicable to the high dimensional setting, where we allow $p, q \to \infty$. However, it is unclear whether this metric can still retain the power to detect the nonlinear dependence or not when the dimension is high. Distance correlation (dCor) is the normalized version of distance covariance, which is defined as

$$\mathrm{dCor}^2(X, Y) = \begin{cases} \dfrac{\mathrm{dCov}^2(X, Y)}{\sqrt{\mathrm{dCov}^2(X, X)\mathrm{dCov}^2(Y, Y)}}, & \mathrm{dCov}^2(X, X)\mathrm{dCov}^2(Y, Y) > 0, \\ 0, & \mathrm{dCov}^2(X, X)\mathrm{dCov}^2(Y, Y) = 0. \end{cases}$$

Following Székely and Rizzo (2014), we introduce the $\mathcal{U}$-centering based unbiased sample distance covariance ($\mathrm{dCov}_n^2$) as follows:

$$\mathrm{dCov}_n^2(\mathbf{X}, \mathbf{Y}) = (\widetilde{\mathbf{A}} \cdot \widetilde{\mathbf{B}}),$$

where $\widetilde{\mathbf{A}}$, $\widetilde{\mathbf{B}}$ are the $\mathcal{U}$-centered versions of $\mathbf{A} = (a_{st})_{s,t=1}^n$, $\mathbf{B} = (b_{st})_{s,t=1}^n$, respectively, and $a_{st} = |X_s - X_t|$, $b_{st} = |Y_s - Y_t|$ for $s, t = 1, \dots, n$. Correspondingly, the sample distance correlation ($\mathrm{dCor}_n^2$) is given as

$$\mathrm{dCor}_n^2(\mathbf{X}, \mathbf{Y}) = \begin{cases} \dfrac{\mathrm{dCov}_n^2(\mathbf{X}, \mathbf{Y})}{\sqrt{\mathrm{dCov}_n^2(\mathbf{X}, \mathbf{X})\mathrm{dCov}_n^2(\mathbf{Y}, \mathbf{Y})}}, & \mathrm{dCov}_n^2(\mathbf{X}, \mathbf{X})\mathrm{dCov}_n^2(\mathbf{Y}, \mathbf{Y}) > 0, \\ 0, & \mathrm{dCov}_n^2(\mathbf{X}, \mathbf{X})\mathrm{dCov}_n^2(\mathbf{Y}, \mathbf{Y}) = 0. \end{cases}$$

Here, for $s \neq t$, we can apply the approximation in Proposition 2.1, that is

$$\frac{a_{st}}{\tau_X} = 1 + \frac{1}{2} L_X(X_s, X_t) + R_X(X_s, X_t), \tag{5}$$

$$\frac{b_{st}}{\tau_Y} = 1 + \frac{1}{2} L_Y(Y_s, Y_t) + R_Y(Y_s, Y_t), \tag{6}$$

where

$$L_X(X_s, X_t) = \frac{|X_s - X_t|^2 - \tau_X^2}{\tau_X^2}, \qquad L_Y(Y_s, Y_t) = \frac{|Y_s - Y_t|^2 - \tau_Y^2}{\tau_Y^2},$$

and $R_X$, $R_Y$ are the remainder terms from the approximation. The approximation of the pairwise $L^2$ distance in equations (5) and (6) is our building block to decompose the unbiased sample (squared) distance covariance ($\mathrm{dCov}_n^2$) into a leading term plus a negligible remainder term under the HDLSS setting. The following main theorem summarizes the decomposition properties of sample distance covariance ($\mathrm{dCov}_n^2$).

THEOREM 2.1. *Under Assumption* D1, *we can show that*

(i)

$$\mathrm{dCov}_n^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{\tau} \sum_{i=1}^p \sum_{j=1}^q \mathrm{cov}_n^2(\mathcal{X}_i, \mathcal{Y}_j) + \mathcal{R}_n. \tag{7}$$

*Here*,

$$\mathrm{cov}_n^2(\mathcal{X}_i, \mathcal{Y}_j) = \frac{1}{\binom{n}{4}} \sum_{s<t<u<v} h(x_{si}, x_{ti}, x_{ui}, x_{vi}; y_{sj}, y_{tj}, y_{uj}, y_{vj}),$$

*and the kernel $h$ is defined as*

$$\begin{aligned} &h(x_{si}, x_{ti}, x_{ui}, x_{vi}; y_{sj}, y_{tj}, y_{uj}, y_{vj}) \\ &\quad = \frac{1}{4!} \sum_{*}^{(s,t,u,v)} \frac{1}{4}(x_{si} - x_{ti})(y_{sj} - y_{tj})(x_{ui} - x_{vi})(y_{uj} - y_{vj}), \end{aligned}$$

*where the summation $\sum_*^{(s,t,u,v)}$ is over all permutations of the* 4*-tuples of indices* $(s, t, u, v)$ *and $\mathcal{R}_n$ is the remainder term.* $\mathrm{cov}_n^2(\mathcal{X}_i, \mathcal{Y}_j)$ *is a fourth-order U-statistic and is an unbiased estimator for the squared covariance between $x_i$ and $y_j$, that is,* $\mathrm{E}[\mathrm{cov}_n^2(\mathcal{X}_i, \mathcal{Y}_j)] = \mathrm{cov}^2(x_i, y_j)$.

(ii) *Further, suppose Assumption* D2 *holds. Then*

$$\frac{1}{\tau} \sum_{i=1}^p \sum_{j=1}^q \mathrm{cov}_n^2(\mathcal{X}_i, \mathcal{Y}_j) = O_p(\tau a_p b_q),$$

$$\mathcal{R}_n = O_p(\tau a_p^2 b_q + \tau a_p b_q^2) = o_p(1),$$

*thus the remainder term is of smaller order compared to the leading term and, therefore, is asymptotically negligible.*

Equation (7) in Theorem 2.1 shows that the leading term for sample distance covariance is the sum of all componentwise squared sample cross-covariances scaled by $\tau$, which depends on the marginal variances of $X$ and $Y$. This theorem suggests that in the HDLSS setting, the sample distance covariance can only measure the componentwise linear dependence between the two random vectors.

As argued previously, sample distance covariance ($\mathrm{dCov}_n^2$) based tests suffer from power loss when $X$ and $Y$ are componentwisely nonlinear dependent but uncorrelated. To remedy this drawback, we can consider the following aggregation of marginal sample distance covariances:

$$\mathrm{mdCov}_n^2(\mathbf{X}, \mathbf{Y}) = \sqrt{\binom{n}{2}} \sum_{i=1}^{p} \sum_{j=1}^{q} \mathrm{dCov}_n^2(\mathcal{X}_i, \mathcal{Y}_j),$$

where $\mathrm{dCov}_n^2(\mathcal{X}_i, \mathcal{Y}_j) = (\widetilde{\mathbf{A}}(i) \cdot \widetilde{\mathbf{B}}(j))$, $\widetilde{\mathbf{A}}(i)$ and $\widetilde{\mathbf{B}}(j)$ are the $\mathcal{U}$-centered versions of $\mathbf{A}(i) = (a_{st}(i))_{s,t=1}^n$, $\mathbf{B}(j) = (b_{st}(j))_{s,t=1}^n$, respectively, and $a_{st}(i) = |x_{si} - x_{ti}|$, $b_{st}(j) = |y_{sj} - y_{tj}|$.

Note that $\mathrm{mdCov}_n^2$ captures the pairwise low dimensional nonlinear dependence, which can be viewed as the main effects of the dependence between two high dimensional random vectors. It is natural in many fields of statistics to test for main effects first before proceeding to high order interactions. See Chakraborty and Zhang (2019a) for some discussions on main effects and high order effects in the context of joint dependence testing. In the testing of mutual independence of a high dimensional vector, Yao, Zhang and Shao (2018) also approached the problem by testing the pairwise independence using distance covariance and demonstrated that there may be intrinsic difficulty to capture the effects beyond main effects (pairwise dependence in the mutual independence testing problem), as the tests that target joint dependence do not perform well in the high dimensional setting.

2.1.2. *Hilbert–Schmidt covariance.* A generalization of the Distance Covariance (dCov) is Hilbert–Schmidt Covariance (hCov), first proposed and aka Hilbert–Schmidt independence criterion (HSIC) by Gretton et al. (2008). In particular, the (squared) Hilbert–Schmidt Covariance (hCov) is obtained by kernelizing the Euclidean distance in equation (4), that is,

$$\begin{aligned}\mathrm{hCov}^2(X, Y) = {} & \mathrm{E}\big[K(X, X')L(Y, Y')\big] \\ & + \mathrm{E}\big[K(X, X')\big]\mathrm{E}\big[L(Y, Y')\big] - 2\mathrm{E}\big[K(X, X')L(Y, Y'')\big],\end{aligned}$$

where $(X', Y')$, $(X'', Y'')$ are independent copies of $(X, Y)$ and $K$, $L$ are user specified kernels. Following the literature, we consider the following widely used kernels:

$$\text{Gaussian kernel: } K(x, y) = \exp\left(-\frac{|x - y|^2}{2\gamma^2}\right),$$

$$\text{Laplacian kernel: } K(x, y) = \exp\left(-\frac{|x - y|}{\gamma}\right),$$

where $\gamma$ is a bandwidth parameter. For later convenience, we focus on the kernels that can be represented compactly as $K(x, y) = f(|x - y|/\gamma)$ for some continuously differentiable function $f$. For example, the Gaussian and Laplacian kernel can be defined by choosing different function $f$,

$$\text{Gaussian kernel: } K(x, y) = f\left(\frac{|x - y|}{\gamma}\right), \quad f(a) = \exp\left(-\frac{a^2}{2}\right),$$

$$\text{Laplacian kernel: } K(x, y) = f\left(\frac{|x - y|}{\gamma}\right), \quad f(a) = \exp(-a).$$

In practice, the bandwidth parameter is usually set as the median of pair-wise sample $L^2$ distance. Thus, a natural estimator for $\mathrm{hCov}^2(X, Y)$ is defined as

$$\mathrm{hCov}_n^2(\mathbf{X}, \mathbf{Y}) = (\widetilde{\mathbf{R}} \cdot \widetilde{\mathbf{H}}),$$

where $\widetilde{\mathbf{R}}$ and $\widetilde{\mathbf{H}}$ are the $\mathcal{U}$-centered versions of $\mathbf{R} = (r_{st})_{s,t=1}^n$, $\mathbf{H} = (h_{st})_{s,t=1}^n$, respectively, and $r_{st} = h_{st} = 0$ if $s = t$, otherwise

$$\begin{cases} r_{st} = K(X_s, X_t, \mathbf{X}) = f\left(\dfrac{|X_s - X_t|}{\gamma_{\mathbf{X}}}\right), & \gamma_{\mathbf{X}} = \mathrm{median}\{|X_s - X_t|, s \neq t\}, \\ h_{st} = L(Y_s, Y_t, \mathbf{Y}) = g\left(\dfrac{|Y_s - Y_t|}{\gamma_{\mathbf{Y}}}\right), & \gamma_{\mathbf{Y}} = \mathrm{median}\{|Y_s - Y_t|, s \neq t\}. \end{cases}$$

Similar to the definition of distance correlation, the Hilbert–Schmidt Correlation (hCor) is defined as

$$\mathrm{hCor}^2(X, Y) = \begin{cases} \dfrac{\mathrm{hCov}^2(X, Y)}{\sqrt{\mathrm{hCov}^2(X, X)\mathrm{hCov}^2(Y, Y)}}, & \mathrm{hCov}^2(X, X)\mathrm{hCov}^2(Y, Y) > 0, \\ 0, & \mathrm{hCov}^2(X, X)\mathrm{hCov}^2(Y, Y) = 0, \end{cases}$$

and the sample Hilbert–Schmidt Correlation ($\mathrm{hCor}_n^2$) is defined in the same way by replacing $\mathrm{hCov}^2$ with the corresponding sample version.

Next, we can extend the decomposition results for sample distance covariance ($\mathrm{dCov}_n^2$) to sample Hilbert–Schmidt covariance ($\mathrm{hCov}_n^2$) as shown in the following theorem. Throughout the paper, we use $f^{(1)}$ and $f^{(2)}$ to denote the first and second derivative of the function $f$.

THEOREM 2.2. *Under Assumption D1, we have*

(i)

$$\begin{aligned} &\tau \times \mathrm{hCov}_n^2(\mathbf{X}, \mathbf{Y}) \\ &\quad = f^{(1)}\left(\frac{\tau_X}{\gamma_{\mathbf{X}}}\right) g^{(1)}\left(\frac{\tau_Y}{\gamma_{\mathbf{Y}}}\right) \frac{\tau_X}{\gamma_{\mathbf{X}}} \frac{\tau_Y}{\gamma_{\mathbf{Y}}} \frac{1}{\tau} \sum_{i=1}^{p} \sum_{j=1}^{q} \mathrm{cov}_n^2(\mathcal{X}_i, \mathcal{Y}_j) + \mathcal{R}_n, \end{aligned} \tag{8}$$

*where* $\mathrm{cov}_n^2$ *is defined the same as in Theorem 2.1 and* $\mathcal{R}_n$ *is the remainder term.*

(ii) *Further, suppose Assumption D2 holds. Then*

$$f^{(1)}\left(\frac{\tau_X}{\gamma_{\mathbf{X}}}\right) g^{(1)}\left(\frac{\tau_Y}{\gamma_{\mathbf{Y}}}\right) \frac{\tau_X}{\gamma_{\mathbf{X}}} \frac{\tau_Y}{\gamma_{\mathbf{Y}}} \asymp_p 1,$$

$$\frac{1}{\tau} \sum_{i=1}^{p} \sum_{j=1}^{q} \mathrm{cov}_n^2(\mathcal{X}_i, \mathcal{Y}_j) = O_p(\tau a_p b_q),$$

$$\mathcal{R}_n = O_p(\tau a_p^2 b_q + \tau a_p b_q^2) = o_p(1).$$

*Thus the remainder term is of smaller order compared to the leading term and is therefore asymptotically negligible.*

Notice that different from the decomposition of $\mathrm{dCov}_n^2(\mathbf{X}, \mathbf{Y})$ as in Theorem 2.1, here we decompose $\mathrm{hCov}_n^2$ multiplied by $\tau = \tau_X \tau_Y$. This is expected, since in $\mathrm{hCov}_n^2$, each pairwise distance is normalized by $\gamma_{\mathbf{X}}$ or $\gamma_{\mathbf{Y}}$, which has asymptotically the same magnitude as $\tau_X$, $\tau_Y$, respectively. In the high dimensional case, the expansion (8) suggests that hCov-based tests also suffer from power loss when $X$ and $Y$ are componentwisely uncorrelated but nonlinearly dependent.

To analyze the asymptotic property of sample Hilbert–Schmidt covariance, most literature would assume the bandwidth parameters to be fixed constants; see, for example, Gretton et al. (2008). In contrast, our approach can handle the case where these bandwidth parameters are selected to be the median of pairwise sample distance, which is random and whose magnitude increases with dimension.

Similar to the marginal distance covariance introduced in Section 2.1.1, we can also aggregate the marginal Hilbert–Schmidt Covariance (mhCov), which is defined as

$$\mathrm{mhCov}_n^2(\mathbf{X},\mathbf{Y}) = \sqrt{\binom{n}{2}} \sum_{i=1}^{p}\sum_{j=1}^{q} \mathrm{hCov}_n^2(\mathcal{X}_i, \mathcal{Y}_j),$$

where $\mathrm{hCov}_n^2(\mathcal{X}_i,\mathcal{Y}_j) = (\widetilde{\mathbf{R}}(i)\cdot\widetilde{\mathbf{H}}(j))$, $\widetilde{\mathbf{R}}(i)$ and $\widetilde{\mathbf{H}}(j)$ are $\mathcal{U}$-centered version of $\mathbf{R}(i) = (r_{st}(i))_{s,t=1}^n$, $\mathbf{H}(j) = (h_{st}(j))_{s,t=1}^n$, respectively, and $r_{st}(i) = h_{st}(j) = 0$ if $s = t$, otherwise

$$\begin{cases} r_{st}(i) = K(x_{si}, x_{ti}, \mathcal{X}_i) = f\left(\dfrac{|x_{si} - x_{ti}|}{\gamma_{\mathcal{X}_i}}\right), & \gamma_{\mathcal{X}_i} = \mathrm{median}\{|x_{si} - x_{ti}|, s \neq t\}, \\ h_{st}(j) = L(y_{sj}, y_{tj}, \mathcal{Y}_j) = g\left(\dfrac{|y_{sj} - y_{tj}|}{\gamma_{\mathcal{Y}_j}}\right), & \gamma_{\mathcal{Y}_j} = \mathrm{median}\{|y_{sj} - y_{tj}|, s \neq t\}. \end{cases}$$

REMARK 2.2. By the multilinearity of the operator $(\widetilde{\cdot},\widetilde{\cdot})$, it can be easily seen that $\mathrm{mdCov}_n^2(\mathbf{X},\mathbf{Y})$ is equal to (up to a constant) $\mathrm{dCov}_n^2(\mathbf{X},\mathbf{Y})$ equipped with $L^1$-distance and $\mathrm{mhCov}_n^2(\mathbf{X},\mathbf{Y})$ is equal to (up to a constant) $\mathrm{hCov}_n^2(\mathbf{X},\mathbf{Y})$ equipped with kernels

$$K'(X_s, X_t, \mathbf{X}) = \sum_{i=1}^{p} K(x_{si}, x_{ti}, \mathcal{X}_i) \quad \text{and} \quad L'(Y_s, Y_t, \mathbf{Y}) = \sum_{j=1}^{q} L(y_{sj}, y_{tj}, \mathcal{Y}_j).$$

2.2. *Studentized test statistics.* In this section, we provide studentized version of the statistics introduced in Section 2.1. It is worth mentioning that we provide a unified approach to the asymptotic analysis of studentized dCov, mdCov and further extend them to the analysis of studentized hCov.

2.2.1. *Unified approach.* First, we will present results that will be useful for deriving the studentized version of the interested statistics, that is, distance covariance (dCov), marginal distance covariance (mdCov), Hilbert–Schmidt Covariance (hCov), marginal Hilbert–Schmidt Covariance (mhCov). It can be shown later that many previously mentioned statistics are asymptotically equal to the unified quantity $\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y})$ multiplied by some normalizing factor. Here, $\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y})$ is defined as

$$\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y}) = \frac{1}{\sqrt{pq}} \sum_{i=1}^{p}\sum_{j=1}^{q} (\widetilde{\mathbf{K}}(i)\cdot\widetilde{\mathbf{L}}(j)),$$

where $\widetilde{\mathbf{K}}(i)$ and $\widetilde{\mathbf{L}}(j)$ are the $\mathcal{U}$-centered versions of $\mathbf{K}(i) = (k_{st}(i))_{s,t=1}^n$, $\mathbf{L}(j) = (l_{st}(j))_{s,t=1}^n$, respectively, and $k_{st}(i) = l_{st}(j) = 0$ if $s = t$, otherwise $k_{st}(i)$, $l_{st}(j)$ are the double centered kernel distances, that is, for bivariate kernels $k$ and $l$,

$$\begin{aligned} k_{st}(i) &= k(x_{si}, x_{ti}) - \mathrm{E}[k(x_{si}, x_{ti})|x_{si}] - \mathrm{E}[k(x_{si}, x_{ti})|x_{ti}] + \mathrm{E}[k(x_{si}, x_{ti})], \\ l_{st}(i) &= l(y_{si}, y_{ti}) - \mathrm{E}[l(y_{si}, y_{ti})|y_{si}] - \mathrm{E}[l(y_{si}, y_{ti})|y_{ti}] + \mathrm{E}[l(y_{si}, y_{ti})]. \end{aligned}$$

The advantage of using the double centering kernel distance is that we can have 0 covariance between $k_{st}(i)$ and $k_{uv}(j)$ ($l_{st}(i)$ and $l_{uv}(j)$) for $\{s,t\} \neq \{u,v\}$ as shown in the following proposition.

PROPOSITION 2.2. *For all* $1 \le i, i' \le p$, $1 \le j, j' \le q$, *if* $\{s,t\} \neq \{u,v\}$, *then*

$$\mathrm{E}\big[k_{st}(i)k_{uv}(i')\big] = \mathrm{E}\big[l_{st}(j)l_{uv}(j')\big] = \mathrm{E}\big[k_{st}(i)l_{uv}(j)\big] = 0.$$

To derive the limiting distribution of the unified quantity, we need the following assumptions.

ASSUMPTION D3. For fixed $n$, as $p \wedge q \to \infty$,

$$\begin{pmatrix} p^{-1/2}\sum_{i=1}^{p} k_{st}(i) \\ q^{-1/2}\sum_{j=1}^{q} l_{uv}(j) \end{pmatrix}_{s<t,u<v} \xrightarrow{d} \begin{pmatrix} c_{st} \\ d_{uv} \end{pmatrix}_{s<t,u<v},$$

where $\{c_{st}, d_{uv}\}_{s<t,u<v}$ are jointly Gaussian. Naturally, we further assume the existence of the following constants that show up in the covariance matrix of $\{c_{st}, d_{uv}\}$:

$$\begin{aligned}
\operatorname{var}[c_{st}] &:= \sigma_x^2 = \lim_p \frac{1}{p}\sum_{i,j=1}^{p} \operatorname{cov}\big[k_{st}(i), k_{st}(j)\big] \\
&= \begin{cases} \lim_p \dfrac{\sum_{i,j=1}^{p} \mathrm{dCov}^2(x_i, x_j)}{p}, & \text{if } k(x,y) = l(x,y) = |x-y|, \\ \lim_p \dfrac{\sum_{i,j=1}^{p} 4\operatorname{cov}^2(x_i, x_j)}{p}, & \text{if } k(x,y) = l(x,y) = |x-y|^2, \end{cases} \\
\operatorname{var}[d_{st}] &:= \sigma_y^2 = \lim_q \frac{1}{q}\sum_{i,j=1}^{q} \operatorname{cov}\big[l_{st}(i), l_{st}(j)\big] \\
&= \begin{cases} \lim_q \dfrac{\sum_{i,j=1}^{q} \mathrm{dCov}^2(y_i, y_j)}{q}, & \text{if } k(x,y) = l(x,y) = |x-y|, \\ \lim_q \dfrac{\sum_{i,j=1}^{q} 4\operatorname{cov}^2(y_i, y_j)}{q}, & \text{if } k(x,y) = l(x,y) = |x-y|^2, \end{cases} \\
\operatorname{cov}[c_{st}, d_{st}] &:= \sigma_{xy}^2 = \lim_{p,q} \frac{1}{\sqrt{pq}}\sum_{i=1}^{p}\sum_{j=1}^{q} \operatorname{cov}\big[k_{st}(i), l_{st}(j)\big] \\
&= \begin{cases} \lim_{p,q} \dfrac{\sum_{i=1}^{p}\sum_{j=1}^{q} \mathrm{dCov}^2(x_i, y_j)}{\sqrt{pq}}, & \text{if } k(x,y) = l(x,y) = |x-y|, \\ \lim_{p,q} \dfrac{\sum_{i=1}^{p}\sum_{j=1}^{q} 4\operatorname{cov}^2(x_i, y_j)}{\sqrt{pq}}, & \text{if } k(x,y) = l(x,y) = |x-y|^2. \end{cases}
\end{aligned}$$

REMARK 2.3. Notice that when $\{s,t\} \neq \{u,v\}$, we do not assume the form of $\operatorname{cov}[c_{st}, c_{uv}]$, $\operatorname{cov}[d_{st}, d_{uv}]$, $\operatorname{cov}[c_{st}, d_{uv}]$ in Assumption D3, since it follows easily from Proposition 2.2 that $\operatorname{cov}[c_{st}, c_{uv}] = 0$, $\operatorname{cov}[d_{st}, d_{uv}] = 0$ and $\operatorname{cov}[c_{st}, d_{uv}] = 0$ if $\{s,t\} \neq \{u,v\}$.

REMARK 2.4. The above Central Limit Theorem (CLT) result can be derived under suitable moment and weak dependence assumptions for the components of $X$ and $Y$ and the weak dependence assumptions can be satisfied by a broad range of time series models such as ARMA models. We refer the reader to Doukhan and Neumann (2008) for a relatively recent survey of weak dependence notions and the CLT results under such weak dependence. It is worth noting that the commonly used weak dependence assumptions in time series analysis, such as $\alpha$-mixing, $\beta$-mixing and variants [Bradley (2007)], near epoch dependence [Gallant and White (1988), Davidson (1994)] and physical dependence measure [Wu (2005)], all require the components have a natural time ordering. In our setting, the components do not necessarily have a natural ordering but our results still hold as long as there exists a permutation of components that satisfy the weak dependence assumption. Furthermore, we remark that the weak dependence assumption typically rules out long range dependence and local strong dependence, under which we might have non-Gaussian limit and different norming rate. We shall examine the validity of our tests in these two scenarios via simulations.

The following theorem is our main result, which shows that the unified quantity converges in distribution to a quadratic form of random variables.

THEOREM 2.3. *Fixing $n$ and letting $p \wedge q \to \infty$, under Assumptions D1 and D3,*

$$\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y}) \xrightarrow{d} \frac{1}{v}\mathbf{c}^T\mathbf{M}\mathbf{d},$$

$$\mathrm{uCov}_n^2(\mathbf{X},\mathbf{X}) \xrightarrow{d} \frac{1}{v}\mathbf{c}^T\mathbf{M}\mathbf{c} \stackrel{d}{=} \frac{\sigma_x^2}{v}\chi_v^2,$$

$$\mathrm{uCov}_n^2(\mathbf{Y},\mathbf{Y}) \xrightarrow{d} \frac{1}{v}\mathbf{d}^T\mathbf{M}\mathbf{d} \stackrel{d}{=} \frac{\sigma_y^2}{v}\chi_v^2,$$

*where $v := n(n-3)/2$, $\mathbf{M}$ is a projection matrix of rank $v$ and*

$$\begin{pmatrix}\mathbf{c}\\ \mathbf{d}\end{pmatrix} \stackrel{d}{=} N\left(\mathbf{0}, \begin{pmatrix}\sigma_x^2\mathbf{I}_{n(n-1)/2} & \sigma_{xy}^2\mathbf{I}_{n(n-1)/2}\\ \sigma_{xy}^2\mathbf{I}_{n(n-1)/2} & \sigma_y^2\mathbf{I}_{n(n-1)/2}\end{pmatrix}\right).$$

For the exact form of $\mathbf{M}$, see the proof of Theorem 2.3 in the Supplemetary Material (Zhu et al. (2020)). Next, we define the quantity $T_u$ as

$$T_u = \sqrt{v-1}\frac{\mathrm{uCor}_n^2(\mathbf{X},\mathbf{Y})}{\sqrt{1-(\mathrm{uCor}_n^2(\mathbf{X},\mathbf{Y}))^2}},$$

where

$$\mathrm{uCor}_n^2(\mathbf{X},\mathbf{Y}) = \frac{\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y})}{\sqrt{\mathrm{uCov}_n^2(\mathbf{X},\mathbf{X})\mathrm{uCov}_n^2(\mathbf{Y},\mathbf{Y})}}.$$

We then define the constants $v$ and $\phi$ that appear in the limiting distribution of $T_u$. Set $v = n(n-3)/2$ and $\phi = \sigma_{xy}^2/\sqrt{\sigma_x^2\sigma_y^2}$ such that

$$\phi = \phi_1\mathbb{I}_{\{k(x,y)=l(x,y)=|x-y|\}} + \phi_2\mathbb{I}_{\{k(x,y)=l(x,y)=|x-y|^2\}},$$

where

$$\phi_1 := \lim_{p,q}\frac{\sum_{i=1}^p\sum_{j=1}^q \mathrm{dCov}^2(x_i,y_j)}{\sqrt{\sum_{i,j=1}^p \mathrm{dCov}^2(x_i,x_j)\sum_{i,j=1}^q \mathrm{dCov}^2(y_i,y_j)}},$$

$$\phi_2 := \lim_{p,q}\frac{\sum_{i=1}^p\sum_{j=1}^q \mathrm{cov}^2(x_i,y_j)}{\sqrt{\sum_{i,j=1}^p \mathrm{cov}^2(x_i,x_j)\sum_{i,j=1}^q \mathrm{cov}^2(y_i,y_j)}}.$$

The limiting distribution of $T_u$ is derived under both null ($H_0$) and alternative ($H_A$) hypothesis, that is,

$$\text{null hypothesis}: H_0 = \{(X, Y)|X \perp Y\},$$
$$\text{alternative hypothesis}: H_A = \{(X, Y)|X \not\perp Y\}.$$

In addition, we also consider the local alternative hypothesis $H_{A_l} \subset H_A$, that is,

$$H_{A_l} = \left\{(X, Y)|X \not\perp Y, \phi = \frac{\phi_0}{\sqrt{v}}\right\},$$

where $v = n(n-3)/2$, $\phi_0 = \phi_{0,1}\mathbb{I}_{\{k(x,y)=l(x,y)=|x-y|\}} + \phi_{0,2}\mathbb{I}_{\{k(x,y)=l(x,y)=|x-y|^2\}}$ and $0 < \phi_{0,1}, \phi_{0,2} < \infty$ are constants with respect to $n$. It is also interesting to compare the asymptotic power under the following class of alternatives $H_{A_s} \subset H_A$, that is,

$$H_{A_s} = \{(X, Y)|x_i \not\perp y_j, \operatorname{cov}(x_i, y_j) = 0 \text{ for all } 1 \le i \le p, 1 \le j \le q\}.$$

In summary, the following table illustrates the value of $\phi$ under different cases we are considering:

| $\phi$ | $H_0$ | $H_A$ | $H_{A_l}$ | $H_{A_s}$ |
|---|---|---|---|---|
| $k(x, y) = l(x, y) = \|x - y\|$ | 0 | $\phi_1$ | $\frac{\phi_{0,1}}{\sqrt{v}}$ | $\phi_1$ |
| $k(x, y) = l(x, y) = \|x - y\|^2$ | 0 | $\phi_2$ | $\frac{\phi_{0,2}}{\sqrt{v}}$ | 0 |

Next, denote by $t_a$ the student $t$-distribution with degrees of freedom $a$. Let $t_a^{(\alpha)}$ be the $(1-\alpha)$th percentile of $t_a$ and $t_{a,b}$ be the noncentral $t$-distribution with degrees of freedom $a$ and noncentral parameter $b$. The asymptotic distribution of $T_u$ is stated in the following proposition.

PROPOSITION 2.3. *Fix $n$ and let $p \wedge q \to \infty$. If Assumptions D1 and D3 hold, then for any fixed $t \in \mathbb{R}$,*

$$P_{H_0}(T_u \le t) \to P(t_{v-1} \le t),$$
$$P_{H_A}(T_u \le t) \to \mathrm{E}[P(t_{v-1,W} \le t)],$$

*where $W \sim \sqrt{\frac{\phi^2}{1-\phi^2}\chi_v^2}$ and $\chi_v^2$ is the chi-square distribution with degrees of freedom $v$.*

REMARK 2.5. For the explicit form of $\mathrm{E}[P(t_{v-1,W} \le t)]$, see Lemma 3 in the Supplementary Material (Zhu et al. (2020)).

Below we derive the large sample approximation of the limiting distribution $\mathrm{E}[P(t_{v-1,W} \le t)]$ under the local alternative hypothesis ($H_{A_l}$).

PROPOSITION 2.4. *Under $H_{A_l}$, if we allow $n$ to grow and $t$ is bounded as $n \to \infty$, $\mathrm{E}[P(t_{v-1,W} \le t)]$ can be approximated as*

$$\mathrm{E}_{H_{A_l}}[P(t_{v-1,W} \le t)] = P(t_{v-1,\phi_0} \le t) + O\left(\frac{1}{v}\right),$$

*where $\phi_0 = \phi_{0,1}\mathbb{I}_{\{k(x,y)=l(x,y)=|x-y|\}} + \phi_{0,2}\mathbb{I}_{\{k(x,y)=l(x,y)=|x-y|^2\}}$. In particular, the result still holds if we replace $t$ with $t_{v-1}^{(\alpha)}$.*

2.2.2. *Studentized tests*. For testing the null, permutation test can be used to determine the critical value of the distance covariance (dCov), Hilbert–Schmidt covariance (hCov), marginal distance covariance (mdCov) and marginal Hilbert–Schmidt covariance (mhCov) respectively. If $\mathrm{dCov}_n^2$, $\mathrm{hCov}_n^2$, $\mathrm{mdCov}_n^2$ or $\mathrm{mhCov}_n^2$ is larger than the corresponding critical value, which can be determined by the empirical permutation distribution function, we reject the null. Alternatively, similar to the construction of $T_u$, we transform each of $\mathrm{dCov}_n^2$, $\mathrm{hCov}_n^2$, $\mathrm{mdCov}_n^2$ and $\mathrm{mhCov}_n^2$ into a statistic that has asymptotic $t$-distribution under the null. Thus, instead of using permutation test, which can be quite computationally expensive, we can determine the critical value using this asymptotic $t$-distribution. For each $R \in \{\mathrm{dCov}, \mathrm{hCov}, \mathrm{mdCov}, \mathrm{mhCov}\}$, the studentized test statistic $T_R$ is defined as

$$T_R = \sqrt{v-1}\frac{R^*(\mathbf{X},\mathbf{Y})}{\sqrt{1-(R^*(\mathbf{X},\mathbf{Y}))^2}},$$

where

$$R^*(\mathbf{X},\mathbf{Y}) = \frac{R_n^2(\mathbf{X},\mathbf{Y})}{\sqrt{R_n^2(\mathbf{X},\mathbf{X})R_n^2(\mathbf{Y},\mathbf{Y})}}.$$

The way to derive the asymptotic distribution of $T_R$ is to show that for each $R \in \{\mathrm{dCov}, \mathrm{hCov}, \mathrm{mdCov}\}$, $R_n^2(\mathbf{X},\mathbf{Y})$ and $\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y})$ are asymptotically equal up to an asymptotically constant factor, as shown below.

PROPOSITION 2.5. *Under Assumption D1*:

(i) *When* $k(x,y) = l(x,y) = |x-y|^2$,

$$\mathrm{dCov}_n^2(\mathbf{X},\mathbf{Y}) = \frac{1}{4}\frac{\sqrt{pq}}{\tau}\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y}) + \mathcal{R}_n',$$
$$\tau \times \mathrm{hCov}_n^2(\mathbf{X},\mathbf{Y}) = \frac{\sqrt{pq}}{4\gamma_{\mathbf{X}}\gamma_{\mathbf{Y}}} f^{(1)}\left(\frac{\tau_X}{\gamma_{\mathbf{X}}}\right) g^{(1)}\left(\frac{\tau_Y}{\gamma_{\mathbf{Y}}}\right)\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y}) + \mathcal{R}_n'',$$

*where* $\mathcal{R}_n'$, $\mathcal{R}_n''$ *are the remainder terms*. *Further, suppose Assumption* D2 *holds*. *Then*

$$\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y}) = O_p(\tau a_p b_q),$$
$$\mathcal{R}_n' = O_p(\tau a_p^2 b_q + \tau a_p b_q^2) = o_p(1),$$
$$\mathcal{R}_n'' = O_p(\tau a_p^2 b_q + \tau a_p b_q^2) = o_p(1).$$

*Thus the remainder term is of smaller order compared to the leading term and*, *therefore*, *is asymptotically negligible*.

(ii) *When* $k(x,y) = l(x,y) = |x-y|$,

$$\mathrm{mdCov}_n^2(\mathbf{X},\mathbf{Y}) = \sqrt{pq}\sqrt{\binom{n}{2}}\mathrm{uCov}_n^2(\mathbf{X},\mathbf{Y}).$$

As shown in Proposition 2.5, $k(x,y) = l(x,y) = |x-y|$ would correspond to the mdCov-based $t$-test and $k(x,y) = l(x,y) = |x-y|^2$ would correspond to the $\{\mathrm{dCov}, \mathrm{hCov}\}$-based $t$-tests. Then, for each $R \in \{\mathrm{dCov}, \mathrm{hCov}, \mathrm{mdCov}\}$ the asymptotic distribution of $T_R$ is given in the following corollary.

COROLLARY 2.1. *If Assumptions D1, D2 and D3 hold, for any fixed $t$ and each $R \in$* {dCov, hCov, mdCov}, *we have*

$$P_{H_0}(T_R \leq t) \to P(t_{v-1} \leq t),$$

$$P_{H_A}(T_R \leq t) \to \mathrm{E}\big[P(t_{v-1,W} \leq t)\big], \quad \textit{where } W \sim \sqrt{\frac{\phi^2}{1-\phi^2}\chi_v^2}.$$

After knowing the asymptotic distribution of $T_R$ under the null, that is, $t$-distribution with degrees of freedom $v-1$, we can set critical value as $t_{v-1}^{(\alpha)}$. Then, from Proposition 2.3, under the alternative, the asymptotic power of testing the null can be written as a function of $\phi$, that is,

$$Power_n(\phi) := \mathrm{E}\big[P\big(t_{v-1,W} > t_{v-1}^{\alpha}\big)\big],$$

and under $H_{A_l}$, if we allow $n$ to grow

$$Power_\infty(\phi_0) := \lim_{n\to\infty} Power_n\left(\frac{\phi_0}{\sqrt{v}}\right) = \lim_{n\to\infty} P\big(t_{v-1,\phi_0} > t_{v-1}^{(\alpha)}\big).$$

Next, we can actually bound the ratio of $\phi_1$ and $\phi_2$ for standard normal random variables.

PROPOSITION 2.6. *Suppose that*

$$\begin{pmatrix} X \\ Y \end{pmatrix} \overset{d}{=} N\left(\mathbf{0}, \begin{pmatrix} \mathbf{I}_p & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{XY}^T & \mathbf{I}_q \end{pmatrix}\right),$$

*where $\mathbf{\Sigma}_{XY} = \mathrm{cov}(X, Y)$. We have*

$$0.89^2\phi_2 \leq \phi_1 \leq \phi_2.$$

It will be shown later that $\phi_1$ corresponds to the mdCov-based test, whereas $\phi_2$ corresponds to the dCov and hCov-based tests. Thus considering models described in Proposition 2.6, we expect a power loss for the mdCov-based test comparing to the dCov and hCov-based tests. On the other hand, since $\phi_1$ is bounded below by $0.89^2\phi_2$, the power loss is expected to be moderate.

Using Corollary 2.1, we can theoretically compare the power of these $t$-tests under different cases and the results are summarized in the following table:

| *Power* | $T_{\mathrm{mdCov}}$ | $T_{\mathrm{dCov}}, T_{\mathrm{hCov}}$ |
|---|---|---|
| under $H_A$ | $Power_n(\phi_1)$ | $Power_n(\phi_2)$ |
| under $H_{A_l}$, allow $n$ growing to infinity | $Power_\infty(\phi_{0,1})$ | $Power_\infty(\phi_{0,2})$ |
| under $H_{A_s}$ | $Power_n(\phi_1)$ | $\alpha$ |

For the studentized version of mhCov, if we consider the bandwidth parameters to be fixed constants, then we can use the unified approach to get the limiting $t$-distribution of the transformed $\mathrm{mhCov}_n^2$. On the other hand, if $\gamma_{\mathcal{X}_i}$ and $\gamma_{\mathcal{Y}_j}$ are treated to be median of sample distance along each dimension and are thus random, we encounter technical difficulties to derive the limiting distribution, as in this case the kernelized pairwise distance along each dimension are correlated with each other. This is due to the choice of the bandwidth parameter and the high dimensional approximation used for $\mathrm{hCov}_n^2$ can not be directly applied, since $\gamma_{\mathcal{X}_i}$ and $\gamma_{\mathcal{Y}_j}$ are calculated componentwisely. Nevertheless, we shall examine the testing efficiency using $t$-distribution approximation when the bandwidth parameters are chosen to be the median of sample distance in simulation.

REMARK 2.6. An anonymous referee inquired about the applicability of our tests to the setting when the $p$-dimensional data vector $X = (x_1, \dots, x_p)^T$ is a growth curve, and thus can be viewed as a stochastic process or random function evaluated at time points $\{t_i\}_{i=1}^p$, say $0 \le t_1 < t_2 < \cdots < t_p \le 1$. Under suitable conditions, one can show that the Euclidean norm of $X$ after proper scaling converges to the $L_2$ norm of the random function when the number of sampling points goes to infinity. It is known that the Hilbert space $L_2([0, 1])$ is of strong negative type (Lyons (2013)), and thus the HSIC or the distance covariance based on the $L_2$ norm completely characterizes dependence. Therefore, the Euclidean norm is a proper norm to use if $X$ is considered to be an element in $L_2([0, 1])$ and we want to use the $L_2$ norm to construct our distance metrics. However, the setting we are considering in this paper assumed the components of $X$ and $Y$ have weak dependence and the above growth curve example falls into the very strong dependent case, and thus our theoretical phenomenon does not apply. In practice, both strongly componentwise correlated high dimensional data and weakly componentwise dependent high dimensional data can be collected depending on the nature of data generating process. We shall illustrate the usefulness of our theory and proposed tests using an earthquake dataset in Section 3.

Our theory demonstrates the limitation of dCov and hCov in the high dimensional environment, which is intimately related to the use of Euclidean norm in their definitions. Similar phenomenon has been discovered for energy distance [Szekely and Rizzo (2004)] and maximum mean discrepancy [Gretton et al. (2012)] in the two sample testing problem recently; see Zhu and Shao (2019) and Chakraborty and Zhang (2019b). It is natural to ask what norm would be desirable to use in the high dimensional setting and in what sense? We shall leave these questions for future study.

**3. Numerical results.** Here, we consider some numerical examples to compare the "joint" tests, where the distance/Hilbert–Schmidt covariance is applied to whole components of data jointly, with the "marginal" tests, where distance/Hilbert–Schmidt covariance is applied to one-dimensional components and then being aggregated. To this end, we consider the following statistics:

$$\text{“Joint”}\begin{cases} \text{dCov : distance covariance (permutation)} \\ T_{\text{dCov}} : \text{studentized distance covariance} \\ \text{hCov : Hilbert–Schmidt covariance (permutation)} \\ T_{\text{hCov}} : \text{studentized Hilbert–Schmidt covariance} \end{cases}$$

$$\text{“Marginal”}\begin{cases} \text{mdCov : marginal distance covariance (permutation)} \\ T_{\text{mdCov}} : \text{studentized marginal distance covariance,} \\ \text{mhCov : marginal Hilbert–Schmidt covariance (permutation)} \\ T_{\text{mhCov}} : \text{studentized marginal Hilbert–Schmidt covariance} \end{cases}$$

In the above display, $\text{dCov}_n^2$ and $\text{hCov}_n^2$ are the two "joint" test statistics to measure the overall dependence between $X$ and $Y$, $\text{mdCov}_n^2$ and $\text{mhCov}_n^2$ are the "marginal" test statistics, and these four test statistics are implemented as permutation tests; $T_{\text{dCov}}$ from Székely and Rizzo (2013) is the studentized version of dCov, our proposed $t$-tests $T_{\text{hCov}}$, $T_{\text{mdCov}}$, $T_{\text{mhCov}}$ are the studentized versions of hCov, mdCov, mhCov, respectively. All of these four tests are implemented using the $t$-distribution based critical value. We examine both the Gaussian kernel and Laplacian kernel for the Hilbert–Schmidt covariance based tests.

For the permutation-based tests, we randomly shuffle the samples $\{X_1, \dots, X_n\}$ and get $(X_{\pi(1)}, \dots, X_{\pi(n)})$, where $\pi$ is the permutation map from $\{1, \dots, n\}$ to $\{1, \dots, n\}$. Then we calculate the test statistic based on the permuted sample $\{(X_{\pi(1)}, \dots, X_{\pi(n)}), (Y_1, \dots, Y_n)\}$.

The $p$-value for permutation-based test is defined as the proportion of times that the test statistic based on the permuted samples is greater than the one based on the original sample. All the numerical results from permutation-based tests are based on 200 permutations and the empirical rejection rate of the tests are based on 5000 Monte Carlo repetitions.

We first examine the size of the aforementioned tests.

EXAMPLE 3.1. Generate i.i.d. samples from the following models for $i = 1, \dots, n$:

(i) $X_i = (x_{i1}, \dots, x_{ip})^T \sim N(\mathbf{0}, \mathbf{I}_p)$, $Y_i = (y_{i1}, \dots, y_{ip})^T \sim N(\mathbf{0}, \mathbf{I}_p)$.

(ii) Let $AR(1)$ denotes the Gaussian autoregressive model of order 1 with parameter $\phi$, $X_i \sim AR(1)$, $\phi = 0.5$, $Y_i \sim AR(1)$, $\phi = -0.5$.

(iii) Let $\boldsymbol{\Sigma} = (\sigma_{ij}) \in \mathbb{R}^{p\times p}$ and $\sigma_{ij} = 0.7^{|i-j|}$, $X_i = (x_{i1}, \dots, x_{ip})^T \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, $Y_i = (y_{i1}, \dots, y_{ip})^T \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

From Table 1, we can see that all the tests have quite accurate size. Although the $t$-tests are derived under the high dimensional scenario, they still have pretty accurate size even for relatively low dimension (e.g., $p = 5$).

As demonstrated in Theorem 2.1 and 2.2, the leading term in (7) and (8) can only measure the linear dependence as $p \wedge q \to \infty$, therefore, we expect the "joint" test based on $\mathrm{dCov}_n^2(\mathbf{X}, \mathbf{Y})$ or $\mathrm{hCov}_n^2(\mathbf{X}, \mathbf{Y})$ may fail to capture the nonlinear dependence in high dimension. On the other hand, we consider the "marginal" test where we take the sum of pairwise sample distance/Hilbert–Schmidt covariances to measure the low dimensional dependence for all the pairs as the test proposed in Sections 2.1.1 and 2.1.2. The "marginal" test statistic measures the dependence marginally in a low-dimensional fashion so that it can preserve the ability to capture component-wise nonlinear dependence. In the following two examples, we demonstrate the superiority of "marginal" tests.

EXAMPLE 3.2. Generate i.i.d. samples from the following models for $i = 1, \dots, n$:

(i) $X_i = (x_{i1}, \dots, x_{ip})^T \sim N(\mathbf{0}, \mathbf{I}_p)$, $Y_i = (y_{i1}, \dots, y_{ip})^T$, where $y_{ij} = x_{ij}^2$ for $j = 1, \dots, p$.

(ii) Let $\boldsymbol{\Sigma} = (\sigma_{ij}) \in \mathbb{R}^{p\times p}$ and $\sigma_{ij} = 0.7^{|i-j|}$, $X_i = (x_{i1}, \dots, x_{ip})^T \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, $Y_i = (y_{i1}, \dots, y_{ip})^T$, where $y_{ij} = x_{ij}^2$ for $j = 1, \dots, p$.

(iii) $X_i = (x_{i1}, \dots, x_{ip})^T \sim N(\mathbf{0}, \mathbf{I}_p)$, $Y_i = (y_{i1}, \dots, y_{ip})^T$, where $y_{ij} = \log|x_{ij}|$ for $j = 1, \dots, p$.

EXAMPLE 3.3. Generate i.i.d. samples from the following models for $i = 1, \dots, n$:

(i) Let $\circ$ denotes the Hadamard product, $X_i = (x_{i1}, \dots, x_{ip})^T \overset{\text{i.i.d.}}{\sim} U(-1, 1)$, $Y_i = X_i \circ X_i$.

(ii) $X_i = (x_{i1}, \dots, x_{ip})^T \overset{\text{i.i.d.}}{\sim} U(0, 1)$, $Y_i = 4X_i \circ X_i \circ X_i - 3.6X_i + 0.8$.

(iii) $Z_i = (z_{i1}, \dots, z_{ip})^T \overset{\text{i.i.d.}}{\sim} U(0, 2\pi)$, $X_i = \sin(Z_i)$, $Y_i = \cos(Z_i)$.

Notice that in the above two examples, $\mathrm{cov}^2(x_i, y_j) = 0$ but $\mathrm{dCov}^2(x_i, y_j) \neq 0$ for all $(i, j)$s, that is, $(X, Y) \in H_{A_s}$. From Table 2, we can observe that for Example 3.2, the "joint" tests suffer substantial power loss as dimension increases for fixed sample size. The power loss is less severe in case (ii) than the ones in cases (i) and (iii), due to the dependence between the components. On the other hand, the powers corresponding to the marginal test statistics consistently outperform their joint counterparts with very little to none power reduction as the dimension increases. Similar phenomenon can be observed for Example 3.3; see Table 3. In addition, for all the cases in both Example 3.2 and Example 3.3, the power

TABLE 1
*Size comparison from Example 3.1*

| | | | | | | | | Gaussian Kernel | | | | Laplacian Kernel | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $p$ | $\alpha$ | dCov | mdCov | $T_{\mathrm{dCov}}$ | $T_{\mathrm{mdCov}}$ | hCov | mhCov | $T_{\mathrm{hCov}}$ | $T_{\mathrm{mhCov}}$ | hCov | mhCov | $T_{\mathrm{hCov}}$ | $T_{\mathrm{mhCov}}$ |
| (i) | 10 | 5 | 0.010 | 0.017 | 0.014 | 0.020 | 0.014 | 0.016 | 0.015 | 0.020 | 0.014 | 0.014 | 0.014 | 0.017 | 0.013 |
| | 10 | 5 | 0.050 | 0.055 | 0.055 | 0.062 | 0.061 | 0.055 | 0.060 | 0.062 | 0.061 | 0.055 | 0.050 | 0.064 | 0.050 |
| | 10 | 5 | 0.100 | 0.105 | 0.107 | 0.110 | 0.110 | 0.103 | 0.106 | 0.109 | 0.109 | 0.102 | 0.099 | 0.105 | 0.101 |
| | 10 | 30 | 0.010 | 0.015 | 0.015 | 0.013 | 0.011 | 0.015 | 0.016 | 0.012 | 0.012 | 0.014 | 0.014 | 0.011 | 0.011 |
| | 10 | 30 | 0.050 | 0.054 | 0.053 | 0.050 | 0.053 | 0.053 | 0.054 | 0.050 | 0.052 | 0.052 | 0.059 | 0.050 | 0.054 |
| | 10 | 30 | 0.100 | 0.102 | 0.104 | 0.099 | 0.102 | 0.102 | 0.105 | 0.100 | 0.103 | 0.102 | 0.107 | 0.101 | 0.105 |
| | 30 | 5 | 0.010 | 0.014 | 0.016 | 0.019 | 0.018 | 0.016 | 0.016 | 0.020 | 0.017 | 0.016 | 0.015 | 0.019 | 0.015 |
| | 30 | 5 | 0.050 | 0.052 | 0.053 | 0.062 | 0.059 | 0.052 | 0.057 | 0.061 | 0.059 | 0.054 | 0.055 | 0.061 | 0.058 |
| | 30 | 5 | 0.100 | 0.105 | 0.104 | 0.105 | 0.107 | 0.103 | 0.107 | 0.106 | 0.106 | 0.105 | 0.104 | 0.109 | 0.104 |
| | 30 | 30 | 0.010 | 0.014 | 0.014 | 0.011 | 0.012 | 0.014 | 0.017 | 0.010 | 0.013 | 0.014 | 0.017 | 0.011 | 0.013 |
| | 30 | 30 | 0.050 | 0.051 | 0.053 | 0.052 | 0.051 | 0.051 | 0.056 | 0.052 | 0.053 | 0.051 | 0.058 | 0.051 | 0.052 |
| | 30 | 30 | 0.100 | 0.097 | 0.105 | 0.096 | 0.103 | 0.097 | 0.105 | 0.095 | 0.101 | 0.099 | 0.104 | 0.100 | 0.102 |
| | 60 | 5 | 0.010 | 0.013 | 0.015 | 0.018 | 0.016 | 0.014 | 0.013 | 0.019 | 0.016 | 0.014 | 0.015 | 0.017 | 0.015 |
| | 60 | 5 | 0.050 | 0.052 | 0.055 | 0.061 | 0.057 | 0.054 | 0.061 | 0.060 | 0.064 | 0.053 | 0.057 | 0.058 | 0.058 |
| | 60 | 5 | 0.100 | 0.103 | 0.104 | 0.109 | 0.104 | 0.107 | 0.108 | 0.110 | 0.110 | 0.102 | 0.101 | 0.103 | 0.102 |
| | 60 | 30 | 0.010 | 0.019 | 0.017 | 0.016 | 0.012 | 0.019 | 0.015 | 0.015 | 0.013 | 0.020 | 0.016 | 0.015 | 0.014 |
| | 60 | 30 | 0.050 | 0.060 | 0.063 | 0.057 | 0.058 | 0.060 | 0.058 | 0.057 | 0.058 | 0.061 | 0.058 | 0.058 | 0.055 |
| | 60 | 30 | 0.100 | 0.113 | 0.112 | 0.110 | 0.107 | 0.113 | 0.109 | 0.111 | 0.105 | 0.110 | 0.111 | 0.107 | 0.107 |
| (ii) | 10 | 5 | 0.010 | 0.015 | 0.015 | 0.023 | 0.023 | 0.014 | 0.016 | 0.023 | 0.019 | 0.015 | 0.017 | 0.022 | 0.021 |
| | 10 | 5 | 0.050 | 0.051 | 0.054 | 0.064 | 0.066 | 0.053 | 0.058 | 0.064 | 0.066 | 0.054 | 0.058 | 0.066 | 0.062 |
| | 10 | 5 | 0.100 | 0.101 | 0.105 | 0.107 | 0.111 | 0.100 | 0.109 | 0.105 | 0.113 | 0.102 | 0.110 | 0.106 | 0.109 |
| | 10 | 30 | 0.010 | 0.014 | 0.018 | 0.013 | 0.016 | 0.014 | 0.017 | 0.014 | 0.013 | 0.017 | 0.018 | 0.017 | 0.013 |
| | 10 | 30 | 0.050 | 0.060 | 0.061 | 0.061 | 0.061 | 0.061 | 0.056 | 0.062 | 0.056 | 0.059 | 0.060 | 0.059 | 0.056 |
| | 10 | 30 | 0.100 | 0.105 | 0.105 | 0.110 | 0.107 | 0.105 | 0.105 | 0.109 | 0.099 | 0.106 | 0.108 | 0.111 | 0.104 |
| | 30 | 5 | 0.010 | 0.012 | 0.011 | 0.022 | 0.023 | 0.012 | 0.014 | 0.021 | 0.020 | 0.013 | 0.013 | 0.019 | 0.016 |
| | 30 | 5 | 0.050 | 0.046 | 0.048 | 0.055 | 0.056 | 0.046 | 0.052 | 0.055 | 0.059 | 0.047 | 0.053 | 0.051 | 0.059 |
| | 30 | 5 | 0.100 | 0.094 | 0.096 | 0.094 | 0.096 | 0.096 | 0.100 | 0.097 | 0.100 | 0.093 | 0.107 | 0.097 | 0.104 |

TABLE 1
(*Continued*)

| | | | | | | | | Gaussian Kernel | | | | Laplacian Kernel | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $p$ | $\alpha$ | dCov | mdCov | $T_{\mathrm{dCov}}$ | $T_{\mathrm{mdCov}}$ | hCov | mhCov | $T_{\mathrm{hCov}}$ | $T_{\mathrm{mhCov}}$ | hCov | mhCov | $T_{\mathrm{hCov}}$ | $T_{\mathrm{mhCov}}$ |
| | 30 | 30 | 0.010 | 0.016 | 0.016 | 0.017 | 0.015 | 0.017 | 0.015 | 0.017 | 0.011 | 0.017 | 0.015 | 0.017 | 0.012 |
| | 30 | 30 | 0.050 | 0.061 | 0.058 | 0.060 | 0.059 | 0.061 | 0.055 | 0.060 | 0.054 | 0.058 | 0.052 | 0.060 | 0.051 |
| | 30 | 30 | 0.100 | 0.109 | 0.105 | 0.110 | 0.107 | 0.111 | 0.101 | 0.110 | 0.098 | 0.111 | 0.102 | 0.113 | 0.097 |
| | 60 | 5 | 0.010 | 0.015 | 0.013 | 0.026 | 0.022 | 0.016 | 0.014 | 0.024 | 0.020 | 0.013 | 0.015 | 0.020 | 0.018 |
| | 60 | 5 | 0.050 | 0.055 | 0.052 | 0.062 | 0.061 | 0.055 | 0.053 | 0.061 | 0.059 | 0.055 | 0.052 | 0.061 | 0.054 |
| | 60 | 5 | 0.100 | 0.101 | 0.100 | 0.103 | 0.100 | 0.102 | 0.100 | 0.104 | 0.099 | 0.101 | 0.097 | 0.103 | 0.099 |
| | 60 | 30 | 0.010 | 0.013 | 0.014 | 0.013 | 0.014 | 0.013 | 0.016 | 0.014 | 0.013 | 0.014 | 0.015 | 0.013 | 0.012 |
| | 60 | 30 | 0.050 | 0.055 | 0.051 | 0.058 | 0.051 | 0.054 | 0.054 | 0.057 | 0.053 | 0.058 | 0.053 | 0.053 | 0.052 |
| | 60 | 30 | 0.100 | 0.105 | 0.102 | 0.105 | 0.100 | 0.106 | 0.103 | 0.105 | 0.102 | 0.107 | 0.105 | 0.107 | 0.104 |
| (iii) | 10 | 5 | 0.010 | 0.012 | 0.013 | 0.025 | 0.024 | 0.012 | 0.014 | 0.024 | 0.022 | 0.016 | 0.013 | 0.025 | 0.019 |
| | 10 | 5 | 0.050 | 0.051 | 0.051 | 0.068 | 0.069 | 0.053 | 0.051 | 0.068 | 0.062 | 0.053 | 0.049 | 0.067 | 0.056 |
| | 10 | 5 | 0.100 | 0.100 | 0.099 | 0.107 | 0.103 | 0.100 | 0.098 | 0.105 | 0.102 | 0.100 | 0.098 | 0.104 | 0.101 |
| | 10 | 30 | 0.010 | 0.014 | 0.015 | 0.016 | 0.014 | 0.014 | 0.015 | 0.016 | 0.013 | 0.015 | 0.015 | 0.017 | 0.013 |
| | 10 | 30 | 0.050 | 0.055 | 0.057 | 0.061 | 0.058 | 0.053 | 0.056 | 0.061 | 0.056 | 0.057 | 0.057 | 0.064 | 0.059 |
| | 10 | 30 | 0.100 | 0.104 | 0.105 | 0.105 | 0.107 | 0.103 | 0.105 | 0.104 | 0.107 | 0.106 | 0.110 | 0.106 | 0.112 |
| | 30 | 5 | 0.010 | 0.015 | 0.014 | 0.028 | 0.029 | 0.015 | 0.014 | 0.025 | 0.024 | 0.014 | 0.014 | 0.024 | 0.019 |
| | 30 | 5 | 0.050 | 0.052 | 0.054 | 0.060 | 0.062 | 0.051 | 0.052 | 0.062 | 0.062 | 0.048 | 0.052 | 0.058 | 0.059 |
| | 30 | 5 | 0.100 | 0.103 | 0.103 | 0.098 | 0.099 | 0.101 | 0.101 | 0.101 | 0.098 | 0.099 | 0.099 | 0.097 | 0.098 |
| | 30 | 30 | 0.010 | 0.017 | 0.015 | 0.019 | 0.017 | 0.016 | 0.015 | 0.019 | 0.015 | 0.013 | 0.016 | 0.018 | 0.012 |
| | 30 | 30 | 0.050 | 0.054 | 0.055 | 0.058 | 0.058 | 0.055 | 0.055 | 0.059 | 0.057 | 0.056 | 0.057 | 0.063 | 0.056 |
| | 30 | 30 | 0.100 | 0.102 | 0.105 | 0.105 | 0.103 | 0.101 | 0.099 | 0.103 | 0.102 | 0.104 | 0.107 | 0.105 | 0.105 |
| | 60 | 5 | 0.010 | 0.012 | 0.012 | 0.029 | 0.027 | 0.014 | 0.012 | 0.028 | 0.024 | 0.016 | 0.011 | 0.023 | 0.021 |
| | 60 | 5 | 0.050 | 0.052 | 0.052 | 0.063 | 0.064 | 0.050 | 0.048 | 0.063 | 0.059 | 0.050 | 0.052 | 0.059 | 0.061 |
| | 60 | 5 | 0.100 | 0.100 | 0.101 | 0.098 | 0.095 | 0.098 | 0.099 | 0.097 | 0.099 | 0.099 | 0.098 | 0.100 | 0.094 |
| | 60 | 30 | 0.010 | 0.017 | 0.015 | 0.020 | 0.019 | 0.016 | 0.017 | 0.020 | 0.017 | 0.016 | 0.015 | 0.019 | 0.014 |
| | 60 | 30 | 0.050 | 0.052 | 0.053 | 0.058 | 0.060 | 0.055 | 0.057 | 0.061 | 0.059 | 0.057 | 0.056 | 0.062 | 0.059 |
| | 60 | 30 | 0.100 | 0.103 | 0.106 | 0.107 | 0.103 | 0.102 | 0.106 | 0.107 | 0.105 | 0.103 | 0.102 | 0.106 | 0.101 |

TABLE 2
*Power comparison under $H_{A_s}$ from Example 3.2*

| | | | | | | | | Gaussian Kernel | | | | Laplacian Kernel | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $p$ | $\alpha$ | dCov | mdCov | $T_{\mathrm{dCov}}$ | $T_{\mathrm{mdCov}}$ | hCov | mhCov | $T_{\mathrm{hCov}}$ | $T_{\mathrm{mhCov}}$ | hCov | mhCov | $T_{\mathrm{hCov}}$ | $T_{\mathrm{mhCov}}$ |
| (i) | 10 | 5 | 0.010 | 0.113 | 0.285 | 0.144 | 0.321 | 0.110 | 0.493 | 0.138 | 0.516 | 0.172 | 0.801 | 0.226 | 0.813 |
| | 10 | 5 | 0.050 | 0.231 | 0.495 | 0.254 | 0.519 | 0.236 | 0.724 | 0.256 | 0.736 | 0.356 | 0.927 | 0.398 | 0.938 |
| | 10 | 5 | 0.100 | 0.325 | 0.618 | 0.332 | 0.628 | 0.325 | 0.828 | 0.336 | 0.834 | 0.495 | 0.968 | 0.506 | 0.969 |
| | 10 | 30 | 0.010 | 0.032 | 0.286 | 0.028 | 0.267 | 0.032 | 0.543 | 0.030 | 0.513 | 0.044 | 0.848 | 0.041 | 0.838 |
| | 10 | 30 | 0.050 | 0.101 | 0.526 | 0.098 | 0.523 | 0.098 | 0.769 | 0.099 | 0.763 | 0.124 | 0.945 | 0.128 | 0.947 |
| | 10 | 30 | 0.100 | 0.158 | 0.669 | 0.162 | 0.666 | 0.160 | 0.858 | 0.160 | 0.858 | 0.203 | 0.978 | 0.205 | 0.977 |
| | 30 | 5 | 0.010 | 0.440 | 0.997 | 0.499 | 0.999 | 0.518 | 1 | 0.583 | 1 | 0.924 | 1 | 0.956 | 1 |
| | 30 | 5 | 0.050 | 0.651 | 1.000 | 0.679 | 1.000 | 0.741 | 1 | 0.768 | 1 | 0.987 | 1 | 0.988 | 1 |
| | 30 | 5 | 0.100 | 0.766 | 1.000 | 0.773 | 1 | 0.836 | 1 | 0.845 | 1 | 0.994 | 1 | 0.995 | 1 |
| | 30 | 30 | 0.010 | 0.084 | 1.000 | 0.082 | 1.000 | 0.085 | 1 | 0.082 | 1 | 0.194 | 1 | 0.192 | 1 |
| | 30 | 30 | 0.050 | 0.190 | 1 | 0.187 | 1 | 0.192 | 1 | 0.192 | 1 | 0.365 | 1 | 0.365 | 1 |
| | 30 | 30 | 0.100 | 0.275 | 1 | 0.272 | 1 | 0.280 | 1 | 0.276 | 1 | 0.476 | 1 | 0.478 | 1 |
| | 60 | 5 | 0.010 | 0.948 | 1 | 0.976 | 1 | 0.983 | 1 | 0.992 | 1 | 1 | 1 | 1 | 1 |
| | 60 | 5 | 0.050 | 0.994 | 1 | 0.996 | 1 | 0.998 | 1 | 0.999 | 1 | 1 | 1 | 1 | 1 |
| | 60 | 5 | 0.100 | 0.999 | 1 | 0.999 | 1 | 1.000 | 1 | 1.000 | 1 | 1 | 1 | 1 | 1 |
| | 60 | 30 | 0.010 | 0.185 | 1 | 0.173 | 1 | 0.194 | 1 | 0.183 | 1 | 0.587 | 1 | 0.587 | 1 |
| | 60 | 30 | 0.050 | 0.346 | 1 | 0.346 | 1 | 0.361 | 1 | 0.360 | 1 | 0.779 | 1 | 0.782 | 1 |
| | 60 | 30 | 0.100 | 0.462 | 1 | 0.459 | 1 | 0.475 | 1 | 0.473 | 1 | 0.861 | 1 | 0.864 | 1 |
| (ii) | 10 | 5 | 0.010 | 0.167 | 0.232 | 0.237 | 0.296 | 0.192 | 0.347 | 0.263 | 0.410 | 0.279 | 0.595 | 0.391 | 0.652 |
| | 10 | 5 | 0.050 | 0.306 | 0.386 | 0.341 | 0.421 | 0.356 | 0.570 | 0.401 | 0.606 | 0.525 | 0.806 | 0.584 | 0.832 |
| | 10 | 5 | 0.100 | 0.401 | 0.489 | 0.409 | 0.500 | 0.479 | 0.699 | 0.487 | 0.709 | 0.674 | 0.892 | 0.689 | 0.901 |
| | 10 | 30 | 0.010 | 0.080 | 0.202 | 0.091 | 0.210 | 0.082 | 0.376 | 0.091 | 0.366 | 0.099 | 0.646 | 0.123 | 0.634 |
| | 10 | 30 | 0.050 | 0.178 | 0.369 | 0.191 | 0.378 | 0.179 | 0.605 | 0.192 | 0.610 | 0.229 | 0.834 | 0.252 | 0.837 |
| | 10 | 30 | 0.100 | 0.257 | 0.492 | 0.259 | 0.492 | 0.264 | 0.728 | 0.265 | 0.730 | 0.342 | 0.906 | 0.351 | 0.909 |
| | 30 | 5 | 0.010 | 0.623 | 0.847 | 0.781 | 0.950 | 0.895 | 0.999 | 0.957 | 1 | 0.995 | 1 | 0.999 | 1 |
| | 30 | 5 | 0.050 | 0.872 | 0.984 | 0.902 | 0.990 | 0.982 | 1 | 0.990 | 1 | 1.000 | 1 | 1 | 1 |
| | 30 | 5 | 0.100 | 0.940 | 0.996 | 0.945 | 0.995 | 0.994 | 1 | 0.994 | 1 | 1 | 1 | 1 | 1 |

TABLE 2
(*Continued*)

| | $n$ | $p$ | $\alpha$ | dCov | mdCov | $T_{\mathrm{dCov}}$ | $T_{\mathrm{mdCov}}$ | Gaussian Kernel hCov | Gaussian Kernel mhCov | Gaussian Kernel $T_{\mathrm{hCov}}$ | Gaussian Kernel $T_{\mathrm{mhCov}}$ | Laplacian Kernel hCov | Laplacian Kernel mhCov | Laplacian Kernel $T_{\mathrm{hCov}}$ | Laplacian Kernel $T_{\mathrm{mhCov}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 30 | 30 | 0.010 | 0.251 | 0.929 | 0.277 | 0.944 | 0.307 | 1 | 0.336 | 1 | 0.629 | 1 | 0.686 | 1 |
| | 30 | 30 | 0.050 | 0.419 | 0.982 | 0.434 | 0.985 | 0.499 | 1 | 0.517 | 1 | 0.830 | 1 | 0.849 | 1 |
| | 30 | 30 | 0.100 | 0.532 | 0.995 | 0.532 | 0.995 | 0.613 | 1 | 0.622 | 1 | 0.905 | 1 | 0.909 | 1 |
| | 60 | 5 | 0.010 | 0.999 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 60 | 5 | 0.050 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 60 | 5 | 0.100 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | 60 | 30 | 0.010 | 0.643 | 1 | 0.684 | 1 | 0.790 | 1 | 0.833 | 1 | 0.996 | 1 | 0.999 | 1 |
| | 60 | 30 | 0.050 | 0.824 | 1 | 0.836 | 1 | 0.918 | 1 | 0.930 | 1 | 1.000 | 1 | 1.000 | 1 |
| | 60 | 30 | 0.100 | 0.894 | 1 | 0.896 | 1 | 0.955 | 1 | 0.958 | 1 | 1 | 1 | 1 | 1 |
| (iii) | 10 | 5 | 0.010 | 0.043 | 0.233 | 0.060 | 0.257 | 0.042 | 0.434 | 0.053 | 0.447 | 0.076 | 0.768 | 0.098 | 0.785 |
| | 10 | 5 | 0.050 | 0.121 | 0.466 | 0.141 | 0.490 | 0.119 | 0.680 | 0.137 | 0.698 | 0.191 | 0.924 | 0.214 | 0.927 |
| | 10 | 5 | 0.100 | 0.201 | 0.616 | 0.212 | 0.624 | 0.203 | 0.808 | 0.210 | 0.810 | 0.291 | 0.963 | 0.298 | 0.964 |
| | 10 | 30 | 0.010 | 0.017 | 0.260 | 0.013 | 0.242 | 0.017 | 0.482 | 0.012 | 0.445 | 0.021 | 0.830 | 0.017 | 0.811 |
| | 10 | 30 | 0.050 | 0.062 | 0.488 | 0.062 | 0.487 | 0.063 | 0.729 | 0.062 | 0.727 | 0.071 | 0.941 | 0.070 | 0.940 |
| | 10 | 30 | 0.100 | 0.120 | 0.632 | 0.116 | 0.630 | 0.118 | 0.837 | 0.115 | 0.836 | 0.131 | 0.972 | 0.130 | 0.975 |
| | 30 | 5 | 0.010 | 0.146 | 0.999 | 0.191 | 1 | 0.153 | 1 | 0.187 | 1 | 0.464 | 1 | 0.529 | 1 |
| | 30 | 5 | 0.050 | 0.346 | 1 | 0.375 | 1 | 0.347 | 1 | 0.380 | 1 | 0.723 | 1 | 0.747 | 1 |
| | 30 | 5 | 0.100 | 0.484 | 1 | 0.497 | 1 | 0.496 | 1 | 0.501 | 1 | 0.835 | 1 | 0.840 | 1 |
| | 30 | 30 | 0.010 | 0.024 | 1.000 | 0.022 | 1.000 | 0.026 | 1 | 0.022 | 1 | 0.038 | 1 | 0.037 | 1 |
| | 30 | 30 | 0.050 | 0.088 | 1 | 0.085 | 1 | 0.086 | 1 | 0.085 | 1 | 0.117 | 1 | 0.115 | 1 |
| | 30 | 30 | 0.100 | 0.149 | 1 | 0.147 | 1 | 0.148 | 1 | 0.144 | 1 | 0.195 | 1 | 0.193 | 1 |
| | 60 | 5 | 0.010 | 0.547 | 1 | 0.630 | 1 | 0.566 | 1 | 0.642 | 1 | 0.978 | 1 | 0.988 | 1 |
| | 60 | 5 | 0.050 | 0.802 | 1 | 0.835 | 1 | 0.808 | 1 | 0.836 | 1 | 0.997 | 1 | 0.998 | 1 |
| | 60 | 5 | 0.100 | 0.907 | 1 | 0.911 | 1 | 0.905 | 1 | 0.913 | 1 | 0.999 | 1 | 0.999 | 1 |
| | 60 | 30 | 0.010 | 0.038 | 1 | 0.030 | 1 | 0.038 | 1 | 0.029 | 1 | 0.089 | 1 | 0.080 | 1 |
| | 60 | 30 | 0.050 | 0.122 | 1 | 0.117 | 1 | 0.119 | 1 | 0.119 | 1 | 0.217 | 1 | 0.214 | 1 |
| | 60 | 30 | 0.100 | 0.198 | 1 | 0.196 | 1 | 0.199 | 1 | 0.197 | 1 | 0.326 | 1 | 0.325 | 1 |

TABLE 3

*Power comparison under $H_{A_s}$ from Example 3.3*

| | | | | | | | | Gaussian Kernel | | | | Laplacian Kernel | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $p$ | $\alpha$ | dCov | mdCov | $T_{\mathrm{dCov}}$ | $T_{\mathrm{mdCov}}$ | hCov | mhCov | $T_{\mathrm{hCov}}$ | $T_{\mathrm{mhCov}}$ | hCov | mhCov | $T_{\mathrm{hCov}}$ | $T_{\mathrm{mhCov}}$ |
| (i) | 10 | 5 | 0.010 | 0.044 | 0.196 | 0.055 | 0.218 | 0.042 | 0.348 | 0.052 | 0.367 | 0.074 | 0.672 | 0.098 | 0.685 |
| | 10 | 5 | 0.050 | 0.120 | 0.390 | 0.136 | 0.416 | 0.114 | 0.582 | 0.129 | 0.604 | 0.183 | 0.859 | 0.209 | 0.870 |
| | 10 | 5 | 0.100 | 0.201 | 0.542 | 0.209 | 0.546 | 0.191 | 0.722 | 0.197 | 0.731 | 0.292 | 0.927 | 0.304 | 0.931 |
| | 10 | 30 | 0.010 | 0.018 | 0.212 | 0.014 | 0.194 | 0.017 | 0.387 | 0.014 | 0.362 | 0.022 | 0.722 | 0.017 | 0.706 |
| | 10 | 30 | 0.050 | 0.066 | 0.434 | 0.064 | 0.428 | 0.066 | 0.627 | 0.064 | 0.625 | 0.075 | 0.892 | 0.077 | 0.891 |
| | 10 | 30 | 0.100 | 0.123 | 0.571 | 0.121 | 0.568 | 0.123 | 0.749 | 0.119 | 0.750 | 0.135 | 0.944 | 0.132 | 0.946 |
| | 30 | 5 | 0.010 | 0.158 | 0.988 | 0.197 | 0.996 | 0.136 | 1 | 0.163 | 1 | 0.486 | 1 | 0.555 | 1 |
| | 30 | 5 | 0.050 | 0.341 | 1.000 | 0.369 | 1 | 0.303 | 1 | 0.328 | 1 | 0.725 | 1 | 0.756 | 1 |
| | 30 | 5 | 0.100 | 0.483 | 1 | 0.488 | 1 | 0.433 | 1 | 0.444 | 1 | 0.838 | 1 | 0.846 | 1 |
| | 30 | 30 | 0.010 | 0.026 | 0.996 | 0.023 | 0.996 | 0.027 | 1.000 | 0.022 | 1.000 | 0.043 | 1 | 0.038 | 1 |
| | 30 | 30 | 0.050 | 0.089 | 1.000 | 0.084 | 0.999 | 0.088 | 1 | 0.083 | 1 | 0.123 | 1 | 0.125 | 1 |
| | 30 | 30 | 0.100 | 0.153 | 1.000 | 0.152 | 1.000 | 0.151 | 1 | 0.152 | 1 | 0.209 | 1 | 0.204 | 1 |
| | 60 | 5 | 0.010 | 0.559 | 1 | 0.637 | 1 | 0.461 | 1 | 0.539 | 1 | 0.989 | 1 | 0.996 | 1 |
| | 60 | 5 | 0.050 | 0.816 | 1 | 0.847 | 1 | 0.738 | 1 | 0.774 | 1 | 1.000 | 1 | 1 | 1 |
| | 60 | 5 | 0.100 | 0.916 | 1 | 0.925 | 1 | 0.861 | 1 | 0.870 | 1 | 1 | 1 | 1 | 1 |
| | 60 | 30 | 0.010 | 0.037 | 1 | 0.032 | 1 | 0.036 | 1 | 0.031 | 1 | 0.091 | 1 | 0.085 | 1 |
| | 60 | 30 | 0.050 | 0.125 | 1 | 0.119 | 1 | 0.122 | 1 | 0.115 | 1 | 0.231 | 1 | 0.228 | 1 |
| | 60 | 30 | 0.100 | 0.208 | 1 | 0.207 | 1 | 0.204 | 1 | 0.202 | 1 | 0.350 | 1 | 0.346 | 1 |
| (ii) | 10 | 5 | 0.010 | 0.044 | 0.217 | 0.059 | 0.242 | 0.040 | 0.393 | 0.055 | 0.413 | 0.077 | 0.713 | 0.106 | 0.732 |
| | 10 | 5 | 0.050 | 0.124 | 0.432 | 0.141 | 0.453 | 0.117 | 0.637 | 0.131 | 0.655 | 0.202 | 0.886 | 0.224 | 0.895 |
| | 10 | 5 | 0.100 | 0.210 | 0.577 | 0.213 | 0.583 | 0.196 | 0.771 | 0.204 | 0.775 | 0.304 | 0.942 | 0.318 | 0.942 |
| | 10 | 30 | 0.010 | 0.020 | 0.247 | 0.013 | 0.224 | 0.019 | 0.439 | 0.013 | 0.409 | 0.022 | 0.774 | 0.018 | 0.763 |
| | 10 | 30 | 0.050 | 0.064 | 0.474 | 0.064 | 0.474 | 0.063 | 0.677 | 0.063 | 0.676 | 0.075 | 0.913 | 0.076 | 0.913 |
| | 10 | 30 | 0.100 | 0.126 | 0.606 | 0.125 | 0.604 | 0.126 | 0.795 | 0.126 | 0.790 | 0.141 | 0.956 | 0.138 | 0.955 |
| | 30 | 5 | 0.010 | 0.178 | 0.995 | 0.221 | 0.999 | 0.148 | 1 | 0.186 | 1 | 0.544 | 1 | 0.608 | 1 |
| | 30 | 5 | 0.050 | 0.376 | 1 | 0.409 | 1 | 0.333 | 1 | 0.358 | 1 | 0.775 | 1 | 0.797 | 1 |
| | 30 | 5 | 0.100 | 0.518 | 1 | 0.526 | 1 | 0.468 | 1 | 0.478 | 1 | 0.871 | 1 | 0.880 | 1 |

TABLE 3
(*Continued*)

| | $n$ | $p$ | $\alpha$ | dCov | mdCov | $T_{\text{dCov}}$ | $T_{\text{mdCov}}$ | Gaussian Kernel hCov | Gaussian Kernel mhCov | Gaussian Kernel $T_{\text{hCov}}$ | Gaussian Kernel $T_{\text{mhCov}}$ | Laplacian Kernel hCov | Laplacian Kernel mhCov | Laplacian Kernel $T_{\text{hCov}}$ | Laplacian Kernel $T_{\text{mhCov}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 30 | 30 | 0.010 | 0.027 | 0.998 | 0.023 | 0.998 | 0.026 | 1.000 | 0.022 | 1 | 0.043 | 1 | 0.038 | 1 |
| | 30 | 30 | 0.050 | 0.088 | 1.000 | 0.087 | 1.000 | 0.088 | 1 | 0.086 | 1 | 0.128 | 1 | 0.128 | 1 |
| | 30 | 30 | 0.100 | 0.155 | 1.000 | 0.152 | 1.000 | 0.154 | 1 | 0.152 | 1 | 0.218 | 1 | 0.213 | 1 |
| | 60 | 5 | 0.010 | 0.632 | 1 | 0.709 | 1 | 0.526 | 1 | 0.609 | 1 | 0.995 | 1 | 0.999 | 1 |
| | 60 | 5 | 0.050 | 0.870 | 1 | 0.895 | 1 | 0.792 | 1 | 0.826 | 1 | 1 | 1 | 1 | 1 |
| | 60 | 5 | 0.100 | 0.946 | 1 | 0.952 | 1 | 0.904 | 1 | 0.911 | 1 | 1 | 1 | 1 | 1 |
| | 60 | 30 | 0.010 | 0.044 | 1 | 0.037 | 1 | 0.043 | 1 | 0.036 | 1 | 0.105 | 1 | 0.096 | 1 |
| | 60 | 30 | 0.050 | 0.126 | 1 | 0.125 | 1 | 0.123 | 1 | 0.121 | 1 | 0.251 | 1 | 0.244 | 1 |
| | 60 | 30 | 0.100 | 0.213 | 1 | 0.211 | 1 | 0.211 | 1 | 0.206 | 1 | 0.368 | 1 | 0.366 | 1 |
| (iii) | 10 | 5 | 0.010 | 0.019 | 0.024 | 0.023 | 0.028 | 0.017 | 0.033 | 0.022 | 0.040 | 0.023 | 0.090 | 0.029 | 0.095 |
| | 10 | 5 | 0.050 | 0.058 | 0.079 | 0.068 | 0.089 | 0.057 | 0.111 | 0.067 | 0.115 | 0.068 | 0.232 | 0.081 | 0.242 |
| | 10 | 5 | 0.100 | 0.113 | 0.148 | 0.117 | 0.151 | 0.114 | 0.194 | 0.118 | 0.196 | 0.124 | 0.351 | 0.129 | 0.355 |
| | 10 | 30 | 0.010 | 0.016 | 0.026 | 0.012 | 0.020 | 0.016 | 0.037 | 0.012 | 0.030 | 0.017 | 0.089 | 0.013 | 0.076 |
| | 10 | 30 | 0.050 | 0.059 | 0.086 | 0.057 | 0.083 | 0.060 | 0.112 | 0.058 | 0.105 | 0.061 | 0.233 | 0.060 | 0.225 |
| | 10 | 30 | 0.100 | 0.111 | 0.156 | 0.108 | 0.153 | 0.112 | 0.199 | 0.108 | 0.193 | 0.112 | 0.357 | 0.109 | 0.346 |
| | 30 | 5 | 0.010 | 0.019 | 0.051 | 0.021 | 0.068 | 0.017 | 0.141 | 0.021 | 0.170 | 0.026 | 0.673 | 0.032 | 0.724 |
| | 30 | 5 | 0.050 | 0.061 | 0.166 | 0.070 | 0.188 | 0.058 | 0.339 | 0.066 | 0.360 | 0.083 | 0.889 | 0.091 | 0.903 |
| | 30 | 5 | 0.100 | 0.117 | 0.283 | 0.117 | 0.288 | 0.117 | 0.488 | 0.116 | 0.497 | 0.153 | 0.953 | 0.153 | 0.955 |
| | 30 | 30 | 0.010 | 0.017 | 0.074 | 0.012 | 0.065 | 0.017 | 0.182 | 0.012 | 0.165 | 0.017 | 0.754 | 0.012 | 0.742 |
| | 30 | 30 | 0.050 | 0.061 | 0.202 | 0.058 | 0.198 | 0.061 | 0.378 | 0.059 | 0.373 | 0.063 | 0.913 | 0.061 | 0.913 |
| | 30 | 30 | 0.100 | 0.112 | 0.309 | 0.110 | 0.307 | 0.113 | 0.518 | 0.110 | 0.517 | 0.117 | 0.960 | 0.114 | 0.959 |
| | 60 | 5 | 0.010 | 0.019 | 0.174 | 0.024 | 0.219 | 0.017 | 0.580 | 0.022 | 0.666 | 0.034 | 1.000 | 0.041 | 1 |
| | 60 | 5 | 0.050 | 0.066 | 0.421 | 0.073 | 0.458 | 0.061 | 0.853 | 0.069 | 0.883 | 0.108 | 1 | 0.119 | 1 |
| | 60 | 5 | 0.100 | 0.123 | 0.600 | 0.128 | 0.612 | 0.119 | 0.941 | 0.122 | 0.949 | 0.179 | 1 | 0.183 | 1 |
| | 60 | 30 | 0.010 | 0.013 | 0.251 | 0.009 | 0.233 | 0.013 | 0.680 | 0.010 | 0.665 | 0.014 | 1.000 | 0.010 | 1 |
| | 60 | 30 | 0.050 | 0.053 | 0.485 | 0.051 | 0.484 | 0.052 | 0.869 | 0.050 | 0.871 | 0.056 | 1 | 0.055 | 1 |
| | 60 | 30 | 0.100 | 0.105 | 0.620 | 0.101 | 0.619 | 0.106 | 0.930 | 0.101 | 0.929 | 0.107 | 1 | 0.106 | 1 |

loss corresponding to Laplacian kernel is consistently less than that for Gaussian kernel. In general, we observe that the tests based on distance covariance, Hilbert–Schmidt covariance with Gaussian kernel and Hilbert–Schmidt covariance with Laplacian kernel, are all admissible, as none of them dominate the others in all situations. In the following example, we examine the aforementioned tests on a real data set.

EXAMPLE 3.4. We consider the Earthquake data set, which is originally from the Northern California Earthquake Data Center and has classes of positive and negative major earthquake events. There are 368 negative and 93 positive cases and each data point is of length 512. This data set can be downloaded from UCR Time Series Classification Archive [Dau et al. (2018)]. Here, we only consider the negative cases. Let $Z_i = (z_{i,1}, z_{i,2}, \dots, z_{i,512})^T$ denote the record of a negative event, then set $X_i = (z_{i,150-p+1}, \dots, z_{i,150})^T$ and $Y_i = (z_{i,151}, \dots, z_{i,150+p})^T$, $i = 1, \dots, 368$. We apply all tests to test the independence between $X_i$ and $Y_i$, which are expected to be dependent due to the serial nature of $Z_i$. For each $p = 5, 30$ and $n = 10, 30, 60$, we randomly sample $n$ rows from the full data set $(X_i, Y_i)_{i=1}^{368}$ without replacement and apply the aforementioned tests based on the subsample. Next, the above procedure is repeated 5000 times to calculate the power.

The results are presented in Table 4. It can be seen that the powers of the marginal tests increase as the dimension grows, whereas the powers of all joint tests experience a decay as $p$ grows and are nearly trivial when $p = 30$. This finding is consistent for all tests including hCov-based ones with Gaussian and Laplacian kernels. In addition, we also note that the marginal tests with Gaussian or Laplacian kernel have consistently higher power as compared to the Euclidean distance based tests.

**4. Conclusion.** In this article, we investigate the behavior of the distance covariance and Hilbert–Schmidt covariance in the high dimensional setting. We discover that the standard distance covariance and Hilbert–Schmidt covariance, which are well known to capture nonlinear dependence in low/fixed dimensional context, can only capture linear componentwise cross-dependence (to the first order) in the high dimensional environment. We believe that this is a new finding that may have significant implications to the design of tests for independence for high dimensional data. On one hand, we reveal the limitation of distance covariance and variants in the high dimensional context, and suggest to use marginally aggregated (sample) distance covariance as a way out, where the latter targets the low dimensional nonlinear dependence. On the other hand, we speculate whether it is possible to capture all kinds of dependence between high dimensional vectors $X$ and $Y$, in a limited sample size framework. If the sample size is fixed, we would conjecture that an omnibus test does not exist; If the sample size can grow faster than the dimension, it seems possible but unclear to us how to develop an omnibus test in an asymptotic sense. We hope the results presented in this paper shed some light on the challenges in the high dimensional dependence testing and will motivate more work in this area.

**Acknowledgments.** We would like to thank three reviewers and Associate Editor for helpful comments, which have improved the presentation of the paper. We are grateful to Dr. Liping Zhu for sending us the R codes for Zhu et al. (2017).

SUPPLEMENTARY MATERIAL

**Supplement to "Distance-based and RKHS-based dependence metrics in high dimension"** (DOI: 10.1214/19-AOS1934SUPP; .pdf). This supplement contains (1) some theoretical results derived under the high dimension and medium sample size setting, where the

TABLE 4

*Power comparison on earthquake data*

| | | | | | | | Gaussian Kernel | | | | Laplacian Kernel | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | $p$ | $\alpha$ | dCov | mdCov | $T_{\mathrm{dCov}}$ | $T_{\mathrm{mdCov}}$ | hCov | mhCov | $T_{\mathrm{hCov}}$ | $T_{\mathrm{mhCov}}$ | hCov | mhCov | $T_{\mathrm{hCov}}$ | $T_{\mathrm{mhCov}}$ |
| 10 | 5 | 0.010 | 0.021 | 0.054 | 0.041 | 0.079 | 0.021 | 0.851 | 0.038 | 0.883 | 0.035 | 0.937 | 0.060 | 0.952 |
| 10 | 5 | 0.050 | 0.070 | 0.144 | 0.085 | 0.160 | 0.065 | 0.927 | 0.080 | 0.937 | 0.091 | 0.975 | 0.112 | 0.979 |
| 10 | 5 | 0.100 | 0.120 | 0.235 | 0.126 | 0.230 | 0.114 | 0.956 | 0.117 | 0.959 | 0.155 | 0.985 | 0.160 | 0.985 |
| 10 | 30 | 0.010 | 0.012 | 0.218 | 0.013 | 0.226 | 0.012 | 1.000 | 0.012 | 1.000 | 0.013 | 1 | 0.016 | 1 |
| 10 | 30 | 0.050 | 0.046 | 0.412 | 0.047 | 0.421 | 0.046 | 1.000 | 0.046 | 1.000 | 0.050 | 1 | 0.054 | 1 |
| 10 | 30 | 0.100 | 0.095 | 0.537 | 0.093 | 0.541 | 0.096 | 1 | 0.095 | 1 | 0.094 | 1 | 0.091 | 1 |
| 30 | 5 | 0.010 | 0.034 | 0.155 | 0.055 | 0.196 | 0.026 | 1 | 0.042 | 1 | 0.078 | 1 | 0.124 | 1 |
| 30 | 5 | 0.050 | 0.106 | 0.333 | 0.128 | 0.356 | 0.082 | 1 | 0.096 | 1 | 0.188 | 1 | 0.210 | 1 |
| 30 | 5 | 0.100 | 0.177 | 0.460 | 0.183 | 0.461 | 0.139 | 1 | 0.146 | 1 | 0.278 | 1 | 0.273 | 1 |
| 30 | 30 | 0.010 | 0.009 | 0.936 | 0.009 | 0.941 | 0.009 | 1 | 0.009 | 1 | 0.011 | 1 | 0.012 | 1 |
| 30 | 30 | 0.050 | 0.040 | 0.977 | 0.041 | 0.976 | 0.043 | 1 | 0.043 | 1 | 0.040 | 1 | 0.041 | 1 |
| 30 | 30 | 0.100 | 0.081 | 0.988 | 0.080 | 0.988 | 0.086 | 1 | 0.085 | 1 | 0.085 | 1 | 0.082 | 1 |
| 60 | 5 | 0.010 | 0.060 | 0.473 | 0.086 | 0.549 | 0.034 | 1 | 0.050 | 1 | 0.171 | 1 | 0.245 | 1 |
| 60 | 5 | 0.050 | 0.147 | 0.722 | 0.171 | 0.749 | 0.096 | 1 | 0.107 | 1 | 0.341 | 1 | 0.370 | 1 |
| 60 | 5 | 0.100 | 0.240 | 0.835 | 0.244 | 0.838 | 0.162 | 1 | 0.167 | 1 | 0.457 | 1 | 0.458 | 1 |
| 60 | 30 | 0.010 | 0.006 | 1 | 0.006 | 1 | 0.006 | 1 | 0.006 | 1 | 0.008 | 1 | 0.008 | 1 |
| 60 | 30 | 0.050 | 0.030 | 1 | 0.031 | 1 | 0.032 | 1 | 0.031 | 1 | 0.033 | 1 | 0.034 | 1 |
| 60 | 30 | 0.100 | 0.066 | 1 | 0.065 | 1 | 0.068 | 1 | 0.070 | 1 | 0.067 | 1 | 0.066 | 1 |

sample size is also allowed to grow as the dimension grows, albeit at a slower rate; (2) additional simulation examples, and comparisons with five nonparametric dependence tests that are not covered by our kernel class; (3) proof details for all theory presented in the paper.

## REFERENCES


AHN, J., MARRON, J. S., MULLER, K. M. and CHI, Y.-Y. (2007). The high-dimension, low-sample-size geometric representation holds under mild conditions. *Biometrika* **94** 760–766. MR2410023 https://doi.org/10.1093/biomet/asm050

AOSHIMA, M., SHEN, D., SHEN, H., YATA, K., ZHOU, Y.-H. and MARRON, J. S. (2018). A survey of high dimension low sample size asymptotics. *Aust. N. Z. J. Stat.* **60** 4–19. MR3780619 https://doi.org/10.1111/anzs.12212

BEDO, J. (2008). Microarray design using the Hilbert–Schmidt independence criterion. In *IAPR International Conference on Pattern Recognition in Bioinformatics* 288–298. Springer, Berlin.

BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli* **20** 1006–1028. MR3178526 https://doi.org/10.3150/13-BEJ514

BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 https://doi.org/10.1093/biomet/asz024

BLUM, J. R., KIEFER, J. and ROSENBLATT, M. (1961). Distribution free tests of independence based on the sample distribution function. *Ann. Math. Stat.* **32** 485–498. MR0125690 https://doi.org/10.1214/aoms/1177705055

BRADLEY, R. C. (2007). *Introduction to Strong Mixing Conditions. Vol.* 1. Kendrick Press, Heber City, UT. MR2325295

CHAKRABORTY, S. and ZHANG, X. (2019a). Distance metrics for measuring joint dependence with application to causal inference. *J. Amer. Statist. Assoc.* **114** 1638–1650. MR4047289 https://doi.org/10.1080/01621459.2018.1513364

CHAKRABORTY, S. and ZHANG, X. (2019b). A new framework for distance and kernel-based metrics in high dimensions. Available at https://arxiv.org/abs/1909.13469.

CSÖRGŐ, S. (1985). Testing for independence by the empirical characteristic function. *J. Multivariate Anal.* **16** 290–299. MR0793494 https://doi.org/10.1016/0047-259X(85)90022-3

DASGUPTA, A. (2008). *Asymptotic Theory of Statistics and Probability. Springer Texts in Statistics.* Springer, New York. MR2664452

DAU, H. A., KEOGH, E., KAMGAR, K., YEH, C.-C. M., ZHU, Y., GHARGHABI, S., RATANAMAHATANA, C. A., YANPING, HU, B. et al. (2018). The UCR time series classification archive. Available at https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

DAVIDSON, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians. Advanced Texts in Econometrics.* Oxford Univ. Press, New York. MR1430804 https://doi.org/10.1093/0198774036.001.0001

DE WET, T. (1980). Cramér–von Mises tests for independence. *J. Multivariate Anal.* **10** 38–50. MR0569795 https://doi.org/10.1016/0047-259X(80)90080-9

DEHEUVELS, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Multivariate Anal.* **11** 102–113. MR0612295 https://doi.org/10.1016/0047-259X(81)90136-6

DOUKHAN, P. and NEUMANN, M. H. (2008). The notion of $\psi$-weak dependence and its applications to bootstrapping time series. *Probab. Surv.* **5** 146–168. MR2426177 https://doi.org/10.1214/06-PS086

DUECK, J., EDELMANN, D., GNEITING, T. and RICHARDS, D. (2014). The affinely invariant distance correlation. *Bernoulli* **20** 2305–2330. MR3263106 https://doi.org/10.3150/13-BEJ558

EDELMANN, D., RICHARDS, D. and VOGEL, D. (2017). The distance standard deviation. Preprint. Available at arXiv:1705.05777.

GALLANT, A. R. and WHITE, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.* Blackwell, Oxford.

GIESER, P. W. and RANDLES, R. H. (1997). A nonparametric test of independence between two vectors. *J. Amer. Statist. Assoc.* **92** 561–567. MR1467849 https://doi.org/10.2307/2965704

GRETTON, A., FUKUMIZU, K., TEO, C. H., SONG, L., SCHÖLKOPF, B. and SMOLA, A. J. (2008). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* 585–592.

GRETTON, A., BORGWARDT, K. M., RASCH, M. J., SCHÖLKOPF, B. and SMOLA, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

HALL, P., MARRON, J. S. and NEEMAN, A. (2005). Geometric representation of high dimension, low sample size data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 427–444. MR2155347 https://doi.org/10.1111/j.1467-9868.2005.00510.x

HETTMANSPERGER, T. P. and OJA, H. (1994). Affine invariant multivariate multisample sign tests. *J. Roy. Statist. Soc. Ser. B* **56** 235–249. MR1257810

HOEFFDING, W. (1948). A non-parametric test of independence. *Ann. Math. Stat.* **19** 546–557. MR0029139 https://doi.org/10.1214/aoms/1177730150

HUA, W.-Y. and GHOSH, D. (2015). Equivalence of kernel machine regression and kernel distance covariance for multidimensional phenotype association studies. *Biometrics* **71** 812–820. MR3402617 https://doi.org/10.1111/biom.12314

JUNG, S. and MARRON, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37** 4104–4130. MR2572454 https://doi.org/10.1214/09-AOS709

KONG, J., KLEIN, B. E., KLEIN, R., LEE, K. E. and WAHBA, G. (2012). Using distance correlation and SS-ANOVA to assess associations of familial relationships, lifestyle factors, diseases, and mortality. *Proc. Natl. Acad. Sci. USA* **109** 20352–20357.

KROUPI, E., YAZDANI, A., VESIN, J.-M. and EBRAHIMI, T. (2012). Multivariate spectral analysis for identifying the brain activations during olfactory perception. In 2012 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 6172–6175. IEEE, New York.

KROUPI, E., YAZDANI, A., VESIN, J.-M. and EBRAHIMI, T. (2014). EEG correlates of pleasant and unpleasant odor perception. *ACM Trans. Multimed. Comput. Commun. Appl.* **11** 13.

LEUNG, D. and DRTON, M. (2018). Testing independence in high dimensions with sums of rank correlations. *Ann. Statist.* **46** 280–307. MR3766953 https://doi.org/10.1214/17-AOS1550

LI, R., ZHONG, W. and ZHU, L. (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc.* **107** 1129–1139. MR3010900 https://doi.org/10.1080/01621459.2012.695654

LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. MR3127883 https://doi.org/10.1214/12-AOP803

MATTESON, D. S. and TSAY, R. S. (2017). Independent component analysis via distance covariance. *J. Amer. Statist. Assoc.* **112** 623–637. MR3671757 https://doi.org/10.1080/01621459.2016.1150851

MIKALSEN, K. Ø., SOGUERO-RUIZ, C., BIANCHI, F. M. and JENSSEN, R. (2019). Noisy multi-label semi-supervised dimensionality reduction. *Pattern Recognit.* **90** 257–270.

PAN, G., GAO, J. and YANG, Y. (2014). Testing independence among a large number of high-dimensional random vectors. *J. Amer. Statist. Assoc.* **109** 600–612. MR3223736 https://doi.org/10.1080/01621459.2013.872037

PARK, T., SHAO, X. and YAO, S. (2015). Partial martingale difference correlation. *Electron. J. Stat.* **9** 1492–1517. MR3367668 https://doi.org/10.1214/15-EJS1047

RAMDAS, A., REDDI, S. J., PÓCZOS, B., SINGH, A. and WASSERMAN, L. (2015). On the decreasing power of kernel and distance based nonparametric hypothesis tests in high dimensions. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. MR3127866 https://doi.org/10.1214/13-AOS1140

SHAO, X. and ZHANG, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *J. Amer. Statist. Assoc.* **109** 1302–1318. MR3265698 https://doi.org/10.1080/01621459.2014.887012

SINHA, B. K. and WIEAND, H. S. (1977). Multivariate nonparametric tests for independence. *J. Multivariate Anal.* **7** 572–583. MR0471173 https://doi.org/10.1016/0047-259X(77)90068-9

SZEKELY, G. J. and RIZZO, M. L. (2004). Testing for equal distributions in high dimension. *InterStat* **5** 1–6.

SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127 https://doi.org/10.1214/09-AOAS312

SZÉKELY, G. J. and RIZZO, M. L. (2013). The distance correlation $t$-test of independence in high dimension. *J. Multivariate Anal.* **117** 193–213. MR3053543 https://doi.org/10.1016/j.jmva.2013.02.012

SZÉKELY, G. J. and RIZZO, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42** 2382–2412. MR3269983 https://doi.org/10.1214/14-AOS1255

SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665 https://doi.org/10.1214/009053607000000505

TASKINEN, S., KANKAINEN, A. and OJA, H. (2003). Sign test of independence between two random vectors. *Statist. Probab. Lett.* **62** 9–21. MR1965367 https://doi.org/10.1016/S0167-7152(02)00399-1

WEI, S., LEE, C., WICHERS, L. and MARRON, J. S. (2016). Direction-projection-permutation for high-dimensional hypothesis tests. *J. Comput. Graph. Statist.* **25** 549–569. MR3499694 https://doi.org/10.1080/10618600.2015.1027773

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215 https://doi.org/10.1073/pnas.0506715102

XU, J., LIU, J., YIN, J. and SUN, C. (2016). A multi-label feature extraction algorithm via maximizing feature variance and feature-label dependence simultaneously. *Knowl.-Based Syst.* **98** 172–184.

YANG, Y. (2017). Source-space analyses in MEG/EEG and applications to explore spatio-temporal neural dynamics in human vision. Ph.D. thesis.

YANG, Y. and PAN, G. (2015). Independence test for high dimensional data based on regularized canonical correlation coefficients. *Ann. Statist.* **43** 467–500. MR3316187 https://doi.org/10.1214/14-AOS1284

YAO, S., ZHANG, X. and SHAO, X. (2018). Testing mutual independence in high dimension via distance covariance. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 455–480. MR3798874 https://doi.org/10.1111/rssb.12259

ZHANG, X., YAO, S. and SHAO, X. (2018). Conditional mean and quantile dependence testing in high dimension. *Ann. Statist.* **46** 219–246. MR3766951 https://doi.org/10.1214/17-AOS1548

ZHANG, Y. and ZHOU, Z.-H. (2010). Multilabel dimensionality reduction via dependence maximization. *ACM Trans. Knowl. Discov. Data* **4** 14.

ZHOU, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.* **33** 438–457. MR2915095 https://doi.org/10.1111/j.1467-9892.2011.00780.x

ZHU, C. and SHAO, X. (2019). Interpoint distance based two sample tests in high dimension. Available at https://arxiv.org/pdf/1902.07279.pdf.

ZHU, L., XU, K., LI, R. and ZHONG, W. (2017). Projection correlation between two random vectors. *Biometrika* **104** 829–843. MR3737307 https://doi.org/10.1093/biomet/asx043

ZHU, C., ZHANG, X., YAO, S. and SHAO, X. (2020). Supplement to "Distance-based and RKHS-based dependence metrics in high dimension." https://doi.org/10.1214/19-AOS1934SUPP.