A new framework for distance and kernel-based metrics in high dimensions

Shubhadeep Chakraborty

Department of Biostatistics, University of Washington
e-mail: deep20@uw.edu

Xianyang Zhang

Department of Statistics, Texas A&M University
e-mail: zhangxiany@stat.tamu.edu

Abstract: The paper presents new metrics to quantify and test for (i) the equality of distributions and (ii) the independence between two high-dimensional random vectors. We show that the energy distance based on the usual Euclidean distance cannot completely characterize the homogeneity of two high-dimensional distributions in the sense that it only detects the equality of means and the traces of covariance matrices in the high-dimensional setup. We propose a new class of metrics which inherits the desirable properties of the energy distance and maximum mean discrepancy/(generalized) distance covariance and the Hilbert-Schmidt Independence Criterion in the low-dimensional setting and is capable of detecting the homogeneity of/completely characterizing independence between the low-dimensional marginal distributions in the high dimensional setup. We further propose t-tests based on the new metrics to perform high-dimensional two-sample testing/independence testing and study their asymptotic behavior under both high dimension low sample size (HDLSS) and high dimension medium sample size (HDMSS) setups. The computational complexity of the t-tests only grows linearly with the dimension and thus is scalable to very high dimensional data. We demonstrate the superior power behavior of the proposed tests for homogeneity of distributions and independence via both simulated and real datasets.

Keywords and phrases: Distance covariance, energy distance, high dimensionality, Hilbert-Schmidt independence criterion, independence test, maximum mean discrepancy, two sample test, U-statistic.

1. Introduction

Nonparametric two-sample testing of homogeneity of distributions has been a classical problem in statistics, finding a plethora of applications in goodness-of-fit testing, clustering, change-point detection and so on. Some of the most traditional tools in this domain are Kolmogorov-Smirnov test, and Wald-Wolfowitz runs test, whose multivariate and multidimensional extensions have been studied by [10, 12, 4] among others. [16] proposed a distribution-free multivariate generalization of the Wald-Wolfowitz runs test applicable for arbitrary but fixed dimensions. [38] proposed another distribution-free test for multivariate

two-sample problem based on k-nearest neighbor (k-NN) graphs. [28] suggested a technique for reducing the dimensionality by examining the distribution of interpoint distances. In a recent novel work, [8] proposed graph-based tests for moderate to high dimensional data and non-Euclidean data. The last two decades have seen an abundance of literature on distance and kernel-based tests for equality of distributions. Energy distance (first introduced by [44]) and maximum mean discrepancy or MMD (see [19]) have been widely studied in both the statistics and machine learning communities. [39] provided a unifying framework establishing the equivalence between the (generalized) energy distance and MMD. Although there have been some very recent works to gain insight on the decaying power of the distance and kernel-based tests for high dimensional inference (see for example [32, 33, 24, 25]), the behavior of these tests in the high dimensional setup is still a pretty unexplored area.

Measuring and testing for independence between two random vectors has been another fundamental problem in statistics, which has found applications in a wide variety of areas such as independent component analysis, feature selection, graphical modeling, causal inference, etc. There has been an enormous amount of literature on developing dependence metrics to quantify non-linear and non-monotone dependence in the low dimensional context. [17, 18] introduced a kernel-based independence measure, namely the Hilbert-Schmidt Independence Criterion (HSIC). [2] proposed a consistent test of independence of two ordinal random variables based on an extension of Kendall's tau. [23] suggested tests of independence based on the RV coefficient. [56] proposed a methodology to characterize independence using projection correlation. For other recent works on measuring and testing for independence in low dimensions, we refer the reader to [51, 36, 37, 3, 7, 53, 27] among others. [47], in their seminal paper, introduced distance covariance (dCov) to characterize dependence between two random vectors of arbitrary dimensions. [26] extended the notion of distance covariance from Euclidean spaces to arbitrary metric spaces. [39] established the equivalence between HSIC and (generalized) distance covariance via the correspondence between positive definite kernels and semi-metrics of negative type. Over the last decade, the idea of distance covariance has been widely extended and analyzed in various ways; see for example [55, 49, 50, 43, 21, 54, 15] among many others. There have been some very recent literature which aims at generalizing distance covariance to quantify the joint dependence among more than two random vectors; see for example [29, 22, 6, 5, 52]. However, in the high dimensional setup, the literature is scarce, and the behavior of the widely used distance and kernel-based dependence metrics is not very well explored till date. [48] proposed a distance correlation based t-test to test for independence in high dimensions. In a very recent work, [57] showed that in the high dimension low sample size (HDLSS) setting, i.e., when the dimensions grow while the sample size is held fixed, the sample distance covariance can only measure the component-wise *linear dependence* between the two vectors. As a consequence, the distance correlation based t-test proposed by [48] for independence between two high dimensional random vectors has trivial power when the two random vectors are nonlinearly dependent but component-wise uncorrelated. As a rem-

edy, [57] proposed a test by aggregating the pairwise squared sample distance covariances and studied its asymptotic behavior under the HDLSS setup.

This paper presents a new class of metrics to quantify the homogeneity of distributions and independence between two high-dimensional random vectors. The core of our methodology is a new way of defining the distance between sample points (interpoint distance) in the high-dimensional Euclidean spaces. In the first part of this work, we show that the energy distance based on the usual Euclidean distance cannot completely characterize the homogeneity of two high-dimensional distributions in the sense that it only detects the *equality of means and the traces of covariance matrices* in the high-dimensional setup. To overcome such a limitation, we propose a new class of metrics based on the new distance which inherits the nice properties of energy distance and MMD in the low-dimensional setting and is capable of detecting the *pairwise homogeneity of the low-dimensional marginal distributions* in the HDLSS setup. We construct a high-dimensional two sample t-test based on the U-statistic type estimator of the proposed metric, which can be viewed as a generalization of the classical two-sample t-test with equal variances. We show under the HDLSS setting that the new two sample t-test converges to a central t-distribution under the null and it has nontrivial power for a broader class of alternatives compared to the energy distance. We further show that the two sample t-test converges to a standard normal limit under the null when the dimension and sample size both grow to infinity with the dimension growing more rapidly. It is worth mentioning that we develop an approach to unify the analysis for the usual energy distance and the proposed metrics. Compared to existing works, we make the following contribution.

- We derive the asymptotic variance of the generalized energy distance under the HDLSS setting and propose a computationally efficient variance estimator (whose computational cost is linear in the dimension). Our analysis is based on a pivotal t-statistic which does not require permutation or resampling-based inference and allows an asymptotic exact power analysis.

In the second part, we propose a new framework to construct dependence metrics to quantify the dependence between two high-dimensional random vectors X and Y of possibly different dimensions. The new metric, denoted by $\mathcal{D}^2(X, Y)$, generalizes both the distance covariance and HSIC. It completely characterizes independence between X and Y and inherits all other desirable properties of the distance covariance and HSIC for fixed dimensions. In the HDLSS setting, we show that the proposed population dependence metric behaves as an aggregation of group-wise (generalized) distance covariances. We construct an unbiased U-statistic type estimator of $\mathcal{D}^2(X, Y)$ and show that with growing dimensions, the unbiased estimator is asymptotically equivalent to the sum of group-wise squared sample (generalized) distance covariances. Thus it can quantify *group-wise non-linear dependence* between two high-dimensional random vectors, going beyond the scope of the distance covariance based on the usual Euclidean distance and HSIC which have been recently shown only to capture the componentwise linear dependence in high dimension, see [57]. We further

propose a t-test based on the new metrics to perform high-dimensional independence testing and study its asymptotic size and power behaviors under both the HDLSS and high dimension medium sample size (HDMSS) setups. In particular, under the HDLSS setting, we prove that the proposed t-test converges to a central t-distribution under the null and a noncentral t-distribution with a random noncentrality parameter under the alternative. Through extensive numerical studies, we demonstrate that the newly proposed t-test can capture group-wise nonlinear dependence which cannot be detected by the usual distance covariance and HSIC in the high dimensional regime. Compared to the marginal aggregation approach in [57], our new method enjoys two major advantages.

- Our approach provides a neater way of generalizing the notion of distance and kernel-based dependence metrics. The newly proposed metrics completely characterize dependence in the low-dimensional case and capture group-wise nonlinear dependence in the high-dimensional case. In this sense, our metric can detect a wider range of dependence compared to the marginal aggregation approach.
- The computational complexity of the t-tests only grows linearly with the dimension and thus is scalable to very high dimensional data.

Notation. Let $X = (X_1, \ldots X_p) \in \mathbb{R}^p$ and $Y = (Y_1, \ldots, Y_q) \in \mathbb{R}^q$ be two random vectors of dimensions p and q respectively. Denote by $\|\cdot\|_p$ the Euclidean norm of $\mathbb{R}^p$ (we shall use it interchangeably with $\|\cdot\|$ when there is no confusion). Let 0_p be the origin of $\mathbb{R}^p$. We use $X \perp\!\!\!\perp Y$ to denote that X is independent of Y, and use "$X \overset{d}{=} Y$" to indicate that X and Y are identically distributed. Let (X', Y'), (X'', Y'') and (X''', Y''') be independent copies of (X, Y). 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 $\asymp$, which is defined as follows: for a sequence of random variables $\{Z_n\}_{n=1}^{\infty}$ and a sequence of real numbers $\{a_n\}_{n=1}^{\infty}$, $Z_n \asymp_p a_n$ if and only if $Z_n/a_n = O_p(1)$ and $a_n/Z_n = O_p(1)$ as $n \to \infty$. For a metric space $(\mathcal{X}, d_{\mathcal{X}})$, let $\mathcal{M}(\mathcal{X})$ and $\mathcal{M}_1(\mathcal{X})$ denote the set of all finite signed Borel measures on $\mathcal{X}$ and the subset of all probability measures on $\mathcal{X}$, respectively. We say that $v \in \mathcal{M}(\mathcal{X})$ has finite first moment if $\int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0) \, d|v|(x) < \infty$ for some $x_0 \in \mathcal{X}$. Define $\mathcal{M}_{d_{\mathcal{X}}}^1(\mathcal{X}) := \{v \in \mathcal{M}(\mathcal{X}) : \exists x_0 \in \mathcal{X} \text{ s.t. } \int_{\mathcal{X}} d_{\mathcal{X}}(x, x_0) \, d|v|(x) < \infty\}$. For $\theta > 0$, define $\mathcal{M}_{\mathcal{K}}^{\theta}(\mathcal{X}) := \{v \in \mathcal{M}(\mathcal{X}) : \int_{\mathcal{X}} \mathcal{K}^{\theta}(x, x) \, d|v|(x) < \infty\}$, where $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a bivariate kernel function. Define $\mathcal{M}_{d_{\mathcal{Y}}}^1(\mathcal{Y})$ and $\mathcal{M}_{\mathcal{K}}^{\theta}(\mathcal{Y})$ in a similar way. For a matrix $A = (a_{kl})_{k,l=1}^n \in \mathbb{R}^{n \times n}$, define its $\mathcal{U}$-centered version $\tilde{A} = (\tilde{a}_{kl}) \in \mathbb{R}^{n \times n}$ as follows

$$\tilde{a}_{kl} = \begin{cases} a_{kl} - \dfrac{1}{n-2}\sum_{j=1}^n a_{kj} - \dfrac{1}{n-2}\sum_{i=1}^n a_{il} + \dfrac{1}{(n-1)(n-2)}\sum_{i,j=1}^n a_{ij}, & k \neq l, \\ 0, & k = l, \end{cases}$$

$$(1.1)$$

for $k, l = 1, \ldots, n$. Define

$$(\tilde{A} \cdot \tilde{B}) := \frac{1}{n(n-3)} \sum_{k \neq l} \tilde{a}_{kl} \tilde{b}_{kl}$$

for $\tilde{A} = (\tilde{a}_{kl})$ and $\tilde{B} = (\tilde{b}_{kl}) \in \mathbb{R}^{n \times n}$. Denote by $\text{tr}(A)$ the trace of a square matrix A. $A \otimes B$ denotes the kronecker product of two matrices A and B. Let $\Phi(\cdot)$ be the cumulative distribution function of the standard normal distribution. Denote by $t_{a,b}$ the noncentral t-distribution with a degrees of freedom and noncentrality parameter b. Write $t_a = t_{a,0}$. Denote by $q_{\alpha,a}$ and Z_α the upper α quantile of the distribution of t_a and the standard normal distribution, respectively, for $\alpha \in (0, 1)$. Also denote by χ_a^2 the chi-square distribution with a degrees of freedom. Denote $U \sim \text{Rademacher}(0.5)$ if $P(U = 1) = P(U = -1) = 0.5$. Let $\mathbb{1}_A$ denote the indicator function associated with a set A. Finally, denote by $\lfloor a \rfloor$ the integer part of $a \in \mathbb{R}$.

2. An overview: distance and kernel-based metrics

2.1. Energy distance and MMD

Energy distance [45, 46, 1] or the Euclidean energy distance between two random vectors $X, Y \in \mathbb{R}^p$ and $X \perp\!\!\!\perp Y$ with $\mathbb{E}\|X\|_p < \infty$ and $\mathbb{E}\|Y\|_p < \infty$, is defined as

$$ED(X, Y) = 2\,\mathbb{E}\|X - Y\|_p - \mathbb{E}\|X - X'\|_p - \mathbb{E}\|Y - Y'\|_p , \qquad (2.1)$$

where (X', Y') is an independent copy of (X, Y). Theorem 1 in [46] shows that $ED(X, Y) \geq 0$ and the equality holds if and only if $X \overset{d}{=} Y$. In general, for an arbitrary metric space $(\mathcal{X}, d)$, the generalized energy distance between $X \sim P_X$ and $Y \sim P_Y$ where $P_X, P_Y \in \mathcal{M}_1(\mathcal{X}) \cap \mathcal{M}_d^1(\mathcal{X})$ is defined as

$$ED_d(X, Y) = 2\,\mathbb{E}\,d(X, Y) - \mathbb{E}\,d(X, X') - \mathbb{E}\,d(Y, Y') . \qquad (2.2)$$

Definition 2.1 (Spaces of negative type). A metric space $(\mathcal{X}, d)$ is said to have negative type if for all $n \geq 1$, $x_1, \ldots, x_n \in \mathcal{X}$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ with $\sum_{i=1}^n \alpha_i = 0$, we have

$$\sum_{i,j=1}^n \alpha_i\, \alpha_j\, d(x_i, x_j) \leq 0 . \qquad (2.3)$$

Suppose $P, Q \in \mathcal{M}_1(\mathcal{X})$ with finite first moments. When $(\mathcal{X}, d)$ has negative type,

$$\int_{\mathcal{X}} d(x_1, x_2)\, d(P - Q)^2(x_1, x_2) \leq 0. \qquad (2.4)$$

We say that $(\mathcal{X}, d)$ has strong negative type if it has negative type and the equality in (2.4) holds only when $P = Q$.

If $(\mathcal{X}, d)$ has strong negative type, then $ED_d(X, Y)$ completely characterizes the homogeneity of the distributions of X and Y (see [26] and [39] for detailed discussions). This quantification of homogeneity of distributions lends itself for reasonable use in one-sample goodness-of-fit testing and two sample testing for equality of distributions.

On the machine learning side, [19] proposed a kernel-based metric, namely maximum mean discrepancy (MMD), to conduct two-sample testing for equality of distributions. We provide some background before introducing MMD.

Definition 2.2 (RKHS). Let $\mathcal{H}$ be a Hilbert space of real valued functions defined on some space $\mathcal{X}$. A bivariate function $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a reproducing kernel of $\mathcal{H}$ if:

1. $\forall x \in \mathcal{X}, \mathcal{K}(\cdot, x) \in \mathcal{H}$
2. $\forall x \in \mathcal{X}, \forall f \in \mathcal{H}, \langle f, \mathcal{K}(\cdot, x) \rangle_{\mathcal{H}} = f(x)$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product associated with $\mathcal{H}$. If $\mathcal{H}$ has a reproducing kernel, it is said to be a reproducing kernel Hilbert space (RKHS).

By Moore-Aronszajn theorem, for every positive definite function (also called a kernel) $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, there is an associated RKHS $\mathcal{H}_{\mathcal{K}}$ with the reproducing kernel $\mathcal{K}$. The map $\Pi : \mathcal{M}_1(\mathcal{X}) \to \mathcal{H}_{\mathcal{K}}$, defined as $\Pi(P) = \int_{\mathcal{X}} \mathcal{K}(\cdot, x) \, dP(x)$ for $P \in \mathcal{M}_1(\mathcal{X})$ is called the mean embedding function associated with $\mathcal{K}$. A kernel $\mathcal{K}$ is said to be characteristic to $\mathcal{M}_1(\mathcal{X})$ if the map Π associated with $\mathcal{K}$ is injective. Suppose $\mathcal{K}$ is a characteristic kernel on $\mathcal{X}$. Then the MMD between $X \sim P_X$ and $Y \sim P_Y$, where $P_X, P_Y \in \mathcal{M}_1(\mathcal{X}) \cap \mathcal{M}_{\mathcal{K}}^{1/2}(\mathcal{X})$ is defined as

$$MMD_{\mathcal{K}}(X, Y) \; = \; \| \Pi(P_X) - \Pi(P_Y) \|_{\mathcal{H}_{\mathcal{K}}} . \tag{2.5}$$

By virtue of $\mathcal{K}$ being a characteristic kernel, $MMD_{\mathcal{K}}(X, Y) = 0$ if and only if $X \stackrel{d}{=} Y$. Lemma 6 in [19] shows that the squared MMD can be equivalently expressed as

$$MMD_{\mathcal{K}}^2(X, Y) \; = \; \mathbb{E}\,\mathcal{K}(X, X') + \mathbb{E}\,\mathcal{K}(Y, Y') - 2\,\mathbb{E}\,\mathcal{K}(X, Y) . \tag{2.6}$$

Theorem 22 in [39] establishes the equivalence between (generalized) energy distance and MMD. Following is the definition of a kernel induced by a distance metric (refer to Section 4.1 in [39] for more details).

Definition 2.3 (Distance-induced kernel and kernel-induced distance). Let $(\mathcal{X}, d)$ be a metric space of negative type and $x_0 \in \mathcal{X}$. Denote $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ as

$$\mathcal{K}(x, x') \; = \; \frac{1}{2} \left\{ d(x, x_0) + d(x', x_0) - d(x, x') \right\} . \tag{2.7}$$

The kernel $\mathcal{K}$ is positive definite if and only if $(\mathcal{X}, d)$ has negative type, and thus $\mathcal{K}$ is a valid kernel on $\mathcal{X}$ whenever d is a metric of negative type. The kernel $\mathcal{K}$ defined in (2.7) is said to be the distance-induced kernel induced by d and centered at x_0. One the other hand, the distance d can be generated by the kernel $\mathcal{K}$ through

$$d(x, x') = \mathcal{K}(x, x) + \mathcal{K}(x', x') - 2\mathcal{K}(x, x'). \tag{2.8}$$

Proposition 29 in [39] establishes that the distance-induced kernel $\mathcal{K}$ induced by d is characteristic to $\mathcal{M}_1(\mathcal{X}) \cap \mathcal{M}_{\mathcal{K}}^1(\mathcal{X})$ if and only if $(\mathcal{X}, d)$ has strong negative type. Therefore, MMD can be viewed as a special case of the generalized energy distance in (2.2) with d being the metric induced by a characteristic kernel.

Suppose $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^m$ are i.i.d samples of X and Y respectively. A U-statistic type estimator of $E_d(X, Y)$ is defined as

$$
\begin{aligned}
E_{n,m}(X, Y) \ = \ & \frac{2}{nm} \sum_{k=1}^{n} \sum_{l=1}^{m} d(X_k, Y_l) - \frac{1}{n(n-1)} \sum_{k \neq l}^{n} d(X_k, X_l) \\
& - \frac{1}{m(m-1)} \sum_{k \neq l}^{m} d(Y_k, Y_l) .
\end{aligned}
\tag{2.9}
$$

2.2. Distance covariance and HSIC

Distance covariance (dCov) was first introduced in the seminal paper by [47] to quantify the dependence between two random vectors of arbitrary (fixed) dimensions. Consider two random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ with $\mathbb{E}\|X\|_p < \infty$ and $\mathbb{E}\|Y\|_q < \infty$. The Euclidean dCov between X and Y is defined as the positive square root of

$$
dCov^2(X, Y) = \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \frac{|f_{X,Y}(t, s) - f_X(t) f_Y(s)|^2}{\|t\|_p^{1+p} \|s\|_q^{1+q}} dt ds,
$$

where f_X, f_Y and $f_{X,Y}$ are the individual and joint characteristic functions of X and Y respectively, and, $c_p = \pi^{(1+p)/2} / \Gamma((1+p)/2)$ is a constant with $\Gamma(\cdot)$ being the complete gamma function.

The key feature of dCov is that it completely characterizes independence between two random vectors of arbitrary dimensions, or in other words $dCov(X, Y) = 0$ if and only if $X \perp\!\!\!\perp Y$. According to Remark 3 in [47], dCov can be equivalently expressed as

$$
\begin{aligned}
dCov^2(X, Y) \ = \ & \mathbb{E}\|X - X'\|_p \|Y - Y'\|_q + \mathbb{E}\|X - X'\|_p \mathbb{E}\|Y - Y'\|_q \\
& - 2\,\mathbb{E}\|X - X'\|_p \|Y - Y''\|_q.
\end{aligned}
\tag{2.10}
$$

Lyons (2013) extends the notion of dCov from Euclidean spaces to general metric spaces. For arbitrary metric spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$, the generalized dCov between $X \sim P_X \in \mathcal{M}_1(\mathcal{X}) \cap \mathcal{M}_{d_{\mathcal{X}}}^1(\mathcal{X})$ and $Y \sim P_Y \in \mathcal{M}_1(\mathcal{Y}) \cap \mathcal{M}_{d_{\mathcal{Y}}}^1(\mathcal{Y})$ is defined as

$$
\begin{aligned}
D_{d_{\mathcal{X}}, d_{\mathcal{Y}}}^2(X, Y) \ = \ & \mathbb{E}\, d_{\mathcal{X}}(X, X') d_{\mathcal{Y}}(Y, Y') + \mathbb{E}\, d_{\mathcal{X}}(X, X')\, \mathbb{E}\, d_{\mathcal{Y}}(Y, Y') \\
& - 2\, \mathbb{E}\, d_{\mathcal{X}}(X, X') d_{\mathcal{Y}}(Y, Y'').
\end{aligned}
\tag{2.11}
$$

Theorem 3.11 in [26] shows that if $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Y}, d_{\mathcal{Y}})$ are both metric spaces of strong negative type, then $D_{d_{\mathcal{X}}, d_{\mathcal{Y}}}(X, Y) = 0$ if and only if $X \perp\!\!\!\perp Y$. In other words, the complete characterization of independence by dCov holds true for

any metric spaces of strong negative type. According to Theorem 3.16 in [26], every separable Hilbert space is of strong negative type. As Euclidean spaces are separable Hilbert spaces, the characterization of independence by dCov between two random vectors in $(\mathbb{R}^p, \| \cdot \|_p)$ and $(\mathbb{R}^q, \| \cdot \|_q)$ is just a special case.

Hilbert-Schmidt Independence Criterion (HSIC) was introduced as a kernel-based independence measure by [17, 18]. Suppose $\mathcal{X}$ and $\mathcal{Y}$ are arbitrary topological spaces, $\mathcal{K}_\mathcal{X}$ and $\mathcal{K}_\mathcal{Y}$ are characteristic kernels on $\mathcal{X}$ and $\mathcal{Y}$ with the respective RKHSs $\mathcal{H}_{\mathcal{K}_\mathcal{X}}$ and $\mathcal{H}_{\mathcal{K}_\mathcal{Y}}$. Let $\mathcal{K} = \mathcal{K}_\mathcal{X} \otimes \mathcal{K}_\mathcal{Y}$ be the tensor product of the kernels $\mathcal{K}_\mathcal{X}$ and $\mathcal{K}_\mathcal{Y}$, and, $\mathcal{H}_\mathcal{K}$ be the tensor product of the RKHSs $\mathcal{H}_{\mathcal{K}_\mathcal{X}}$ and $\mathcal{H}_{\mathcal{K}_\mathcal{Y}}$. The HSIC between $X \sim P_X \in \mathcal{M}_1(\mathcal{X}) \cap \mathcal{M}_\mathcal{K}^{1/2}(\mathcal{X})$ and $Y \sim P_Y \in \mathcal{M}_1(\mathcal{Y}) \cap \mathcal{M}_\mathcal{K}^{1/2}(\mathcal{Y})$ is defined as

$$HSIC_{\mathcal{K}_\mathcal{X}, \mathcal{K}_\mathcal{Y}}(X, Y) = \| \Pi(P_{XY}) - \Pi(P_X P_Y) \|_{\mathcal{H}_\mathcal{K}}, \qquad (2.12)$$

where P_{XY} denotes the joint probability distribution of X and Y. The HSIC between X and Y is essentially the MMD between the joint distribution P_{XY} and the product of the marginals P_X and P_Y. Clearly, $HSIC_{\mathcal{K}_\mathcal{X}, \mathcal{K}_\mathcal{Y}}(X, Y) = 0$ if and only if $X \perp\!\!\!\perp Y$. [17] shows that the squared HSIC can be equivalently expressed as

$$\begin{aligned} HSIC_{\mathcal{K}_\mathcal{X}, \mathcal{K}_\mathcal{Y}}^2(X, Y) = \ & \mathbb{E}\mathcal{K}_\mathcal{X}(X, X')\mathcal{K}_\mathcal{Y}(Y, Y') + \mathbb{E}\mathcal{K}_\mathcal{X}(X, X')\, \mathbb{E}\mathcal{K}_\mathcal{Y}(Y, Y') \\ & - 2\, \mathbb{E}\mathcal{K}_\mathcal{X}(X, X')\mathcal{K}_\mathcal{Y}(Y, Y''). \end{aligned}$$

$$(2.13)$$

Theorem 24 in [39] establishes the equivalence between the generalized dCov and HSIC.

For an observed random sample $(X_i, Y_i)_{i=1}^n$ from the joint distribution of X and Y, a U-statistic type estimator of the generalized dCov in (2.11) can be defined as

$$\widetilde{D}_{n\,;\,d_\mathcal{X}, d_\mathcal{Y}}^2(X, Y) = (\tilde{A} \cdot \tilde{B}) = \frac{1}{n(n-3)} \sum_{k \neq l} \tilde{a}_{kl} \tilde{b}_{kl}\,, \qquad (2.14)$$

where $\tilde{A}, \tilde{B}$ are the $\mathcal{U}$-centered versions (see (1.1)) of $A = \left(d_\mathcal{X}(X_k, X_l)\right)_{k,l=1}^n$ and $B = \left(d_\mathcal{Y}(Y_k, Y_l)\right)_{k,l=1}^n$, respectively. We denote $\widetilde{D}_{n\,;\,d_\mathcal{X}, d_\mathcal{Y}}^2(X, Y)$ by $dCov_n^2(X, Y)$ when $d_\mathcal{X}$ and $d_\mathcal{Y}$ are Euclidean distances.

3. New distance for Euclidean space

We introduce a family of distances for Euclidean space, which shall play a central role in the subsequent developments. For $x \in \mathbb{R}^{\tilde{p}}$, we partition x into p subvectors or groups, namely $x = (x_{(1)}, \ldots, x_{(p)})$, where $x_{(i)} \in \mathbb{R}^{d_i}$ with $\sum_{i=1}^p d_i = \tilde{p}$. Let ρ_i be a metric or semimetric (see for example Definition 1 in [39]) defined on $\mathbb{R}^{d_i}$ for $1 \leq i \leq p$. We define a family of distances for $\mathbb{R}^{\tilde{p}}$ as

$$K_{\mathbf{d}}(x, x') := \sqrt{\rho_1(x_{(1)}, x'_{(1)}) + \ldots + \rho_p(x_{(p)}, x'_{(p)})}\,, \qquad (3.1)$$

where $x, x' \in \mathbb{R}^{\tilde{p}}$ with $x = (x_{(1)}, \ldots, x_{(p)})$ and $x' = (x'_{(1)}, \ldots, x'_{(p)})$, and $\mathbf{d} = (d_1, d_2, \ldots, d_p)$ with $d_i \in \mathbb{Z}_+$ and $\sum_{i=1}^p d_i = \tilde{p}$.

Proposition 3.1. *Suppose each ρ_i is a metric of strong negative type on $\mathbb{R}^{d_i}$. Then $(\mathbb{R}^{\tilde{p}}, K_{\mathbf{d}})$ satisfies the following two properties:*

1. *$K_{\mathbf{d}} : \mathbb{R}^{\tilde{p}} \times \mathbb{R}^{\tilde{p}} \to [0, \infty)$ is a valid metric on $\mathbb{R}^{\tilde{p}}$;*
2. *$(\mathbb{R}^{\tilde{p}}, K_{\mathbf{d}})$ has strong negative type.*

In a special case, suppose ρ_i is the Euclidean distance on $\mathbb{R}^{d_i}$. By Theorem 3.16 in [26], $(\mathbb{R}^{d_i}, \rho_i)$ is a separable Hilbert space, and hence has strong negative type. Then the Euclidean space equipped with the metric

$$K_{\mathbf{d}}(x, x') = \sqrt{\|x_{(1)} - x'_{(1)}\| + \ldots + \|x_{(p)} - x'_{(p)}\|} \,. \tag{3.2}$$

is of strong negative type. Further, if all the components $x_{(i)}$ are unidimensional, i.e., $d_i = 1$ for $1 \le i \le p$, then the metric boils down to

$$K_{\mathbf{d}}(x, x') = \|x - x'\|_1^{1/2} = \sqrt{\sum_{j=1}^p |x_j - x'_j|} \,, \tag{3.3}$$

where $\|x\|_1 = \sum_{j=1}^p |x_j|$ is the l_1 or the absolute norm on $\mathbb{R}^p$. If

$$\rho_i(x_{(i)}, x'_{(i)}) = \|x_{(i)} - x'_{(i)}\|^2, \quad 1 \le i \le p, \tag{3.4}$$

then $K_{\mathbf{d}}$ reduces to the usual Euclidean distance. We shall unify the analysis of our new metrics with the classical metrics by considering $K_{\mathbf{d}}$ which is defined in (3.1) with

S1 each ρ_i being a metric of strong negative type on $\mathbb{R}^{d_i}$;
S2 each ρ_i being a semimetric defined in (3.4).

The first case corresponds to the newly proposed metrics while the second case leads to the classical metrics based on the usual Euclidean distance.

Remark 3.1. *The square root in the definition of $K_{\mathbf{d}}$ in (3.1) is important not only to affirm that $(\mathbb{R}^{\tilde{p}}, K_{\mathbf{d}})$ has strong negative type, but also in deriving Proposition 4.1 via some suitable Taylor's expansions. Proposition 4.1 in turn plays a key role in the subsequent developments of the paper.*

Remarks 3.2 and 3.3 provide two different ways of generalizing the class in (3.1). To be focused, our analysis below shall only concern about the distances defined in (3.1). In the numerical studies in Section 6, we consider ρ_i to be the Euclidean distance and the distances induced by the Laplace and Gaussian kernels (see Definition 2.3) which are of strong negative type on $\mathbb{R}^{d_i}$ for $1 \le i \le p$.

Remark 3.2. *A more general family of distances can be defined as*

$$K_{\mathbf{d},r}(x, x') = \left(\rho_1(x_{(1)}, x'_{(1)}) + \cdots + \rho_p(x_{(p)}, x'_{(p)}) \right)^r, \quad 0 < r < 1.$$

According to Remark 3.19 of [26], the space $(\mathbb{R}^{\tilde{p}}, K_{\mathbf{d},r})$ is of strong negative type. The proposed distance is a special case with $r = 1/2$.

Remark 3.3. *Based on the proposed distance, one can construct the generalized Gaussian and Laplacian kernels as*

$$f(K_{\mathbf{d}}(x,x')/\gamma) = \begin{cases} \exp(-K_{\mathbf{d}}^2(x,x')/\gamma^2), & f(x) = \exp(-x^2) \text{ for Gaussian kernel,} \\ \exp(-K_{\mathbf{d}}(x,x')/\gamma), & f(x) = \exp(-x) \text{ for Laplacian kernel.} \end{cases}$$

If $K_{\mathbf{d}}$ is translation invariant, then by Theorem 9 in [42] it can be verified that $f(K_{\mathbf{d}}(x,x')/\gamma)$ is a characteristic kernel on $\mathbb{R}^{\tilde{p}}$. As a consequence, the Euclidean space equipped with the distance

$$K_{\mathbf{d},f}(x,x') = f(K_{\mathbf{d}}(x,x)/\gamma) + f(K_{\mathbf{d}}(x',x')/\gamma) - 2f(K_{\mathbf{d}}(x,x')/\gamma)$$

is of strong negative type.

Remark 3.4. *In Sections 4 and 5 we develop new classes of homogeneity and dependence metrics to quantify the pairwise homogeneity of distributions or the pairwise non-linear dependence of the low-dimensional groups. A natural question to arise in this regard is how to partition the random vectors optimally in practice. We present some real data examples in Section 6.3 of the main paper where all the group sizes have been considered to be one (as a special case of the general theory proposed in this paper), and an additional real data example in Section C of the appendix where the data admits some natural grouping. We believe this partitioning can be very much problem specific and may require subject knowledge. It would be intriguing to develop an algorithm to find the optimal groups using the data and perhaps some auxiliary information. One potential way to do grouping might be to estimate and exploit the underlying directed acyclic graph (DAG) structure representing the underlying distributions of the random vectors. Further research along this line is on the way.*

4. Homogeneity metrics

Consider $X, Y \in \mathbb{R}^{\tilde{p}}$. Suppose X and Y can be partitioned into p sub-vectors or groups, viz. $X = \left(X_{(1)}, X_{(2)}, \ldots, X_{(p)}\right)$ and $Y = \left(Y_{(1)}, Y_{(2)}, \ldots, Y_{(p)}\right)$, where the groups $X_{(i)}$ and $Y_{(i)}$ are d_i dimensional, $1 \leq i \leq p$, and p might be fixed or growing. We assume that $X_{(i)}$ and $Y_{(i)}$'s are finite (low) dimensional vectors, i.e., $\{d_i\}_{i=1}^p$ is a bounded sequence. Clearly $\tilde{p} = \sum_{i=1}^p d_i = O(p)$. Denote the mean vectors and the covariance matrices of X and Y by μ_X and μ_Y, and, Σ_X and Σ_Y, respectively. We propose the following class of metrics $\mathcal{E}$ to quantify the homogeneity of the distributions of X and Y:

$$\mathcal{E}(X,Y) = 2\,\mathbb{E}\,K_{\mathbf{d}}(X,Y) - \mathbb{E}\,K_{\mathbf{d}}(X,X') - \mathbb{E}\,K_{\mathbf{d}}(Y,Y'), \qquad (4.1)$$

with $\mathbf{d} = (d_1, \ldots, d_p)$. We shall drop the subscript $\mathbf{d}$ below for the ease of notation.

Assumption 4.1. *Assume that* $\sup_{1 \le i \le p} \mathbb{E}\, \rho_i^{1/2}(X_{(i)}, 0_{d_i}) < \infty$ *and* $\sup_{1 \le i \le p} \mathbb{E}\, \rho_i^{1/2}(Y_{(i)}, 0_{d_i}) < \infty$.

Under Assumption 4.1, $\mathcal{E}$ is finite. In Section A.1 of the appendix we illustrate that in the low-dimensional setting, $\mathcal{E}(X, Y)$ completely characterizes the homogeneity of the distributions of X and Y.

Consider i.i.d. samples $\{X_k\}_{k=1}^n$ and $\{Y_l\}_{l=1}^m$ from the respective distributions of X and $Y \in \mathbb{R}^{\tilde{p}}$, where $X_k = (X_{k(1)}, \ldots, X_{k(p)})$, $Y_l = (Y_{l(1)}, \ldots, Y_{l(p)})$ for $1 \le k \le n$, $1 \le l \le m$ and $X_{k(i)}, Y_{l(i)} \in \mathbb{R}^{d_i}$. We propose an unbiased U-statistic type estimator $\mathcal{E}_{n,m}(X, Y)$ of $\mathcal{E}(X, Y)$ as in equation (2.9) with d being the new metric K. We refer the reader to Section A.1 of the appendix, where we show that $\mathcal{E}_{n,m}(X, Y)$ essentially inherits all the nice properties of the U-statistic type estimator of generalized energy distance and MMD.

We define the following quantities which will play an important role in our subsequent analysis:

$$\tau_X^2 = \mathbb{E}\, K(X, X')^2, \quad \tau_Y^2 = \mathbb{E}\, K(Y, Y')^2, \quad \tau^2 = \mathbb{E}\, K(X, Y)^2. \qquad (4.2)$$

In Case S2 (i.e., when K is the Euclidean distance), we have

$$\tau_X^2 = 2\mathrm{tr}\Sigma_X, \quad \tau_Y^2 = 2\mathrm{tr}\Sigma_Y, \quad \tau^2 = \mathrm{tr}\Sigma_X + \mathrm{tr}\Sigma_Y + \|\mu_X - \mu_Y\|^2. \qquad (4.3)$$

Under the null hypothesis $H_0 : X \stackrel{d}{=} Y$, it is clear that $\tau_X^2 = \tau_Y^2 = \tau^2$.

In the subsequent discussion, we study the asymptotic behavior of $\mathcal{E}$ in the high-dimensional framework, i.e., when p grows to ∞ with fixed n and m (discussed in Subsection 4.1) and when n and m grow to ∞ as well (discussed in Subsection B.1 in the appendix). We point out some limitations of the test for homogeneity of distributions in the high-dimensional setup based on the usual Euclidean energy distance. Consequently we propose a test based on the proposed metric and justify its consistency for growing dimension.

4.1. High dimension low sample size (HDLSS)

In this subsection, we study the asymptotic behavior of the Euclidean energy distance and our proposed metric $\mathcal{E}$ when the dimension grows to infinity while the sample sizes n and m are held fixed. We make the following moment assumption.

Assumption 4.2. *There exist constants* a, a', a'', A, A', A'' *such that uniformly over p,*

$$0 < a \le \inf_{1 \le i \le p} \mathbb{E}\, \rho_i(X_{(i)}, X'_{(i)}) \le \sup_{1 \le i \le p} \mathbb{E}\, \rho_i(X_{(i)}, X'_{(i)}) \le A < \infty,$$

$$0 < a' \le \inf_{1 \le i \le p} \mathbb{E}\, \rho_i(Y_{(i)}, Y'_{(i)}) \le \sup_{1 \le i \le p} \mathbb{E}\, \rho_i(Y_{(i)}, Y'_{(i)}) \le A' < \infty,$$

$$0 < a'' \le \inf_{1 \le i \le p} \mathbb{E}\, \rho_i(X_{(i)}, Y_{(i)}) \le \sup_{1 \le i \le p} \mathbb{E}\, \rho_i(X_{(i)}, Y_{(i)}) \le A'' < \infty.$$

Under Assumption 4.2, it is not hard to see that $\tau_X, \tau_Y, \tau \asymp p^{1/2}$. The proposition below provides an expansion for K evaluated at random samples.

Proposition 4.1. *Under Assumption 4.2, we have*

$$\frac{K(X, X')}{\tau_X} = 1 + \frac{1}{2}L_X(X, X') + R_X(X, X'), \tag{4.4}$$

$$\frac{K(Y, Y')}{\tau_Y} = 1 + \frac{1}{2}L_Y(Y, Y') + R_Y(Y, Y'), \tag{4.5}$$

and

$$\frac{K(X, Y)}{\tau} = 1 + \frac{1}{2}L(X, Y) + R(X, Y), \tag{4.6}$$

where $L_X(X, X') := \frac{K^2(X,X')-\tau_X^2}{\tau_X^2}$, $L_Y(Y, Y') := \frac{K^2(Y,Y')-\tau_Y^2}{\tau_Y^2}$, $L(X, Y) := \frac{K^2(X,Y)-\tau^2}{\tau^2}$, *and* $R_X(X, X'), R_Y(Y, Y'), R(X, Y)$ *are the remainder terms. In addition, if* $L_X(X, X'), L_Y(Y, Y')$ *and* $L(X, Y)$ *are* $o_p(1)$ *random variables as* $p \to \infty$, *then* $R_X(X, X') = O_p\left(L_X^2(X, X')\right)$, $R_Y(Y, Y') = O_p\left(L_Y^2(Y, Y')\right)$ *and* $R(X, Y) = O_p\left(L^2(X, Y)\right)$.

Henceforth we will drop the subscripts X and Y from L_X, L_Y, R_X and R_Y for notational convenience. Theorem 4.1 and Lemma 4.1 below provide insights into the behavior of $\mathcal{E}(X, Y)$ in the high-dimensional framework.

Assumption 4.3. *Assume that* $L(X, Y) = O_p(a_p)$, $L(X, X') = O_p(b_p)$ *and* $L(Y, Y') = O_p(c_p)$, *where* a_p, b_p, c_p *are positive real sequences satisfying* $a_p = o(1)$, $b_p = o(1)$, $c_p = o(1)$ *and* $\tau a_p^2 + \tau_X b_p^2 + \tau_Y c_p^2 = o(1)$.

Remark 4.1. *To illustrate Assumption 4.3, we observe that under assumption 4.2 we can write*

$$\begin{aligned} var\left(L(X, X')\right) &= O\left(\frac{1}{p^2}\right) \sum_{i,j=1}^{p} cov\left(\rho_i(X_{(i)}, X'_{(i)}), \rho_j(X_{(j)}, X'_{(j)})\right) \\ &= O\left(\frac{1}{p^2}\right) \sum_{i,j=1}^{p} cov\left(Z_i, Z_j\right), \end{aligned}$$

where $Z_i := \rho_i(X_{(i)}, X'_{(i)})$ *for* $1 \le i \le p$. *Assume that* $\sup_{1 \le i \le p} \mathbb{E}\, \rho_i^2(X_{(i)}, 0_{d_i}) < \infty$, *which implies* $\sup_{1 \le i \le p} \mathbb{E}\, Z_i^2 < \infty$. *Under certain strong mixing conditions or in general certain weak dependence assumptions, it is not hard to see that* $\sum_{i,j=1}^{p} cov(Z_i, Z_j) = O(p)$ *as* $p \to \infty$ *(see for example Theorem 1.2 in [35] or Theorem 1 in [13]). Therefore we have* $var\left(L(X, X')\right) = O(\frac{1}{p})$ *and hence by Chebyshev's inequality, we have* $L(X, X') = O_p(\frac{1}{\sqrt{p}})$. *We refer the reader to Remark 2.1.1 in [57] for illustrations when each* ρ_i *is the squared Euclidean distance.*

Theorem 4.1. *Suppose Assumptions [4.2](#) and [4.3](#) hold. Further assume that the following three sequences*

$$\left\{ \frac{\sqrt{p}L^2(X,Y)}{1+L(X,Y)} \right\}, \quad \left\{ \frac{\sqrt{p}L^2(X,X')}{1+L(X,X')} \right\}, \quad \left\{ \frac{\sqrt{p}L^2(Y,Y')}{1+L(Y,Y')} \right\}$$

indexed by p are all uniformly integrable. Then we have

$$\mathcal{E}(X,Y) \;=\; 2\tau - \tau_X - \tau_Y \;+\; o(1). \tag{4.7}$$

Remark 4.2. *Remark [D.1](#) in the appendix provides some illustrations on certain sufficient conditions under which $\{\sqrt{p}L^2(X,Y)/(1+L(X,Y))\}$, $\{\sqrt{p}L^2(X,X')/(1+L(X,X'))\}$ and $\{\sqrt{p}L^2(Y,Y')/(1+L(Y,Y'))\}$ are uniformly integrable.*

Remark 4.3. *To illustrate that the leading term in equation ([4.7](#)) indeed gives a close approximation of the population $\mathcal{E}(X,Y)$, we consider the special case when K is the Euclidean distance. Suppose $X \sim N_p(0,I_p)$ and $Y = X+W$ where $W \sim N_p(0,I_p)$ with $W \perp\!\!\!\perp X$. Clearly from ([4.3](#)) we have $\tau_X^2 = 2p$, $\tau_Y^2 = 4p$ and $\tau^2 = 3p$. We simulate large samples of sizes $m = n = 5000$ from the distributions of X and Y for $p = 20, 40, 60, 80$ and 100. The large sample sizes are to ensure that the U-statistic type estimator of $\mathcal{E}(X,Y)$ gives a very close approximation of the population $\mathcal{E}(X,Y)$. In Table [1](#) we list the ratio between $\mathcal{E}(X,Y)$ and the leading term in ([4.7](#)) for the different values of p, which turn out to be very close to 1, demonstrating that the leading term in ([4.7](#)) indeed approximates $\mathcal{E}(X,Y)$ reasonably well.*

TABLE 1
Ratio of $\mathcal{E}(X,Y)$ and the leading term in ([4.7](#)) for different values of p.

$p = 20$	$p = 40$	$p = 60$	$p = 80$	$p = 100$
0.995	0.987	0.992	0.997	0.983

Lemma 4.1. *Assume $\tau, \tau_X, \tau_Y < \infty$. We have*

1. *In Case S1, $2\tau - \tau_X - \tau_Y = 0$ if and only if $X_{(i)} \stackrel{d}{=} Y_{(i)}$ for $i \in \{1,\dots,p\}$;*
2. *In Case S2, $2\tau - \tau_X - \tau_Y = 0$ if and only if $\mu_X = \mu_Y$ and $tr\Sigma_X = tr\Sigma_Y$.*

It is to be noted that assuming $\tau, \tau_X, \tau_Y < \infty$ does not contradict with the growth rate $\tau, \tau_X, \tau_Y = O(p^{1/2})$. Clearly under H_0, $2\tau - \tau_X - \tau_Y = 0$ irrespective of the choice of K. In view of Lemma [4.1](#) and Theorem [4.1](#), in Case S2, the leading term of $\mathcal{E}(X,Y)$ becomes zero if and only if $\mu_X = \mu_Y$ and $tr\Sigma_X = tr\Sigma_Y$. In other words, when dimension grows high, the Euclidean energy distance can only capture the equality of the means and the first spectral means, whereas our proposed metric captures the pairwise homogeneity of the low dimensional marginal distributions of $X_{(i)}$ and $Y_{(i)}$. Clearly $X_{(i)} \stackrel{d}{=} Y_{(i)}$ for $1 \leq i \leq p$ implies $\mu_X = \mu_Y$ and $tr\Sigma_X = tr\Sigma_Y$. Thus the proposed metric can capture a wider range of inhomogeneity of distributions than the Euclidean energy distance.

Define

$$d_{kl}(i) := \rho_i(X_{k(i)}, Y_{l(i)}) - \mathbb{E}\left[\rho_i(X_{k(i)}, Y_{l(i)})|X_{k(i)}\right] - \mathbb{E}\left[\rho_i(X_{k(i)}, Y_{l(i)})|Y_{l(i)}\right]$$
$$+ \mathbb{E}\left[\rho_i(X_{k(i)}, Y_{l(i)})\right],$$

as the double-centered distance between $X_{k(i)}$ and $Y_{l(i)}$ for $1 \leq i \leq p$, $1 \leq k \leq n$ and $1 \leq l \leq m$. Similarly define $d_{kl}^X(i)$ and $d_{kl}^Y(i)$ as the double-centered distances between $X_{k(i)}$ and $X_{l(i)}$ for $1 \leq k \neq l \leq n$, and, $Y_{k(i)}$ and $Y_{l(i)}$ for $1 \leq k \neq l \leq m$, respectively. Further define $H(X_k, Y_l) := \frac{1}{\tau}\sum_{i=1}^p d_{kl}(i)$ for $1 \leq k \leq n$, $1 \leq l \leq m$, $H(X_k, X_l) := \frac{1}{\tau_X}\sum_{i=1}^p d_{kl}^X(i)$ for $1 \leq k \neq l \leq n$ and $H(Y_k, Y_l)$ in a similar way.

We impose the following conditions to study the asymptotic behavior of the (unbiased) U-statistic type estimator of $\mathcal{E}(X, Y)$ in the HDLSS setup.

Assumption 4.4. *For fixed n and m, as $p \to \infty$,*

$$\begin{pmatrix} H(X_k, Y_l) \\ H(X_s, X_t) \\ H(Y_u, Y_v) \end{pmatrix}_{k,l,\, s<t,\, u<v} \xrightarrow{d} \begin{pmatrix} a_{kl} \\ b_{st} \\ c_{uv} \end{pmatrix}_{k,l,\, s<t,\, u<v},$$

where $\{a_{kl}, b_{st}, c_{uv}\}_{k,l,\, s<t,\, u<v}$ are jointly Gaussian with zero mean. Further we assume that

$$\begin{aligned} var(a_{kl}) &:= \sigma^2 = \lim_{p\to\infty} \mathbb{E}\left[H^2(X_k, Y_l)\right], \\ var(b_{st}) &:= \sigma_X^2 = \lim_{p\to\infty} \mathbb{E}\left[H^2(X_s, X_t)\right], \\ var(c_{uv}) &:= \sigma_Y^2 = \lim_{p\to\infty} \mathbb{E}\left[H^2(Y_u, Y_v)\right]. \end{aligned}$$

$\{a_{kl}, b_{st}, c_{uv}\}_{k,l,\, s<t,\, u<v}$ *are all independent with each other.*

Due to the double-centering property and the independence between the two samples, it is straightforward to verify that $\{H(X_k, Y_l), H(X_s, X_t), H(Y_u, Y_v)\}_{k,l,s<t,u<t}$ are uncorrelated with each other. So it is natural to expect that the limit $\{a_{kl}, b_{st}, c_{uv}\}_{k,l,\, s<t,\, u<v}$ are all independent with each other.

Remark 4.4. *The above multi-dimensional central limit theorem is classic and can be derived under suitable moment and weak dependence assumptions on the components of X and Y, such as mixing or near epoch dependence conditions. We refer the reader to [14] for a review on central limit theorem results under weak dependence assumptions.*

We describe a new two-sample t-test for testing the null hypothesis $H_0 : X \stackrel{d}{=} Y$. The t statistic can be constructed based on either the Euclidean energy distance or the new homogeneity metrics. We show that the t-tests based on different metrics can have strikingly different power behaviors under the HDLSS setup. The major difficulty here is to introduce a consistent and computationally efficient variance estimator. Towards this end, we define a quantity called Cross

Distance Covariance (cdCov) between X and Y, which plays an important role in the construction of the t-test statistic:

$$cdCov_{n,m}^2(X,Y) := \frac{1}{(n-1)(m-1)} \sum_{k=1}^{n} \sum_{l=1}^{m} \widehat{K}(X_k,Y_l)^2,$$

where

$$\widehat{K}(X_k,Y_l) = K(X_k,Y_l) - \frac{1}{n}\sum_{i=1}^{n} K(X_i,Y_l) - \frac{1}{m}\sum_{j=1}^{m} K(X_k,Y_j)$$

$$+ \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} K(X_i,Y_j).$$

Let $v_s := s(s-3)/2$ for $s = m,n$. We introduce the following quantities

$$
\begin{aligned}
m_0 &:= \frac{\sigma^2\,(n-1)(m-1) + \sigma_X^2\,v_n + \sigma_Y^2\,v_m}{(n-1)(m-1)+v_n+v_m}, \\
\sigma_{nm} &:= \sqrt{\frac{\sigma^2}{nm} + \frac{\sigma_X^2}{2n(n-1)} + \frac{\sigma_Y^2}{2m(m-1)}}, \\
a_{nm} &:= \sqrt{\frac{1}{nm} + \frac{1}{2n(n-1)} + \frac{1}{2m(m-1)}}, \\
\Delta &:= \lim_{p\to\infty} 2\tau - \tau_X - \tau_Y,
\end{aligned}
\tag{4.8}
$$

where σ^2, σ_X^2 and σ_Y^2 are defined in Assumption 4.4. Under Assumption 4.5, further define

$$
\begin{aligned}
m_0^* &:= \lim_{m,n\to\infty} m_0 = \frac{2\alpha_0\,\sigma^2 + \sigma_X^2 + \sigma_Y^2\,\alpha_0^2}{2\alpha_0 + 1 + \alpha_0^2}, \\
a_0^* &:= \lim_{m,n\to\infty} \frac{a_{nm}}{\sigma_{nm}} = \Big(\frac{2\alpha_0 + \alpha_0^2 + 1}{2\alpha_0\,\sigma^2 + \alpha_0^2\,\sigma_X^2 + \sigma_Y^2}\Big)^{1/2}.
\end{aligned}
$$

We are now ready to introduce the two-sample t-test

$$T_{n,m} := \frac{\mathcal{E}_{n,m}(X,Y)}{a_{nm}\sqrt{S_{n,m}}},$$

where

$$S_{n,m} := \frac{4(n-1)(m-1)\,cdCov_{n,m}^2(X,Y) + 4v_n\,\widetilde{\mathcal{D}_n^2}(X,X) + 4v_m\,\widetilde{\mathcal{D}_m^2}(Y,Y)}{(n-1)(m-1)+v_n+v_m}$$

is the pool variance estimator with $\widetilde{\mathcal{D}_n^2}(X,X)$ and $\widetilde{\mathcal{D}_m^2}(Y,Y)$ being the unbiased estimators of the (squared) distance variances defined in equation (2.14). It is interesting to note that the variability of the sample generalized energy distance

depends on the distance variances as well as the cdCov. It is also worth mentioning that the computational complexity of the pool variance estimator and thus the t-statistic is linear in p.

To study the asymptotic behavior of the test, we consider the following class of distributions on (X, Y):

$$\mathcal{P} = \Big\{(P_X, P_Y): \ X \sim P_X, \ Y \sim P_Y, \ E[\tau L(X,Y) - \tau_X L(X,X')|X] = o_p(1),$$
$$E[\tau L(X,Y) - \tau_Y L(Y,Y')|Y] = o_p(1)\Big\}.$$

If $P_X = P_Y$ (i.e., under the H_0), it is clear that $(P_X, P_Y) \in \mathcal{P}$ irrespective of the metrics in the definition of L. Suppose $\|X - \mu_X\|^2 - \mathrm{tr}(\Sigma_X) = O_p(\sqrt{p})$ and $\|Y - \mu_Y\|^2 - \mathrm{tr}(\Sigma_Y) = O_p(\sqrt{p})$, which hold under weak dependence assumptions on the components of X and Y. Then in Case S2 (i.e., K is the Euclidean distance), a set of sufficient conditions for $(P_X, P_Y) \in \mathcal{P}$ is given by

$$(\mu_X - \mu_Y)^\top (\Sigma_X + \Sigma_Y)(\mu_X - \mu_Y) = o(p), \quad \tau - \tau_X = o(\sqrt{p}), \quad \tau - \tau_Y = o(\sqrt{p}), \tag{4.9}$$

which suggests that the first two moments of P_X and P_Y are not too far away from each other. In this sense, $\mathcal{P}$ defines a class of local alternative distributions (with respect to the null $H_0: P_X = P_Y$). We now state the main result of this subsection.

Theorem 4.2. *In both Cases S1 and S2, under Assumptions 4.2, 4.3 and 4.4 as $p \to \infty$ with n and m remaining fixed, and further assuming that $(P_X, P_Y) \in \mathcal{P}$, we have*

$$\frac{\mathcal{E}_{n,m}(X,Y) - (2\tau - \tau_X - \tau_Y)}{a_{nm}\sqrt{S_{n,m}}} \xrightarrow{d} \frac{\sigma_{nm}\, Z}{a_{nm}\sqrt{M}},$$

where

$$M \stackrel{d}{=} \frac{\sigma^2\, \chi^2_{(n-1)(m-1)} + \sigma_X^2 \chi^2_{v_n} + \sigma_Y^2 \chi^2_{v_m}}{(n-1)(m-1) + v_n + v_m},$$

$\chi^2_{(n-1)(m-1)}, \chi^2_{v_n}, \chi^2_{v_m}$ *are independent chi-squared random variables, and $Z \sim N(0,1)$. In other words,*

$$T_{n,m} \xrightarrow{d} \frac{\sigma_{nm}\, N(\Delta/\sigma_{nm}, 1)}{a_{nm}\sqrt{M}},$$

where σ_{nm} and a_{nm} are defined in equation (4.8). In particular, under H_0, we have

$$T_{n,m} \xrightarrow{d} t_{(n-1)(m-1)+v_n+v_m}.$$

Based on the asymptotic behavior of $T_{n,m}$ for growing dimensions, we propose a test for H_0 as follows: at level $\alpha \in (0, 1)$, reject H_0 if

$$T_{n,m} > q_{\alpha,(n-1)(m-1)+v_n+v_m}$$

and fail to reject H_0 otherwise, where

$$P(t_{(n-1)(m-1)+v_n+v_m} > q_{\alpha,(n-1)(m-1)+v_n+v_m}) = \alpha.$$

For a fixed real number t, define

$$
\begin{aligned}
\phi_{n,m}(t) &:= \lim_{p\to\infty} P(T_{n,m} \le t) = \mathbb{E}\left[P\left(\frac{\sigma_{nm} N(\Delta/\sigma_{nm}, 1)}{a_{nm}\sqrt{M}} \le t \mid M\right)\right] \\
&= \mathbb{E}\left[\Phi\left(\frac{a_{nm}\sqrt{M}\,t - \Delta}{\sigma_{nm}}\right)\right].
\end{aligned}
$$

$$(4.10)$$

The asymptotic power curve for testing H_0 based on $T_{n,m}$ is given by $1-\phi_{m,n}(t)$. The following proposition gives a large sample approximation of the power curve.

Assumption 4.5. *As $m, n \to \infty$, $m/n \to \alpha_0$ where $\alpha_0 > 0$.*

Proposition 4.2. *Suppose $\Delta = \Delta_0/\sqrt{nm}$ where Δ_0 is a constant with respect to n, m. Then for any bounded real number t as $n, m \to \infty$ and under Assumption 4.5, we have*

$$\lim_{m,n\to\infty} \phi_{n,m}(t) = \Phi\left(a_0^*\sqrt{m_0^*}\,t - \Delta_0^*\right),$$

where

$$\Delta_0^* = \Delta_0 \lim_{m,n\to\infty} \frac{1}{\sigma_{nm}\sqrt{nm}} = \Delta_0\left(\frac{2\alpha_0}{2\sigma^2\,\alpha_0 + \sigma_X^2\,\alpha_0^2 + \sigma_Y^2}\right)^{1/2}.$$

Under the alternative, if $\Delta_0 \to \infty$ as $n, m \to \infty$, we have

$$\lim_{m,n\to\infty}\left\{1 - \phi_{n,m}(q_{\alpha,(n-1)(m-1)+v_n+v_m})\right\} = 1,$$

thereby justifying the consistency of the test.

Remark 4.5. *We first derive the power function $1 - \phi_{n,m}(t)$ under the assumption that n and m are fixed. The main idea behind Proposition 4.2 where we let $n, m \to \infty$ is to see whether we get a reasonably good approximation of power when n, m are large. In a sense we are doing sequential asymptotics, first letting $p \to \infty$ and deriving the power function, and then deriving the leading term by letting $n, m \to \infty$. This is a quite common practice in Econometrics (see for example [31]). The aim is to derive a leading term for the power when n, m are fixed but large. Consider $\Delta = s/\sqrt{nm}$ (as in Proposition 4.2) and set $\sigma^2 = \sigma_X^2 = \sigma_Y^2 = 1$. In Figure 1 below, we plot the exact power (computed from (4.10) with $50,000$ Monte Carlo samples from the distribution of M) with $n = m = 5$ and 10, $t = q_{\alpha,(n-1)(m-1)+v_n+v_m}$ and $\alpha = 0.05$, over different values of s. We overlay the large sample approximation of the power function (given in Proposition 4.2) and observe that the approximation works reasonably well even for small sample sizes. Clearly larger s results in better power and $s = 0$ corresponds to trivial power.*

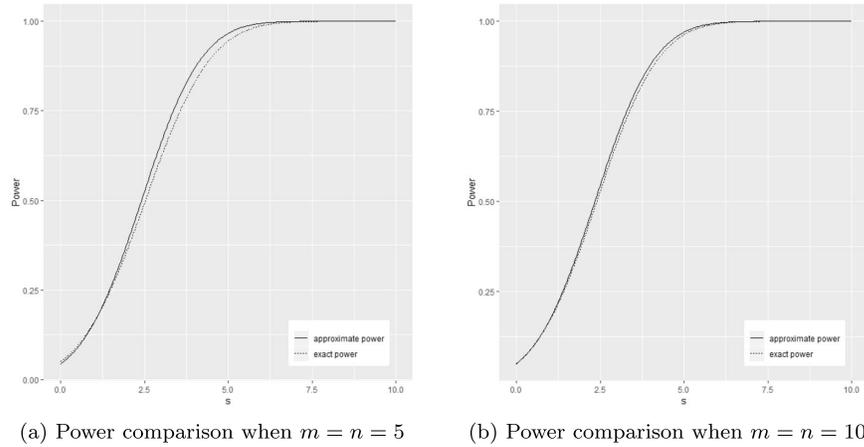(a) Power comparison when $m = n = 5$ (b) Power comparison when $m = n = 10$

FIG 1. *Comparison of exact and approximate power.*

We now discuss the power behavior of $T_{n,m}$ based on the Euclidean energy distance. In Case S2, it can be seen that

$$\sigma_X^2 = \lim_{p \to \infty} \frac{1}{\tau_X^2} \sum_{i,i'=1}^{p} 4 \operatorname{tr} \Sigma_X^2(i, i'), \tag{4.11}$$

where $\Sigma_X^2(i, i')$ is the covariance matrix between $X_{(i)}$ and $X_{(i')}$, and similar expressions for σ_Y^2. In case S2 (i.e., when K is the Euclidean distance), if we further assume $\mu_X = \mu_Y$, it can be verified that

$$\sigma^2 = \lim_{p \to \infty} \frac{1}{\tau^2} \sum_{i,i'=1}^{p} 4 \operatorname{tr}\left(\Sigma_X(i, i') \, \Sigma_Y(i, i')\right). \tag{4.12}$$

Hence in Case S2, under the assumptions that $\mu_X = \mu_Y$, $\operatorname{tr} \Sigma_X = \operatorname{tr} \Sigma_Y$ and $\operatorname{tr} \Sigma_X^2 = \operatorname{tr} \Sigma_Y^2 = \operatorname{tr} \Sigma_X \Sigma_Y$, it can be easily seen from equations (4.3), (4.11) and (4.12) that

$$\tau_X^2 = \tau_Y^2 = \tau^2, \quad \sigma_X^2 = \sigma_Y^2 = \sigma^2, \tag{4.13}$$

which implies that $\Delta_0^* = 0$ in Proposition 4.2. Consider the following class of alternative distributions

$$\begin{aligned} H_A &= \{(P_X, P_Y) : P_X \neq P_Y, \ \mu_X = \mu_Y, \ \operatorname{tr} \Sigma_X = \operatorname{tr} \Sigma_Y, \\ &\quad \operatorname{tr} \Sigma_X^2 = \operatorname{tr} \Sigma_Y^2 = \operatorname{tr} \Sigma_X \Sigma_Y\}. \end{aligned}$$

According to Theorem 4.2, the t-test $T_{n,m}$ based on Euclidean energy distance has trivial power against H_A. In contrast, the t-test based on the proposed metrics has non-trivial power against H_A as long as $\Delta_0^* > 0$.

To summarize our contributions:

- We show that the Euclidean energy distance can only detect the equality of means and the traces of covariance matrices in the high-dimensional setup. To the best of our knowledge, such a limitation of the Euclidean energy distance has not been pointed out in the literature before.
- We propose a new class of homogeneity metrics which completely characterizes homogeneity of two distributions in the low-dimensional setup and has nontrivial power against a broader range of alternatives, or in other words, can detect a wider range of inhomogeneity of two distributions in the high-dimensional setup.
- Grouping allows us to detect homogeneity beyond univariate marginal distributions, as the difference between two univariate marginal distributions is automatically captured by the difference between the marginal distributions of the groups that contain these two univariate components.
- Consequently we construct a high-dimensional two-sample t-test whose computational cost is linear in p. Owing to the pivotal nature of the limiting distribution of the test statistic, no resampling-based inference is needed.

Remark 4.6. *Although the test based on our proposed statistic is asymptotically powerful against the alternative H_A unlike the Euclidean energy distance, it can be verified that it has trivial power against the alternative $H_{A'} = \{(X, Y) : X_{(i)} \overset{d}{=} Y_{(i)}, 1 \leq i \leq p\}$. Thus although it can detect differences between two high-dimensional distributions beyond the first two moments (as a significant improvement to the Euclidean energy distance), it cannot capture differences beyond the equality of the low-dimensional marginal distributions. We conjecture that there might be some intrinsic difficulties for distance and kernel-based metrics to completely characterize the discrepancy between two high-dimensional distributions.*

5. Dependence metrics

In this section, we focus on dependence testing of two random vectors $X \in \mathbb{R}^{\tilde{p}}$ and $Y \in \mathbb{R}^{\tilde{q}}$. Suppose X and Y can be partitioned into p and q groups, viz. $X = (X_{(1)}, X_{(2)}, \ldots, X_{(p)})$ and $Y = (Y_{(1)}, Y_{(2)}, \ldots, Y_{(q)})$, where the components $X_{(i)}$ and $Y_{(j)}$ are d_i and g_j dimensional, respectively, for $1 \leq i \leq p, 1 \leq j \leq q$. Here p, q might be fixed or growing. We assume that $X_{(i)}$ and $Y_{(j)}$'s are finite (low) dimensional vectors, i.e., $\{d_i\}_{i=1}^p$ and $\{g_j\}_{j=1}^q$ are bounded sequences. Clearly, $\tilde{p} = \sum_{i=1}^p d_i = O(p)$ and $\tilde{q} = \sum_{j=1}^q g_j = O(q)$. We define a class of dependence metrics $\mathcal{D}$ between X and Y as the positive square root of

$$
\begin{aligned}
\mathcal{D}^2(X, Y) \ := \ & \mathbb{E}\,K_{\mathbf{d}}(X, X')\,K_{\mathbf{g}}(Y, Y') + \mathbb{E}\,K_{\mathbf{d}}(X, X')\,\mathbb{E}\,K_{\mathbf{g}}(Y, Y') \\
& - 2\,\mathbb{E}\,K_{\mathbf{d}}(X, X')\,K_{\mathbf{g}}(Y, Y''),
\end{aligned}
\tag{5.1}
$$

where $\mathbf{d} = (d_1, \ldots, d_p)$ and $\mathbf{g} = (g_1, \ldots, g_q)$. We drop the subscripts $\mathbf{d}, \mathbf{g}$ of K for notational convenience.

To ensure the existence of $\mathcal{D}$, we make the following assumption.

Assumption 5.1. *Assume that* $\sup_{1 \leq i \leq p} \mathbb{E}\rho_i^{1/2}(X_{(i)}, 0_{d_i}) < \infty$ *and* $\sup_{1 \leq i \leq q} \mathbb{E}\rho_i^{1/2}(Y_{(i)}, 0_{g_i}) < \infty$.

In Section A.2 of the appendix we demonstrate that in the low-dimensional setting, $\mathcal{D}(X, Y)$ completely characterizes independence between X and Y. For an observed random sample $(X_k, Y_k)_{k=1}^n$ from the joint distribution of X and Y, define $D^X := (d_{kl}^X) \in \mathbb{R}^{n \times n}$ with $d_{kl}^X := K(X_k, X_l)$ and $k, l \in \{1, \ldots, n\}$. Define d_{kl}^Y and D^Y in a similar way. With some abuse of notation, we consider the U-statistic type estimator $\widetilde{\mathcal{D}_n^2}(X, Y)$ of $\mathcal{D}^2$ as defined in (2.14) with $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ being $K_{\mathbf{d}}$ and $K_{\mathbf{g}}$ respectively. In Section A.2 of the appendix, we illustrate that $\widetilde{\mathcal{D}_n^2}(X, Y)$ essentially inherits all the nice properties of the U-statistic type estimator of generalized dCov and HSIC.

In the subsequent discussion we study the asymptotic behavior of $\mathcal{D}$ in the high-dimensional framework, i.e., when p and q grow to ∞ with fixed n (discussed in Subsection 5.1) and when n grows to ∞ as well (discussed in Subsection B.2 in the appendix).

5.1. High dimension low sample size (HDLSS)

In this subsection, our goal is to explore the behavior of $\mathcal{D}^2(X, Y)$ and its unbiased U-statistic type estimator in the HDLSS setting where p and q grow to ∞ while the sample size n is held fixed. Denote $\tau_{XY}^2 = \tau_X^2 \tau_Y^2 = \mathbb{E}\, K^2(X, X')\, \mathbb{E}\, K^2(Y, Y')$. We impose the following conditions.

Assumption 5.2. $\mathbb{E}[L^2(X, X')] = O(a_p'^2)$ *and* $\mathbb{E}[L^2(Y, Y')] = O(b_q'^2)$, *where* a_p' *and* b_q' *are positive real sequences satisfying* $a_p' = o(1)$, $b_q' = o(1)$, $\tau_{XY}\, a_p'^2 b_q' = o(1)$ *and* $\tau_{XY}\, a_p' b_q'^2 = o(1)$. *Further assume that* $\mathbb{E}[R^2(X, X')] = O(a_p'^4)$ *and* $\mathbb{E}[R^2(Y, Y')] = O(b_q'^4)$.

Remark 5.1. *We refer the reader to Remark 4.1 in Section 4 for illustrations about some sufficient conditions under which we have* $var(L(X, X')) = \mathbb{E}\, L^2(X, X') = O(\frac{1}{p})$, *and similarly for* $L(Y, Y')$. *Remark D.1 in the appendix illustrates certain sufficient conditions under which* $\mathbb{E}[R^2(X, X')] = O(\frac{1}{p^2})$, *and similarly for* $R(Y, Y')$.

Theorem 5.1. *Under Assumptions 4.2 and 5.2, we have*

$$\mathcal{D}^2(X, Y) = \frac{1}{4\tau_{XY}} \sum_{i=1}^p \sum_{j=1}^q D_{\rho_i, \rho_j}^2(X_{(i)}, Y_{(j)}) + \mathcal{R}, \qquad (5.2)$$

where $\mathcal{R}$ *is the remainder term such that* $\mathcal{R} = O(\tau_{XY}\, a_p'^2 b_q' + \tau_{XY}\, a_p' b_q'^2) = o(1)$.

Theorem 5.1 shows that when dimensions grow high, the population $\mathcal{D}^2(X, Y)$ behaves as an aggregation of group-wise generalized dCov and thus essentially captures group-wise non-linear dependencies between X and Y.

Remark 5.2. *Consider a special case where $d_i = 1$ and $g_j = 1$, and ρ_i and ρ_j are Euclidean distances for all $1 \le i \le p$ and $1 \le j \le q$. Then Theorem 5.1 essentially boils down to*

$$\mathcal{D}^2(X,Y) = \frac{1}{4\tau_{XY}} \sum_{i=1}^{p} \sum_{j=1}^{q} dCov^2(X_i, Y_j) + \mathcal{R} , \qquad (5.3)$$

where $\mathcal{R} = o(1)$. This shows that in a special case (when we have unit group sizes), $\mathcal{D}^2(X,Y)$ essentially behaves as an aggregation of cross-component dCov between X and Y. If $K_{\boldsymbol{d}}$ and $K_{\boldsymbol{g}}$ are Euclidean distances, or in other words if each ρ_i and ρ_j are squared Euclidean distances, then using equation (2.10) it is straightforward to verify that $D^2_{\rho_i,\rho_j}(X_i, Y_j) = 4\, cov^2(X_i, Y_j)$ for all $1 \le i \le p$ and $1 \le j \le q$. Consequently we have

$$\mathcal{D}^2(X,Y) = dCov^2(X,Y) = \frac{1}{\tau_{XY}} \sum_{i=1}^{p} \sum_{j=1}^{q} cov^2(X_i, Y_j) + \mathcal{R}_1 , \qquad (5.4)$$

where $\mathcal{R}_1 = o(1)$, which essentially presents a population version of Theorem 2.1.1 in [57] as a special case of Theorem 5.1. It is to be noted that if each ρ_i and ρ_j are squared Euclidean distances, then the quantities $D^2_{\rho_i,\rho_j}(X_i, Y_j)$ are indeed well-defined in view of the extension of the notion of distance covariance to semimetric spaces of negative type by [39]. Following Proposition 3 in their paper, it is not hard to argue that $\mathbb{R}^p$ equipped with the squared Euclidean distance is indeed a semimetric space of negative type.

Remark 5.3. *To illustrate that the leading term in equation (5.2) indeed gives a close approximation of the population $\mathcal{D}^2(X,Y)$, we consider the special case when $K_{\boldsymbol{d}}$ and $K_{\boldsymbol{g}}$ are Euclidean distances and $p = q$. Suppose $X \sim N_p(0, I_p)$ and $Y = X + W$ where $W \sim N_p(0, I_p)$ with $W \perp\!\!\!\perp X$. Clearly we have $\tau_X^2 = 2p$, $\tau_Y^2 = 4p$, $D^2_{\rho_i,\rho_j}(X_i, Y_j) = 4\, cov^2(X_i, Y_j) = 4$ for all $1 \le i = j \le p$ and $D^2_{\rho_i,\rho_j}(X_i, Y_j) = 0$ for all $1 \le i \ne j \le p$. From Remark 5.2, it is clear that in this case we essentially have $\mathcal{D}^2(X,Y) = dCov^2(X,Y)$. We simulate a large sample of size $n = 5000$ from the distribution of (X,Y) for $p = 20, 40, 60, 80$ and 100. The large sample size is to ensure that the U-statistic type estimator of $\mathcal{D}^2(X,Y)$ (given in (2.14)) gives a very close approximation of the population $\mathcal{D}^2(X,Y)$. We list the ratio between $\mathcal{D}^2(X,Y)$ and the leading term in (5.2) for the different values of p, which turn out to be very close to 1, demonstrating that the leading term in (5.2) indeed approximates $\mathcal{D}^2(X,Y)$ reasonably well.*

TABLE 2
Ratio of $\mathcal{D}^2(X,Y)$ and the leading term in (5.2) for different values of p.

$p = 20$	$p = 40$	$p = 60$	$p = 80$	$p = 100$
0.980	0.993	0.994	0.989	0.997

The following theorem explores the behavior of the population $\mathcal{D}^2(X,Y)$ when p is fixed and q grows to infinity, while the sample size is held fixed.

As far as we know, this asymptotic regime has not been previously considered in the literature. In this case, the Euclidean distance covariance behaves as an aggregation of martingale difference divergences proposed in [43] which measures conditional mean dependence. We summarize below the curse of dimensionality for the Euclidean distance covariance under different asymptotic regimes.

Theorem 5.2. *Under Assumption 4.2 and the assumption that $\mathbb{E}\left[R^2(Y, Y')\right] = O(b_q'^4)$ with $\tau_Y\, b_q'^2 = o(1)$, as $q \to \infty$ with p and n remaining fixed, we have*

$$\mathcal{D}^2(X, Y) \;=\; \frac{1}{2\tau_Y} \sum_{j=1}^{q} D^2_{K_d, \rho_j}(X, Y_{(j)}) \;+\; \mathcal{R},$$

where $\mathcal{R}$ is the remainder term such that $\mathcal{R} = O(\tau_Y\, b_q'^2) = o(1)$.

Remark 5.4. *In particular, when both K_d and K_g are Euclidean distances, we have*

$$\mathcal{D}^2(X, Y) \;=\; dCov^2(X, Y) \;=\; \frac{1}{\tau_Y} \sum_{j=1}^{\tilde{q}} MDD^2(Y_j | X) \;+\; \mathcal{R},$$

where $MDD^2(Y_j | X) = -\mathbb{E}[(Y_j - \mathbb{E}Y_j)(Y_j' - \mathbb{E}Y_j)\|X - X'\|]$ is the martingale difference divergence which completely characterizes the conditional mean dependence of Y_j given X in the sense that $E[Y_j | X] = E[Y_j]$ almost surely if and only if $MDD^2(Y_j | X) = 0$.

To summarize the curse of dimensionality for the Euclidean distance covariance under different asymptotic regimes:

- When $p, q \to \infty$, $dCov^2(X, Y)$ is asymptotically equivalent to $\frac{1}{\tau_{XY}} \sum_{i=1}^{p} \sum_{j=1}^{q} cov^2(X_i, Y_j)$. In other words, as $p, q \to \infty$, the Euclidean dCov can merely capture component-wise linear dependencies between X and Y.
- When $q \to \infty$ with p remaining fixed, $dCov^2(X, Y)$ is asymptotically equivalent to $\frac{1}{\tau_Y} \sum_{j=1}^{\tilde{q}} MDD^2(Y_j | X)$. In other words, the Euclidean dCov can only detect component-wise conditional mean independence of Y given X (which is even weaker than independence).
- When $p \to \infty$ with q remaining fixed, $dCov^2(X, Y)$ is asymptotically equivalent to $\frac{1}{\tau_X} \sum_{i=1}^{\tilde{p}} MDD^2(X_i | Y)$. In other words, the Euclidean dCov can only detect component-wise conditional mean independence of X given Y.

Next we study the asymptotic behavior of the sample version $\widetilde{\mathcal{D}_n^2}(X, Y)$.

Assumption 5.3. *Assume that $L(X, X') = O_p(a_p)$ and $L(Y, Y') = O_p(b_q)$, where a_p and b_q are positive real sequences satisfying $a_p = o(1)$, $b_q = o(1)$, $\tau_{XY}\, a_p^2 b_q = o(1)$ and $\tau_{XY}\, a_p b_q^2 = o(1)$.*

Remark 5.5. *We refer the reader to Remark 4.1 in Section 4 for illustrations about Assumption 5.3.*

Theorem 5.3. *Under Assumptions [4.2] and [5.3], it can be shown that*

$$\widetilde{\mathcal{D}_n^2}(X,Y) = \frac{1}{4\tau_{XY}} \sum_{i=1}^{p} \sum_{j=1}^{q} \widetilde{D_n^2}_{\,;\,\rho_i,\rho_j}(X_{(i)}, Y_{(j)}) + \mathcal{R}_n \,, \tag{5.5}$$

where $X_{(i)}, Y_{(j)}$ are the ith and j^{th} groups of X and Y, respectively, $1 \leq i \leq p$, $1 \leq j \leq q$, and $\mathcal{R}_n$ is the remainder term. Moreover $\mathcal{R}_n = O_p(\tau_{XY}\, a_p^2 b_q + \tau_{XY}\, a_p b_q^2) = o_p(1)$, i.e., $\mathcal{R}_n$ is of smaller order compared to the leading term and hence is asymptotically negligible.

The above theorem generalizes Theorem 2.1.1 in [57] by showing that the leading term of $\widetilde{\mathcal{D}_n^2}(X,Y)$ is the sum of all the group-wise (unbiased) squared sample generalized dCov scaled by τ_{XY}. In other words, in the HDLSS setting, $\widetilde{\mathcal{D}_n^2}(X,Y)$ is asymptotically equivalent to the aggregation of group-wise squared sample generalized dCov. Thus $\widetilde{\mathcal{D}_n^2}(X,Y)$ can quantify group-wise non-linear dependencies between X and Y, going beyond the scope of the usual Euclidean dCov.

Remark 5.6. *Consider a special case where $d_i = 1$ and $g_j = 1$, and ρ_i and ρ_j are Euclidean distances for all $1 \leq i \leq p$ and $1 \leq j \leq q$. Then Theorem [5.3] essentially states that*

$$\widetilde{\mathcal{D}_n^2}(X,Y) = \frac{1}{4\tau_{XY}} \sum_{i=1}^{p} \sum_{j=1}^{q} dCov_n^2(X_i, Y_j) + \mathcal{R}_n \,, \tag{5.6}$$

where $\mathcal{R}_n = o_p(1)$. This demonstrates that in a special case (when we have unit group sizes), $\widetilde{\mathcal{D}_n^2}(X,Y)$ is asymptotically equivalent to the marginal aggregation of cross-component distance covariances proposed by [57] as dimensions grow high. If K_d and K_g are Euclidean distances, then Theorem [5.3] essentially boils down to Theorem 2.1.1 in [57] as a special case.

Remark 5.7. *To illustrate the approximation of $\widetilde{\mathcal{D}_n^2}(X,Y)$ by the aggregation of group-wise squared sample generalized dCov given by Theorem [5.3], we simulated the datasets in Examples [6.4].1, [6.4].2, [6.5].1 and [6.5].2 100 times each with $n = 50$ and $p = q = 50$. For each of the datasets, the difference between $\widehat{\mathcal{D}_n^2}(X,Y)$ and the leading term in the RHS of equation [(5.5)] is smaller than 0.01 100% of the times, which illustrates that the approximation works reasonably well.*

The following theorem illustrates the asymptotic behavior of $\widetilde{\mathcal{D}_n^2}(X,Y)$ when p is fixed and q grows to infinity while the sample size is held fixed. Under this setup, if both K_d and K_g are Euclidean distances, the leading term of $\widetilde{\mathcal{D}_n^2}(X,Y)$ is the sum of the group-wise unbiased U-statistic type estimators of $MDD^2(Y_j|X)$ for $1 \leq j \leq q$, scaled by τ_Y. In other words, the sample Euclidean distance covariance behaves as an aggregation of sample martingale difference divergences.

Theorem 5.4. *Under Assumption [4.2] and the assumption that $L(Y, Y') = O_p(b_q)$ with $b_q = o(1)$ and $\tau_Y b_q^2 = o(1)$, as $q \to \infty$ with p and n remaining fixed, we have*

$$\widetilde{\mathcal{D}_n^2}(X, Y) = \frac{1}{2\tau_Y} \sum_{j=1}^{q} \widetilde{\mathcal{D}_n^2}_{; K_d, \rho_j}(X, Y_{(j)}) + \mathcal{R}_n,$$

where $\mathcal{R}_n$ is the remainder term such that $\mathcal{R}_n = O_p(\tau_Y b_q^2) = o_p(1)$.

Remark 5.8. *In particular, when both K_d and K_g are Euclidean distances, we have*

$$\widetilde{\mathcal{D}_n^2}(X, Y) = dCov_n^2(X, Y) = \frac{1}{\tau_Y} \sum_{j=1}^{\tilde{q}} MDD_n^2(Y_j | X) + \mathcal{R}_n,$$

where $MDD_n^2(Y_j | X)$ is the unbiased U-statistic type estimator of $MDD^2(Y_j | X)$ defined as in (2.14) with $d_{\mathcal{X}}(x, x') = \|x - x'\|$ for $x, x' \in \mathbb{R}^{\tilde{p}}$ and $d_{\mathcal{Y}}(y, y') = |y - y'|^2 / 2$ for $y, y' \in \mathbb{R}$, respectively.

Now denote $X_k = (X_{k(1)}, \ldots, X_{k(p)})$ and $Y_k = (Y_{k(1)}, \ldots, Y_{k(q)})$ for $1 \le k \le n$. Define the leading term of $\widetilde{\mathcal{D}_n^2}(X, Y)$ in equation (5.5) as

$$L := \frac{1}{4\tau_{XY}} \sum_{i=1}^{p} \sum_{j=1}^{q} \widetilde{D_n^2}_{; \rho_i, \rho_j}(X_{(i)}, Y_{(j)}).$$

It can be verified that

$$L = \frac{1}{4\tau_{XY}} \sum_{i=1}^{p} \sum_{j=1}^{q} \left(\tilde{D}^X(i) \cdot \tilde{D}^Y(j) \right),$$

where $\tilde{D}^X(i), \tilde{D}^Y(j)$ are the $\mathcal{U}$-centered versions of $D^X(i) = \left(d_{kl}^X(i)\right)_{k,l=1}^{n}$ and $D^Y(j) = \left(d_{kl}^Y(j)\right)_{k,l=1}^{n}$, respectively. As an advantage of using the double-centered distances, we have for all $1 \le i, i' \le p$, $1 \le j, j' \le q$ and $\{k, l\} \ne \{u, v\}$,

$$\mathbb{E}\left[d_{kl}^X(i) d_{uv}^X(i')\right] = \mathbb{E}\left[d_{kl}^Y(j) d_{uv}^Y(j')\right] = \mathbb{E}\left[d_{kl}^X(i) d_{uv}^Y(j)\right] = 0. \quad (5.7)$$

See for example the proof of Proposition 2.2.1 in [57] for a detailed explanation.

Assumption 5.4. *For fixed n, as $p, q \to \infty$,*

$$\begin{pmatrix} \frac{1}{2\tau_X} \sum_{i=1}^{p} d_{kl}^X(i) \\ \frac{1}{2\tau_Y} \sum_{j=1}^{q} d_{uv}^Y(j) \end{pmatrix}_{k<l, \, u<v} \xrightarrow{d} \begin{pmatrix} d_{kl}^1 \\ d_{uv}^2 \end{pmatrix}_{k<l, \, u<v},$$

where $\{d_{kl}^1,\, d_{uv}^2\}_{k<l,\,u<v}$ are jointly Gaussian. Further we assume that

$$var(d_{kl}^1) := \sigma_X^2 = \lim_{p\to\infty} \frac{1}{4\tau_X^2} \sum_{i,i'=1}^{p} D_{\rho_i,\rho_{i'}}^2 \left(X_{(i)}, X_{(i')}\right),$$

$$var(d_{kl}^2) := \sigma_Y^2 = \lim_{q\to\infty} \frac{1}{4\tau_Y^2} \sum_{j,j'=1}^{q} D_{\rho_j,\rho_{j'}}^2 \left(Y_{(j)}, Y_{(j')}\right),$$

$$cov(d_{kl}^1, d_{kl}^2) := \sigma_{XY}^2 = \lim_{p,q\to\infty} \frac{1}{4\tau_{XY}} \sum_{i=1}^{p}\sum_{j=1}^{q} D_{\rho_i,\rho_j}^2 \left(X_{(i)}, Y_{(j)}\right).$$

In view of (5.7), we have $\mathrm{cov}\,(d_{kl}^1, d_{uv}^1) = \mathrm{cov}\,(d_{kl}^2, d_{uv}^2) = \mathrm{cov}\,(d_{kl}^1, d_{uv}^2) = 0$ for $\{k,l\} \neq \{u,v\}$. Theorem 5.3 states that for growing p and q and fixed n, $\widetilde{\mathcal{D}_n^2}(X,Y)$ and L are asymptotically equivalent. By studying the leading term, we obtain the limiting distribution of $\widetilde{\mathcal{D}_n^2}(X,Y)$ as follows.

Theorem 5.5. *Under Assumptions 4.2, 5.3 and 5.4, for fixed n and $p,q \to \infty$,*

$$\widetilde{\mathcal{D}_n^2}(X,Y) \xrightarrow{d} \frac{1}{\nu} d^{1\top} M d^2,$$

$$\widetilde{\mathcal{D}_n^2}(X,X) \xrightarrow{d} \frac{1}{\nu} d^{1\top} M d^1 \overset{d}{=} \frac{\sigma_X^2}{\nu} \chi_\nu^2,$$

$$\widetilde{\mathcal{D}_n^2}(Y,Y) \xrightarrow{d} \frac{1}{\nu} d^{2\top} M d^2 \overset{d}{=} \frac{\sigma_Y^2}{\nu} \chi_\nu^2,$$

where M is a projection matrix of rank $\nu = \frac{n(n-3)}{2}$, and

$$\begin{pmatrix} d^1 \\ d^2 \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_X^2\, I_{\frac{n(n-1)}{2}} & \sigma_{XY}^2\, I_{\frac{n(n-1)}{2}} \\ \sigma_{XY}^2\, I_{\frac{n(n-1)}{2}} & \sigma_Y^2\, I_{\frac{n(n-1)}{2}} \end{pmatrix}\right).$$

To perform independence testing, in the spirit of [49], we define the studentized test statistic

$$\mathcal{T}_n := \sqrt{\nu - 1}\, \frac{\widetilde{\mathcal{DC}_n^2}(X,Y)}{\sqrt{1 - \left(\widetilde{\mathcal{DC}_n^2}(X,Y)\right)^2}}, \tag{5.8}$$

where

$$\widetilde{\mathcal{DC}_n^2}(X,Y) = \frac{\widetilde{\mathcal{D}_n^2}(X,Y)}{\sqrt{\widetilde{\mathcal{D}_n^2}(X,X)\,\widetilde{\mathcal{D}_n^2}(Y,Y)}}.$$

Define $\psi = \sigma_{XY}^2 / \sqrt{\sigma_X^2 \sigma_Y^2}$. The following theorem states the asymptotic distributions of the test statistic $\mathcal{T}_n$ under the null hypothesis $\tilde{H}_0 : X \perp\!\!\!\perp Y$ and the alternative hypothesis $\tilde{H}_A : X \not\perp\!\!\!\perp Y$.

Theorem 5.6. *Under Assumptions 4.2, 5.3 and 5.4, for fixed n and $p, q \to \infty$,*

$$P_{\tilde{H}_0}\left(\mathcal{T}_n \leq t\right) \longrightarrow P\left(t_{\nu-1} \leq t\right),$$
$$P_{\tilde{H}_A}\left(\mathcal{T}_n \leq t\right) \longrightarrow \mathbb{E}\left[P\left(t_{\nu-1,W} \leq t|W\right)\right],$$

where t is any fixed real number and $W \sim \sqrt{\frac{\psi^2}{1-\psi^2}\chi_\nu^2}$.

For an explicit form of $\mathbb{E}\left[P\left(t_{\nu-1,W} \leq t|W\right)\right]$, we refer the reader to Lemma 3 in the appendix of [57]. Now consider the local alternative hypothesis $\tilde{H}_A^*$: $X \not\perp Y$ with $\psi = \psi_0/\sqrt{\nu}$, where ψ_0 is a constant with respect to n. The following proposition gives an approximation of $\mathbb{E}\left[P\left(t_{\nu-1,W} \leq t|W\right)\right]$ under the local alternative hypothesis $\tilde{H}_A^*$ when n is allowed to grow.

Proposition 5.1. *Under $\tilde{H}_A^*$, as $n \to \infty$ and $t = O(1)$,*

$$\mathbb{E}\left[P\left(t_{\nu-1,W} \leq t|W\right)\right] = P\left(t_{\nu-1,\psi_0} \leq t\right) + O\left(\frac{1}{\nu}\right).$$

The following summarizes our key findings in this section.

- **Advantages of our proposed metrics over the Euclidean dCov and HSIC:**

  i) Our proposed dependence metrics completely characterize independence between X and Y in the low-dimensional setup, and can detect group-wise non-linear dependencies between X and Y in the high-dimensional setup as opposed to merely detecting component-wise linear dependencies by the Euclidean dCov and HSIC (in light of Theorem 2.1.1 in [57]).

  ii) We also showed that with p remaining fixed and q growing high, the Euclidean dCov can only quantify conditional mean independence of the components of Y given X (which is weaker than independence). To the best of our knowledge, this has not been pointed out in the literature before.

- **Advantages over the marginal aggregation approach by [57]:**

  i) In the low-dimensional setup, our proposed dependence metrics can completely characterize independence between X and Y, whereas the metric proposed by [57] can only capture pairwise dependencies between the components of X and Y.

  ii) We provide a neater way of generalizing dCov and HSIC between X and Y which is shown to be asymptotically equivalent to the marginal aggregation of cross-component distance covariances proposed by [57] as dimensions grow high. Also grouping or partitioning the two high-dimensional random vectors (which again may be problem specific) allows us to detect a wider range of alternatives compared to only detecting component-wise non-linear dependencies, as independence of two univariate marginals is implied from independence of two higher dimensional marginals containing the two univariate marginals.

iii) The computational complexity of the (unbiased) squared sample $\mathcal{D}(X, Y)$ is $O(n^2(p + q))$. Thus the computational cost of our proposed two-sample t-test only grows linearly with the dimension and therefore is scalable to very high-dimensional data. Although a naive aggregation of marginal distance covariances has a computational complexity of $O(n^2 pq)$, the approach of [57] essentially corresponds to the use of an additive kernel and the computational cost of their proposed estimator can also be made linear in the dimensions if properly implemented.

TABLE 3
Summary of the behaviors of the proposed homogeneity/dependence metrics for different choices of $\rho_i(x, x')$ in high dimension.

Choice of $\rho_i(x, x')$	Asymptotic behavior of the proposed homogeneity metric	Asymptotic behavior of the proposed dependence metric
the semi-metric $\|x - x'\|^2$	Behaves as a sum of squared Euclidean distances	Behaves as a sum of squared Pearson correlations
metric of strong negative type on $\mathbb{R}^{d_i}$	Behaves as a sum of groupwise energy distances with the metric ρ_i	Behaves as a sum of groupwise dCov with the metric ρ_i
$k_i(x, x) + k_i(x', x') - 2k_i(x, x')$, where k_i is a characteristic kernel on $\mathbb{R}^{d_i} \times \mathbb{R}^{d_i}$	Behaves as a sum of groupwise MMD with the kernel k_i	Behaves as a sum of groupwise HSIC with the kernel k_i

6. Numerical studies

6.1. Testing for homogeneity of distributions

We investigate the empirical size and power of the tests for homogeneity of two high dimensional distributions. For comparison, we consider the t-tests based on the following metrics:

   I. $\mathcal{E}$ with ρ_i as the Euclidean distance for $1 \leq i \leq p$;
  II. $\mathcal{E}$ with ρ_i as the distance induced by the Laplace kernel for $1 \leq i \leq p$;
 III. $\mathcal{E}$ with ρ_i as the distance induced by the Gaussian kernel for $1 \leq i \leq p$;
 IV. the usual Euclidean energy distance;
  V. MMD with the Laplace kernel;
 VI. MMD with the Gaussian kernel.

We set $d_i = 1$ in Examples 6.1 and 6.2, and $d_i = 2$ in Example 6.3 for $1 \leq i \leq p$.

Example 6.1. *Consider* $X_k = (X_{k1}, \ldots, X_{kp})$ *and* $Y_l = (Y_{l1}, \ldots, Y_{lp})$ *with* $k = 1, \ldots, n$ *and* $l = 1, \ldots, m$. *We generate i.i.d. samples from the following models:*

1. $X_k \sim N(0, I_p)$ *and* $Y_l \sim N(0, I_p)$.
2. $X_k \sim N(0, \Sigma)$ *and* $Y_l \sim N(0, \Sigma)$, *where* $\Sigma = (\sigma_{ij})_{i,j=1}^p$ *with* $\sigma_{ii} = 1$ *for* $i = 1, \ldots, p$, $\sigma_{ij} = 0.25$ *if* $1 \leq |i - j| \leq 2$ *and* $\sigma_{ij} = 0$ *otherwise.*
3. $X_k \sim N(0, \Sigma)$ *and* $Y_l \sim N(0, \Sigma)$, *where* $\Sigma = (\sigma_{ij})_{i,j=1}^p$ *with* $\sigma_{ij} = 0.7^{|i-j|}$.

Example 6.2. *Consider* $X_k = (X_{k1}, \ldots, X_{kp})$ *and* $Y_l = (Y_{l1}, \ldots, Y_{lp})$ *with* $k = 1, \ldots, n$ *and* $l = 1, \ldots, m$. *We generate i.i.d. samples from the following models:*

1. $X_k \sim N(\mu, I_p)$ *with* $\mu = (1, \ldots, 1) \in \mathbb{R}^p$ *and* $Y_{li} \overset{ind}{\sim}$ *Poisson* (1) *for* $i = 1, \ldots, p$, *i.e., the components of* Y_l *independently follow Poisson* (1) *distribution.*
2. $X_k \sim N(\mu, I_p)$ *with* $\mu = (1, \ldots, 1) \in \mathbb{R}^p$ *and* $Y_{li} \overset{ind}{\sim}$ *Exponential* (1) *for* $i = 1, \ldots, p$,, *i.e., the components of* Y_l *independently follow Exponential* (1) *distribution.*
3. $X_k \sim N(0, I_p)$ *and* $Y_l = (Y_{l1}, \ldots, Y_{l\lfloor \beta p \rfloor}, Y_{l(\lfloor \beta p \rfloor + 1)}, \ldots, Y_{lp})$, *where* Y_{l1}, $\ldots, Y_{l\lfloor \beta p \rfloor} \overset{i.i.d.}{\sim}$ *Rademacher* (0.5) *and* $Y_{l(\lfloor \beta p \rfloor + 1)}, \ldots, Y_{lp} \overset{i.i.d.}{\sim} N(0,1)$. *In other words, the first* $\lfloor \beta p \rfloor$ *components of* Y *independently follow Rademacher* (0.5) *distribution, whereas the last* $p - \lfloor \beta p \rfloor$ *components are independently generated from a standard normal distribution.*
4. $X_k \sim N(0, I_p)$ *and* $Y_l = (Y_{l1}, \ldots, Y_{l\lfloor \beta p \rfloor}, Y_{l(\lfloor \beta p \rfloor + 1)}, \ldots, Y_{lp})$, *where* Y_{l1}, $\ldots, Y_{l\lfloor \beta p \rfloor} \overset{i.i.d.}{\sim}$ *Uniform* $(-\sqrt{3}, \sqrt{3})$ *and* $Y_{l(\lfloor \beta p \rfloor + 1)}, \ldots, Y_{lp} \overset{i.i.d.}{\sim} N(0,1)$. *In other words, the first* $\lfloor \beta p \rfloor$ *components of* Y *independently follow Uniform* $(-\sqrt{3}, \sqrt{3})$ *distribution, whereas the last* $p - \lfloor \beta p \rfloor$ *components are independently generated from a standard normal distribution.*
5. $X_k = R^{1/2} Z_{1k}$ *and* $Y_l = R^{1/2} Z_{2l}$, *where* $R = (r_{ij})_{i,j=1}^p$ *with* $r_{ii} = 1$ *for* $i = 1, \ldots, p$, $r_{ij} = 0.25$ *if* $1 \leq |i - j| \leq 2$ *and* $r_{ij} = 0$ *otherwise,* $Z_{1k} \sim N(0, I_p)$ *and* $Z_{2l} = \underbrace{(Z_{2l1}, \ldots, Z_{2lp})}_{\overset{i.i.d.}{\sim} Exponential(1)} - 1$.

Example 6.3. Consider $X_k = (X_{k(1)}, \ldots, X_{k(p)})$ and $Y_l = (Y_{l(1)}, \ldots, Y_{l(p)})$ with $k = 1, \ldots, n$ and $l = 1, \ldots, m$ and $d_i = 2$ for $1 \leq i \leq p$. We generate i.i.d. samples from the following models:

1. $X_{k(i)} \sim N(\mu, \Sigma_1)$ and $Y_{l(i)} \sim N(\mu, \Sigma_2)$ with $X_{k(i)} \perp\!\!\!\perp X_{k(j)}$ and $Y_{l(i)} \perp\!\!\!\perp Y_{l(j)}$ for $1 \leq i \neq j \leq p$, where $\mu = (1,1)^\top$, $\Sigma_1 = \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}$.
2. $X_{k(i)} \sim N(\mu, \Sigma)$ with $X_{k(i)} \perp\!\!\!\perp X_{k(j)}$ for $1 \leq i \neq j \leq p$, where $\mu = (1,1)^\top$, $\Sigma = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$. The components of Y_l are i.i.d. Exponential (1).

Note that for Examples 6.1 and 6.2, the metric defined in equation (3.1) essentially boils down to the special case in equation (3.3). We try small sample sizes $n = m = 50$, dimensions $p = q = 50, 100$ and 200, and $\beta = 1/2$. Table 4 reports the proportion of rejections out of 1000 simulation runs for the different tests. For the tests V and VI, we chose the bandwidth parameter heuristically as the median distance between the aggregated sample observations. For tests II and III, the bandwidth parameters are chosen using the median heuristic separately for each group.

In Example 6.1, the data generating scheme suggests that the variables X and Y are identically distributed. The results in Table 4 show that the tests based on both the proposed homogeneity metrics and the usual Euclidean energy distance and MMD perform more or less equally good, and the rejection probabilities are quite close to the 10% or 5% nominal level. In Example 6.2, clearly X and Y have different distributions but $\mu_X = \mu_Y$ and $\Sigma_X = \Sigma_Y$. The results in Table 4 indicate that the tests based on the proposed homogeneity metrics are able to detect the differences between the two high-dimensional distributions beyond the first two moments unlike the tests based on the usual Euclidean energy distance and MMD, and thereby outperform the latter in terms of empirical power. In Example 6.3, clearly $\mu_X = \mu_Y$ and $\operatorname{tr}\Sigma_X = \operatorname{tr}\Sigma_Y$ and the results show that the tests based on the proposed homogeneity metrics are able to detect the in-homogeneity of the low-dimensional marginal distributions unlike the tests based on the usual Euclidean energy distance and MMD.

Remark 6.1. *In Example 6.3.1, marginally the p-many two-dimensional groups of X and Y are not identically distributed, but each of the $2p$ unidimensional components of X and Y have identical distributions. Consequently, choosing $d_i = 1$ for $1 \leq i \leq p$ leads to trivial power of even our proposed tests, as is evident from Table 5 below. This demonstrates that grouping allows us to detect a wider range of alternatives.*

6.2. Testing for independence

We study the empirical size and power of tests for independence between two high dimensional random vectors. We consider the t-tests based on the following metrics:

   I. $\mathcal{D}$ with $d_i = 1$ and ρ_i be the Euclidean distance for $1 \leq i \leq p$;
  II. $\mathcal{D}$ with $d_i = 1$ and ρ_i be the distance induced by the Laplace kernel for $1 \leq i \leq p$;
 III. $\mathcal{D}$ with $d_i = 1$ and ρ_i be the distance induced by the Gaussian kernel for $1 \leq i \leq p$;
 IV. the usual Euclidean distance covariance;
  V. HSIC with the Laplace kernel;
 VI. HSIC with the Gaussian kernel.

We also compare the empirical size and power of the above tests with the

VII. projection correlation based test for independence proposed by [56],

TABLE 4
Empirical size and power for the different tests of homogeneity of distributions.

			I		II		III		IV		V		VI	
		p	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
	(1)	50	0.109	0.062	0.109	0.058	0.106	0.063	0.109	0.068	0.110	0.069	0.109	0.070
	(1)	100	0.124	0.073	0.119	0.053	0.121	0.063	0.116	0.067	0.114	0.068	0.117	0.068
	(1)	200	0.086	0.043	0.099	0.048	0.088	0.035	0.090	0.045	0.086	0.043	0.090	0.045
	(2)	50	0.114	0.069	0.108	0.054	0.118	0.068	0.116	0.077	0.115	0.073	0.116	0.078
Ex 6.1	(2)	100	0.130	0.069	0.133	0.073	0.124	0.070	0.126	0.067	0.123	0.068	0.124	0.067
	(2)	200	0.099	0.048	0.103	0.041	0.092	0.047	0.097	0.040	0.095	0.039	0.097	0.040
	(3)	50	0.100	0.064	0.107	0.057	0.099	0.060	0.112	0.072	0.105	0.067	0.110	0.073
	(3)	100	0.103	0.062	0.113	0.061	0.113	0.063	0.097	0.060	0.100	0.057	0.098	0.059
	(3)	200	0.108	0.062	0.115	0.062	0.117	0.064	0.091	0.055	0.093	0.056	0.090	0.055
	(1)	50	1	1	1	1	0.995	0.994	0.102	0.067	0.111	0.069	0.103	0.066
	(1)	100	1	1	1	1	1	1	0.120	0.066	0.120	0.071	0.119	0.066
	(1)	200	1	1	1	1	1	1	0.111	0.057	0.111	0.057	0.111	0.057
	(2)	50	1	1	1	1	1	1	0.126	0.085	0.154	0.105	0.119	0.073
	(2)	100	1	1	1	1	1	1	0.098	0.058	0.108	0.066	0.094	0.055
	(2)	200	1	1	1	1	1	1	0.111	0.055	0.114	0.056	0.108	0.054
Ex 6.2	(3)	50	1	1	1	1	1	0.999	0.118	0.069	0.117	0.072	0.120	0.070
	(3)	100	1	1	1	1	1	1	0.102	0.067	0.106	0.065	0.103	0.067
	(3)	200	1	1	1	1	1	1	0.103	0.046	0.103	0.049	0.102	0.046
	(4)	50	0.452	0.328	0.863	0.771	0.552	0.421	0.114	0.061	0.111	0.061	0.114	0.061
	(4)	100	0.640	0.491	0.990	0.967	0.761	0.637	0.098	0.063	0.104	0.063	0.098	0.062
	(4)	200	0.840	0.733	1	0.999	0.933	0.876	0.105	0.042	0.108	0.042	0.105	0.043
	(5)	50	1	1	1	1	1	1	0.128	0.078	0.163	0.098	0.115	0.077
	(5)	100	1	1	1	1	1	1	0.098	0.053	0.115	0.063	0.091	0.051
	(5)	200	1	1	1	1	1	1	0.100	0.050	0.103	0.054	0.098	0.050
	(1)	50	1	1	1	1	1	1	0.157	0.098	0.223	0.137	0.156	0.098
	(1)	100	1	1	1	1	1	1	0.158	0.089	0.188	0.124	0.157	0.090
Ex 6.3	(1)	200	1	1	1	1	1	1	0.122	0.074	0.161	0.091	0.121	0.074
	(2)	50	1	1	1	1	1	1	0.140	0.078	0.190	0.118	0.137	0.075
	(2)	100	1	1	1	1	1	1	0.139	0.080	0.171	0.105	0.136	0.080
	(2)	200	1	1	1	1	1	1	0.109	0.053	0.127	0.069	0.108	0.053

TABLE 5
Empirical power in Example 6.3.1 if we choose $d_i = 1$ for $1 \leq i \leq p$.

			I		II		III		IV		V		VI	
		p	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
	(1)	50	0.144	0.087	0.133	0.076	0.143	0.086	0.174	0.107	0.266	0.170	0.175	0.105
Ex 6.3	(1)	100	0.145	0.085	0.134	0.070	0.142	0.085	0.157	0.098	0.223	0.137	0.156	0.098
	(1)	200	0.126	0.063	0.101	0.058	0.111	0.065	0.158	0.089	0.188	0.124	0.157	0.090

which is shown to have higher empirical power compared to the usual Euclidean distance covariance when the dimensions are relatively large. The numerical examples we consider are motivated from [57].

Example 6.4. Consider $X_k = (X_{k1}, \ldots, X_{kp})$ and $Y_k = (Y_{k1}, \ldots, Y_{kp})$ for $k = 1, \ldots, n$. We generate i.i.d. samples from the following models:

1. $X_k \sim N(0, I_p)$ and $Y_k \sim N(0, I_p)$.
2. $X_k \sim AR(1), \phi = 0.5, Y_k \sim AR(1), \phi = -0.5$, where $AR(1)$ denotes the autoregressive model of order 1 with parameter ϕ.
3. $X_k \sim N(0, \Sigma)$ and $Y_k \sim N(0, \Sigma)$, where $\Sigma = (\sigma_{ij})_{i,j=1}^p$ with $\sigma_{ij} = 0.7^{|i-j|}$.

Example 6.5. Consider $X_k = (X_{k1}, \ldots, X_{kp})$ and $Y_k = (Y_{k1}, \ldots, Y_{kp})$, $k = 1, \ldots, n$. We generate i.i.d. samples from the following models:

1. $X_k \sim N(0, I_p)$ and $Y_{kj} = X_{kj}^2$ for $j = 1, \ldots, p$.
2. $X_k \sim N(0, I_p)$ and $Y_{kj} = \log|X_{kj}|$ for $j = 1, \ldots, p$.
3. $X_k \sim N(0, \Sigma)$ and $Y_{kj} = X_{kj}^2$ for $j = 1, \ldots, p$, where $\Sigma = (\sigma_{ij})_{i,j=1}^p$ with $\sigma_{ij} = 0.7^{|i-j|}$.

Example 6.6. Consider $X_k = (X_{k1}, \ldots, X_{kp})$ and $Y_k = (Y_{k1}, \ldots, Y_{kp})$, $k = 1, \ldots, n$. Let $\circ$ denote the Hadamard product of matrices. We generate i.i.d. samples from the following models:

1. $X_{kj} \sim U(-1, 1)$ for $j = 1, \ldots, p$, and $Y_k = X_k \circ X_k$.
2. $X_{kj} \sim U(0, 1)$ for $j = 1, \ldots, p$, and $Y_k = 4X_k \circ X_k - 4X_k + 2$.
3. $X_{kj} = \sin(Z_{kj})$ and $Y_{kj} = \cos(Z_{kj})$ with $Z_{kj} \sim U(0, 2\pi)$ and $j = 1, \ldots, p$.

For each example, we draw 1000 simulated datasets and perform tests for independence between the two variables based on the proposed dependence metrics, and the usual Euclidean dCov and HSIC. We try a small sample size $n = 50$ and dimensions $p = 50, 100$ and 200. For the tests II, III, V and VI, we chose the bandwidth parameter heuristically as the median distance between the sample observations. Table 6 reports the proportion of rejections out of the 1000 simulation runs for the different tests. For VII, we conduct a permutation based test with 500 replicates.

In Example 6.4, the data generating scheme suggests that the variables X and Y are independent. The results in Table 6 show that the tests based on the proposed dependence metrics perform almost equally good as the other competitors, and the rejection probabilities are quite close to the 10% or 5% nominal level. In Examples 6.5 and 6.6, the variables are clearly (componentwise non-linearly) dependent by virtue of the data generating scheme. The results indicate that the tests based on the proposed dependence metrics are able to detect the componentwise non-linear dependence between the two high-dimensional random vectors unlike the tests based on the usual Euclidean dCov and HSIC, and thereby outperform the latter in terms of empirical power. Also, our proposed tests clearly perform far better compared to the projection correlation based test.

TABLE 6
Empirical size and power for the different tests of independence.

			I		II		III		IV		V		VI		VII	
		p	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
	(1)	50	0.115	0.053	0.109	0.055	0.106	0.053	0.112	0.060	0.112	0.053	0.111	0.061	0.119	0.059
	(1)	100	0.106	0.057	0.090	0.046	0.095	0.048	0.111	0.060	0.112	0.059	0.113	0.060	0.116	0.062
	(1)	200	0.076	0.031	0.084	0.046	0.084	0.042	0.096	0.035	0.090	0.038	0.095	0.035	0.091	0.038
	(2)	50	0.101	0.052	0.096	0.061	0.094	0.053	0.096	0.050	0.103	0.054	0.096	0.052	0.094	0.050
Ex 6.4	(2)	100	0.080	0.036	0.083	0.035	0.086	0.042	0.081	0.041	0.088	0.044	0.083	0.041	0.081	0.037
	(2)	200	0.117	0.051	0.098	0.056	0.103	0.052	0.104	0.048	0.103	0.052	0.106	0.048	0.101	0.050
	(3)	50	0.093	0.056	0.098	0.052	0.097	0.056	0.091	0.052	0.080	0.050	0.087	0.052	0.094	0.044
	(3)	100	0.104	0.052	0.085	0.046	0.091	0.054	0.104	0.048	0.105	0.051	0.102	0.048	0.098	0.045
	(3)	200	0.105	0.059	0.110	0.057	0.103	0.051	0.106	0.055	0.099	0.052	0.105	0.056	0.109	0.058
	(1)	50	1	1	1	1	1	1	0.267	0.172	0.534	0.398	0.277	0.182	0.388	0.280
	(1)	100	1	1	1	1	1	1	0.171	0.102	0.284	0.180	0.167	0.102	0.323	0.204
	(1)	200	1	1	1	1	1	1	0.130	0.075	0.194	0.108	0.128	0.073	0.302	0.188
	(2)	50	1	1	1	1	1	1	0.154	0.092	0.199	0.130	0.154	0.091	0.147	0.077
Ex 6.5	(2)	100	1	1	1	1	1	1	0.109	0.050	0.128	0.064	0.108	0.049	0.108	0.048
	(2)	200	1	1	1	1	1	1	0.099	0.057	0.107	0.060	0.097	0.057	0.101	0.048
	(3)	50	1	1	1	1	1	1	0.654	0.546	0.981	0.959	0.708	0.631	0.661	0.545
	(3)	100	1	1	1	1	1	1	0.418	0.309	0.790	0.700	0.455	0.343	0.535	0.419
	(3)	200	1	1	1	1	1	1	0.277	0.188	0.504	0.391	0.284	0.193	0.454	0.345
	(1)	50	1	1	1	1	1	1	0.129	0.072	0.193	0.105	0.130	0.071	0.141	0.076
	(1)	100	1	1	1	1	1	1	0.145	0.069	0.158	0.091	0.145	0.069	0.155	0.084
	(1)	200	1	1	1	1	1	1	0.113	0.065	0.123	0.067	0.113	0.065	0.130	0.068
	(2)	50	1	1	1	1	1	1	0.129	0.072	0.193	0.105	0.130	0.071	0.141	0.076
Ex 6.6	(2)	100	1	1	1	1	1	1	0.145	0.069	0.158	0.091	0.145	0.069	0.155	0.084
	(2)	200	1	1	1	1	1	1	0.113	0.065	0.123	0.067	0.113	0.065	0.130	0.068
	(3)	50	0.540	0.388	1	1	0.859	0.760	0.110	0.057	0.108	0.063	0.111	0.056	0.092	0.049
	(3)	100	0.550	0.416	1	1	0.857	0.761	0.108	0.063	0.112	0.063	0.108	0.062	0.097	0.051
	(3)	200	0.542	0.388	1	1	0.872	0.765	0.106	0.049	0.111	0.051	0.106	0.050	0.089	0.044

Remark 6.2. *From the results in Table 6, it is clear that when we have $n = 50$ observations, we start noticing a dramatic improvement of the empirical power even for $p = q = 50$, when we implement tests for independence based on our proposed dependence metrics. In Section C.1 in the appendix, we compare our tests with the tests IV-VI for much smaller choices of p and q, for example $p/n = 0.1$ (similar for q). The empirical power for all the tests turn out to be quite comparable, which is quite expected as both the Euclidean dCov and HSIC can completely characterize independence in the low-dimensional setting.*

6.3. Real data analysis

6.3.1. Testing for homogeneity of distributions

We consider the two sample testing problem of homogeneity of two high-dimensional distributions on Earthquakes data. The dataset has been downloaded from UCR Time Series Classification Archive (https://www.cs.ucr.edu/~eamonn/time_series_data_2018/). The data are taken from Northern California Earthquake Data Center. There are 368 negative and 93 positive earthquake events and each data point is of length 512.

Table 7 shows the p-values corresponding to the different tests for the homogeneity of distributions between the two classes. Here we set $d_i = 1$ for tests I-III. The p-values corresponding to the tests based on our proposed homogeneity metrics turn out to be extremely small, so we approximate them by zeroes. Clearly the tests based on the proposed homogeneity metrics reject the null hypothesis of equality of distributions at 5% level. However the tests based on the usual Euclidean energy distance and MMD fail to reject the null at 5% level, thereby indicating no significant difference between the distributions of the two classes.

TABLE 7
p-values corresponding to the different tests for homogeneity of distributions for Earthquakes data.

I	II	III	IV	V	VI
≈ 0	≈ 0	≈ 0	0.070	0.068	0.070

Remark 6.3. *In Section C.2 in the appendix, we present histograms corresponding to the first four variables for both the negative and positive earthquake event groups. Figure 3 graphically illustrates inhomogeneities between their marginal distributions, which sheds an intuitively explanation that why the tests based on the proposed homogeneity metrics reject the null hypothesis of equality of distributions.*

6.3.2. Testing for independence

We consider the daily closed stock prices of $p = 126$ companies under the finance sector and $q = 122$ companies under the healthcare sector on the first dates of

each month during the time period between January 1, 2017 and December 31, 2018. The data has been downloaded from Yahoo Finance via the R package 'quantmod'. At each time t, denote the closed stock prices of these companies from the two different sectors by $X_t = (X_{1t}, \ldots, X_{pt})$ and $Y_t = (Y_{1t}, \ldots, Y_{qt})$ for $1 \le t \le 24$. We consider the stock returns $S_t^X = (S_{1t}^X, \ldots, S_{pt}^X)$ and $S_t^Y = (S_{1t}^Y, \ldots, S_{qt}^Y)$ for $1 \le t \le 23$, where $S_{it}^X = \log \frac{X_{i,t+1}}{X_{it}}$ and $S_{jt}^Y = \log \frac{Y_{j,t+1}}{Y_{jt}}$ for $1 \le i \le p$ and $1 \le j \le q$. It seems intuitive that the stock returns for the companies under two different sectors are not totally independent, especially when a large number of companies are being considered. Table 8 shows the p-values corresponding to the different tests for independence between $\{S_t^X\}_{t=1}^{23}$ and $\{S_t^Y\}_{t=1}^{23}$, where we set $d_i = g_i = 1$ for the proposed tests. The tests based on the proposed dependence metrics deliver much smaller p-values compared to the tests based on traditional metrics. We note that the tests based on the usual dCov and HSIC as well as projection correlation fail to reject the null at 5% level, thereby indicating cross-sector independence of stock return values. These results are consistent with the fact that the dependence among financial asset returns is usually nonlinear and thus cannot be fully characterized by traditional metrics in the high dimensional setup.

TABLE 8
p-values corresponding to the different tests for cross-sector independence of stock returns data.

I	II	III	IV	V	VI	VII
4.91×10^{-12}	4.29×10^{-11}	1.12×10^{-11}	0.093	0.084	0.099	0.154

We present an additional real data example on testing for independence in high dimensions in Section C of the appendix. There the data admits a natural grouping, and our results indicate that our proposed tests for independence exhibit better power when we consider the natural grouping than when we consider unit group sizes. It is to be noted that considering unit group sizes makes our proposed statistics essentially equivalent to the marginal aggregation approach proposed by [57]. This indicates that grouping or clustering might improve the power of testing as they are capable of detecting a wider range of dependencies.

7. Discussions

In this paper, we introduce a family of distances for high dimensional Euclidean spaces. Built on the new distances, we propose a class of distance and kernel-based metrics for high-dimensional two-sample and independence testing. The proposed metrics overcome certain limitations of the traditional metrics constructed based on the Euclidean distance. The new distance we introduce corresponds to a semi-norm given by

$$B(x) = \sqrt{\rho_1(x_{(1)}) + \ldots, \rho_p(x_{(p)})},$$
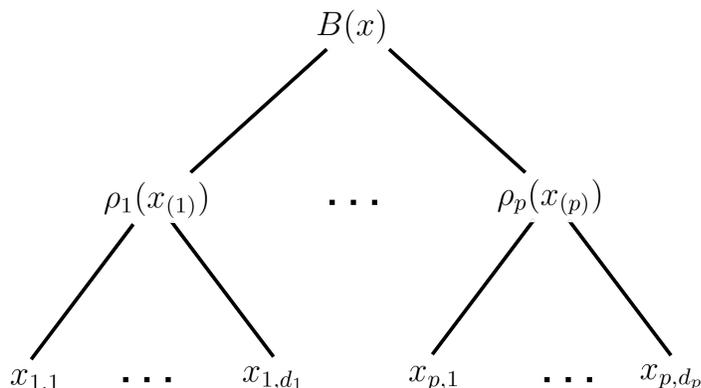
FIG 2. *An interpretation of the semi-norm $B(\cdot)$ based on a tree*

where $\rho_i(x_{(i)}) = \rho_i(x_{(i)}, 0_{d_i})$ and $x = (x_{(1)}, \ldots, x_{(p)}) \in \mathbb{R}^{\tilde{p}}$ with $x_{(i)} = (x_{i,1}, \ldots, x_{i,d_i})$. Such a semi-norm has an interpretation based on a tree as illustrated by Figure 2.

The idea can be generalized to a general tree. Suppose the tree has p internal nodes and each of the variables x_i for $1 \leq i \leq \tilde{p}$ corresponds to a leaf node of the tree. Then for the internal node i, we have a group of leaf nodes (denote the corresponding index set by $\mathcal{S}_i$) that are the off-springs of the internal node. In other words, each internal node induces a group of variables that can overlap with the other groups. In this case, we can define the tree-based norm as

$$B(x) = \sqrt{\sum_{i=1}^{p} w_i \rho_i(x_{\mathcal{S}_i})},$$

where $x_{\mathcal{S}_i} = \{x_j : j \in \mathcal{S}_i\}$, ρ_i is a metric on $\mathbb{R}^{|\mathcal{S}_i|}$ with $|\mathcal{S}_i|$ being the cardinality of $\mathcal{S}_i$, and $w_i \geq 0$ is a sequence of non-negative weights. It is not hard to see that $B(x)$ defined in this way is indeed a semi-norm.

Tree structure provides useful information for doing grouping at different levels/depths. Theoretically, grouping allows us to detect a wider range of alternatives. For example, in two-sample testing, the difference between two one-dimensional marginals is always captured by the difference between two higher dimensional marginals that contain the two one-dimensional marginals. The same thing is true for dependence testing. Generally, one would like to find blocks which are nearly independent, but the variables inside a block have significant dependence among themselves. It is interesting to develop an algorithm for finding the optimal groups using the data and perhaps some auxiliary information such as DAGs representing the underlying distributions. Another interesting direction is to study the semi-norm and distance constructed based on a more sophisticated tree structure. For example, in microbiome-wide association studies, phylogenetic tree or evolutionary tree which is a branching diagram or

"tree" showing the evolutionary relationships among various biological species. Distance and kernel-based metrics constructed based on the distance utilizing the phylogenetic tree information is expected to be more powerful in signal detection. We leave these topics for future investigation.

Appendix

The appendix is organized as follows. In Appendix A we explore our proposed homogeneity and dependence metrics in the low-dimensional setup. In Appendix B we study the asymptotic behavior of our proposed homogeneity and dependence metrics in the high dimension medium sample size (HDMSS) framework where both the dimension(s) and the sample size(s) grow. Appendix C contains some additional numerical results. Finally, Appendix D contains additional proofs of the main results in the paper and Appendix A and B.

Appendix A: Low-dimensional setup

In this section we illustrate that the new class of homogeneity metrics proposed in this paper inherits all the nice properties of generalized energy distance and MMD in the low-dimensional setting. Likewise, the proposed dependence metrics inherit all the desirable properties of generalized dCov and HSIC in the low-dimensional framework.

A.1. Homogeneity metrics

Note that in either Case S1 or S2, the Euclidean space equipped with distance K is of strong negative type. As a consequence, we have the following result.

Theorem A.1. $\mathcal{E}(X, Y) = 0$ *if and only if* $X \overset{d}{=} Y$*, in other words* $\mathcal{E}(X, Y)$ *completely characterizes the homogeneity of the distributions of* X *and* Y*.*

The following proposition shows that $\mathcal{E}_{n,m}(X, Y)$ is a two-sample U-statistic and an unbiased estimator of $\mathcal{E}(X, Y)$.

Proposition A.1. *The U-statistic type estimator enjoys the following properties:*

1. *$\mathcal{E}_{n,m}$ is an unbiased estimator of the population $\mathcal{E}$.*
2. *$\mathcal{E}_{n,m}$ admits the following form:*

$$\mathcal{E}_{n,m}(X, Y) = \frac{1}{\binom{n}{2}\binom{m}{2}} \sum_{1 \leq i < j \leq n} \sum_{1 \leq k < l \leq m} h(X_i, X_j; Y_k, Y_l),$$

where

$$h(X_i, X_j; Y_k, Y_l) = \frac{1}{2}\Big(K(X_i, Y_k) + K(X_i, Y_l) + K(X_j, Y_k) + K(X_j, Y_l)\Big) - K(X_i, X_j) - K(Y_k, Y_l).$$

The following theorem shows the asymptotic behavior of the U-statistic type estimator of $\mathcal{E}$ for fixed p and growing n.

Theorem A.2. *Under Assumption 4.5 and the assumption that* $\sup_{1 \leq i \leq p} \mathbb{E} \rho_i(X_{(i)}, 0_{d_i}) < \infty$ *and* $\sup_{1 \leq i \leq p} \mathbb{E} \rho_i(Y_{(i)}, 0_{d_i}) < \infty$, *as* $m, n \to \infty$ *with* p *remaining fixed, we have the following:*

1. $\mathcal{E}_{n,m}(X, Y) \xrightarrow{a.s.} \mathcal{E}(X, Y)$.
2. *When* $X \overset{d}{=} Y$, $\mathcal{E}_{n,m}$ *has degeneracy of order* $(1, 1)$, *and*

$$\frac{(m-1)(n-1)}{n+m} \mathcal{E}_{n,m}(X, Y) \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k^2 \left(Z_k^2 - 1 \right),$$

where $\{Z_k\}$ *is a sequence of independent* $N(0, 1)$ *random variables and* λ_k*'s depend on the distribution of* (X, Y).

Proposition A.1, Theorem A.1 and Theorem A.2 demonstrate that $\mathcal{E}$ inherits all the nice properties of generalized energy distance and MMD in the low-dimensional setting.

A.2. Dependence metrics

Note that Proposition 3.1 in Section 3 and Proposition 3.7 in [26] ensure that $\mathcal{D}(X, Y)$ completely characterizes independence between X and Y, which leads to the following result.

Theorem A.3. *Under Assumption 5.1,* $\mathcal{D}(X, Y) = 0$ *if and only if* $X \perp\!\!\!\perp Y$.

The following proposition shows that $\widetilde{\mathcal{D}_n^2}(X, Y)$ is an unbiased estimator of $\mathcal{D}^2(X, Y)$ and is a U-statistic of order four.

Proposition A.2. *The U-statistic type estimator* $\widetilde{\mathcal{D}_n^2}$ *(defined in (2.14) in the main paper) has the following properties:*

1. $\widetilde{\mathcal{D}_n^2}$ *is an unbiased estimator of the squared population* $\mathcal{D}^2$.
2. $\widetilde{\mathcal{D}_n^2}$ *is a fourth-order U-statistic which admits the following form:*

$$\widetilde{\mathcal{D}_n^2} = \frac{1}{\binom{n}{4}} \sum_{i<j<k<l} h_{i,j,k,l},$$

where

$$h_{i,j,k,l} = \frac{1}{4!} \sum_{(s,t,u,v)}^{(i,j,k,l)} (d_{st}^X d_{st}^Y + d_{st}^X d_{uv}^Y - 2 d_{st}^X d_{su}^Y)$$

$$= \frac{1}{6} \sum_{s<t,u<v}^{(i,j,k,l)} (d_{st}^X d_{st}^Y + d_{st}^X d_{uv}^Y) - \frac{1}{12} \sum_{(s,t,u)}^{(i,j,k,l)} d_{st}^X d_{su}^Y,$$

the summation is over all possible permutations of the 4-tuple of indices (i, j, k, l). For example, when $(i, j, k, l) = (1, 2, 3, 4)$, there exist 24 permutations, including $(1, 2, 3, 4), \ldots, (4, 3, 2, 1)$. Furthermore, $\widetilde{\mathcal{D}_n^2}$ has degeneracy of order 1 when X and Y are independent.

The following theorem shows the asymptotic behavior of the U-statistic type estimator of $\mathcal{D}^2$ for fixed p, q and growing n.

Theorem A.4. *Under Assumption 5.1, with fixed p, q and $n \to \infty$, we have the following as $n \to \infty$:*

1. *$\widetilde{\mathcal{D}_n^2}(X, Y) \xrightarrow{a.s.} \mathcal{D}^2(X, Y)$;*

2. *When $\mathcal{D}^2(X, Y) = 0$ (i.e., $X \perp\!\!\!\perp Y$), $n\widetilde{\mathcal{D}_n^2}(X, Y) \xrightarrow{d} \sum_{i=1}^{\infty} \tilde{\lambda}_i^2(Z_i^2 - 1)$,*

   where $Z_i's$ are i.i.d. standard normal random variables and $\tilde{\lambda}_i$'s depend on the distribution of (X, Y);

3. *When $\mathcal{D}^2(X, Y) > 0$, $n\widetilde{\mathcal{D}_n^2}(X, Y) \xrightarrow{a.s.} \infty$.*

Proposition A.2, Theorem A.3 and Theorem A.4 demonstrate that in the low-dimensional setting, $\mathcal{D}$ inherits all the nice properties of generalized dCov and HSIC.

Appendix B: High dimension medium sample size (HDMSS)

B.1. Homogeneity metrics

In this subsection, we consider the HDMSS setting where $p \to \infty$ and $n, m \to \infty$ at a slower rate than p. Under H_0, we impose the following conditions to obtain the asymptotic null distribution of the statistic $T_{n,m}$ under the HDMSS setup.

Assumption B.1. *As n, m and $p \to \infty$,*

$$\frac{1}{n^2} \frac{\mathbb{E}\left[H^4(X, X')\right]}{\left(\mathbb{E}\left[H^2(X, X')\right]\right)^2} = o(1), \quad \frac{1}{n} \frac{\mathbb{E}\left[H^2(X, X'') H^2(X', X'')\right]}{\left(\mathbb{E}\left[H^2(X, X')\right]\right)^2} = o(1),$$

$$\frac{\mathbb{E}\left[H(X, X'') H(X', X'') H(X, X''') H(X', X''')\right]}{\left(\mathbb{E}\left[H^2(X, X')\right]\right)^2} = o(1).$$

Remark B.1. *We refer the reader to Section 2.2 in [54] and Remark A.2.2 in [57] for illustrations of Assumption B.1 where ρ_i has been considered to be the Euclidean distance or the squared Euclidean distance, respectively, for $1 \leq i \leq p$.*

Assumption B.2. *Suppose $\mathbb{E}\left[L^2(X, X')\right] = O(\alpha_p^2)$ where α_p is a positive real sequence such that $\tau_X \alpha_p^2 = o(1)$ as $p \to \infty$. Further assume that as $n, p \to \infty$,*

$$\frac{n^4 \tau_X^4 \mathbb{E}\left[R^4(X, X')\right]}{\left(\mathbb{E}\left[H^2(X, X')\right]\right)^2} = o(1).$$

Remark B.2. *We refer the reader to Remark 4.1 in the main paper which illustrates some sufficient conditions under which $\alpha_p = O(\frac{1}{\sqrt{p}})$ and consequently $\tau_X \alpha_p^2 = o(1)$ holds, as $\tau_X \asymp p^{1/2}$. In similar lines of Remark D.1 in Section D of the appendix, it can be argued that $\mathbb{E}\left[R^4(X, X')\right] = O\left(\frac{1}{p^4}\right)$. If we further assume that Assumption 4.4 holds, then we have $\mathbb{E}\left[H^2(X, X')\right] \asymp 1$. Combining all the above, it is easy to verify that $\frac{n^4 \tau_X^4 \mathbb{E}\left[R^4(X, X')\right]}{(\mathbb{E}\left[H^2(X, X')\right])^2} = o(1)$ holds provided $n = o(p^{1/2})$.*

The following theorem illustrates the limiting null distribution of $T_{n,m}$ under the HDMSS setup. We refer the reader to Section D of the appendix for a detailed proof.

Theorem B.1. *Under H_0 and Assumptions 4.5, B.1 and B.2, as n, m and $p \to \infty$, we have*

$$T_{n,m} \xrightarrow{d} N(0, 1).$$

B.2. Dependence metrics

In this subsection, we consider the HDMSS setting where $p, q \to \infty$ and $n \to \infty$ at a slower rate than p, q. The following theorem shows that similar to the HDLSS setting, under the HDMSS setup, $\widetilde{\mathcal{D}}_n^2$ is asymptotically equivalent to the aggregation of group-wise generalized dCov. In other words $\widetilde{\mathcal{D}}_n^2(X, Y)$ can quantify group-wise nonlinear dependence between X and Y in the HDMSS setup as well.

Assumption B.3. $\mathbb{E}[L_X(X, X')^2] = \alpha_p^2$, $\mathbb{E}[L_X(X, X')^4] = \gamma_p^2$, $\mathbb{E}[L_Y(Y, Y')^2] = \beta_q^2$ *and* $\mathbb{E}[L_Y(Y, Y')^4] = \lambda_q^2$, *where* $\alpha_p, \gamma_p, \beta_q, \lambda_q$ *are positive real sequences satisfying* $n\alpha_p = o(1)$, $n\beta_q = o(1)$, $\tau_X^2(\alpha_p\gamma_p + \gamma_p^2) = o(1)$, $\tau_Y^2(\beta_q\lambda_q + \lambda_q^2) = o(1)$, *and* $\tau_{XY}(\alpha_p\lambda_q + \gamma_p\beta_q + \gamma_p\lambda_q) = o(1)$.

Remark B.3. *Following Remark 4.1 in the main paper, we can write $L(X, X') = O(\frac{1}{p}) \sum_{i=1}^p (Z_i - \mathbb{E} Z_i)$, where $Z_i = \rho_i(X_{(i)}, X'_{(i)})$ for $1 \le i \le p$. Assume that $\sup_{1 \le i \le p} \mathbb{E} \rho_i^4(X_{(i)}, 0_{d_i}) < \infty$, which implies $\sup_{1 \le i \le p} \mathbb{E} Z_i^4 < \infty$. Under certain weak dependence assumptions, it can be shown that $\mathbb{E}\left(\sum_{i=1}^p (Z_i - \mathbb{E} Z_i)\right)^4 = O(p^2)$ as $p \to \infty$ (see for example Theorem 1 in [13]). Therefore we have $\mathbb{E}[L(X, X')^4] = O(\frac{1}{p^2})$. It follows from Hölder's inequality that $\mathbb{E}[L(X, X')^2] = O(\frac{1}{p})$. Similar arguments can be made about $\mathbb{E}[L(Y, Y')^4]$ and $\mathbb{E}[L(Y, Y')^2]$ as well.*

Theorem B.2. *Under Assumptions 4.2 and B.3, we can show that*

$$\widetilde{\mathcal{D}}_n^2(X, Y) = \frac{1}{4\tau_{XY}} \sum_{i=1}^p \sum_{j=1}^q \widetilde{D}_{n\,;\,\rho_i,\rho_j}^2(X_{(i)}, Y_{(j)}) + \mathcal{R}_n, \tag{B.1}$$

where $\mathcal{R}_n$ is the remainder term satisfying that $\mathcal{R}_n = O_p(\tau_{XY} (\alpha_p \lambda_q + \gamma_p \beta_q + \gamma_p \lambda_q)) = o_p(1)$, i.e., $\mathcal{R}_n$ is of smaller order compared to the leading term and hence is asymptotically negligible.

The following theorem states the asymptotic null distribution of the studentized test statistic $\mathcal{T}_n$ (given in equation (5.8) in the main paper) under the HDMSS setup. Define

$$U(X_k, X_l) := \frac{1}{\tau_X} \sum_{i=1}^{p} d_{kl}^X(i), \quad \text{and} \quad V(Y_k, Y_l) := \frac{1}{\tau_Y} \sum_{i=1}^{q} d_{kl}^Y(i).$$

Assumption B.4. *Assume that*

$$\frac{\mathbb{E}[U(X, X')]^4}{\sqrt{n} (\mathbb{E}[U(X, X')]^2)^2} = o(1),$$

$$\frac{\mathbb{E}[U(X, X') U(X', X'') U(X'', X''') U(X''', X)]}{(\mathbb{E}[U(X, X')]^2)^2} = o(1),$$

and the same conditions hold for Y in terms of $V(Y, Y')$.

Remark B.4. *We refer the reader to Section 2.2 in [54] and Remark A.2.2 in [57] for illustrations of Assumption B.1 where ρ_i has been considered to be the Euclidean distance or the squared Euclidean distance, respectively.*

We can show that under H_0, the studentized test $\mathcal{T}_n$ converge to the standard normal distribution under the HDMSS setup.

Theorem B.3. *Under H_0 and Assumptions B.3-B.4, as $n, p, q \to \infty$, we have $\mathcal{T}_n \xrightarrow{d} N(0, 1)$.*

Appendix C: Additional numerical results

C.1. Additional simulation study

In this subsection, we compare the tests for independence based on our proposed dependence metrics with the tests based on the Euclidean dCov and HSIC when p and q are much smaller compared to n, for example, $p/n = 0.1$ (and similar for q). We try the Examples 6.4.1-6.4.2 and 6.5.1-6.5.3, and consider $n = 50$ and $p = q = 5$. In Examples 6.4.1-6.4.2, the empirical sizes corresponding to all the tests are quite close to the 10% or 5% nominal level. In Examples 6.5.1-6.5.3, the empirical power for all the tests turn out to be quite comparable as well, which is quite expected as both the Euclidean dCov and HSIC can completely characterize independence in the low-dimensional setting.

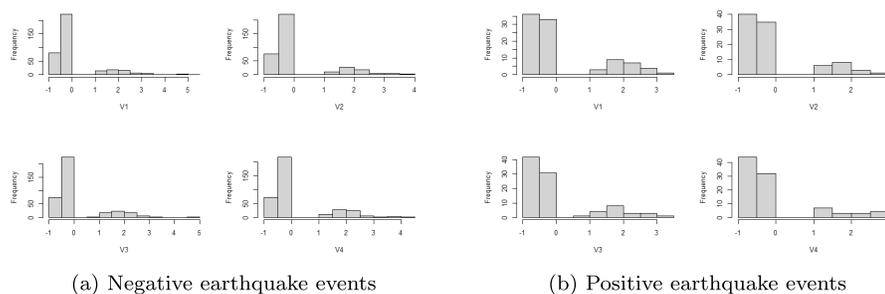### C.2. Additional discussions on the real data example for homogeneity testing

(a) Negative earthquake events

(b) Positive earthquake events

FIG 3. *Comparison of histograms corresponding to the first four variables for the negative and positive earthquake event groups.*

C.3. Additional real data example for independence testing

We consider the monthly closed stock prices of $\tilde{p} = 36$ companies under the transport sector and $\tilde{q} = 41$ companies under the utilities sector between January 1, 2017 and December 31, 2018. The companies under both the sectors are clustered or grouped according to their countries. The data has been downloaded from Yahoo Finance via the R package 'quantmod'. Under the transport sector, we have $q = 14$ countries or groups, viz. USA, Brazil, Canada, Greece, China, Panama, Belgium, Bermuda, UK, Mexico, Chile, Monaco, Ireland and Hong Kong, with $d = (5, 1, 2, 8, 4, 1, 1, 3, 1, 3, 1, 4, 1, 1)$. And under the utilities sector, we have $q = 21$ countries or groups, viz. USA, Mexico, UK, India, Canada, China, Hong Kong, Taiwan, Brazil, Cayman Islands, Israel, Argentina, Chile, Singapore, South Korea, Russia, France, Phillipines, Indonesia, Spain and Turkey, with $g = (5, 1, 3, 1, 5, 2, 3, 1, 4, 1, 1, 4, 1, 1, 2, 1, 1, 1, 1, 1, 1)$. At each time t, denote the closed stock prices of these companies from the two different sectors by $X_t = (X_{1t}, \ldots, X_{pt})$ and $Y_t = (Y_{1t}, \ldots, Y_{qt})$ for $1 \leq t \leq 24$. We consider the stock returns $S_t^X = (S_{1t}^X, \ldots, S_{pt}^X)$ and $S_t^Y = (S_{1t}^Y, \ldots, S_{qt}^Y)$ for $1 \leq t \leq 23$, where $S_{itl}^X = \log \frac{X_{i,t+1,l}}{X_{itl}}$ and $S_{jtl'}^Y = \log \frac{Y_{j,t+1,l'}}{Y_{jtl'}}$ for $1 \leq l \leq d_i$, $1 \leq i \leq p$, $1 \leq l' \leq g_j$ and $1 \leq j \leq q$.

The intuitive idea is, stock returns of transport companies should affect the stock returns of companies under the utilities sector, and here both the random vectors admit a natural grouping based on the countries. Table 10 below shows the p-values corresponding to the different tests for independence between $\{S_t^X\}_{t=1}^{23}$ and $\{S_t^Y\}_{t=1}^{23}$. The tests based on the proposed dependence metrics considering the natural grouping deliver much smaller p-values compared to the tests based on the usual dCov and HSIC, as well as the projection correlation based test, which fail to reject the null hypothesis of independence between $\{S_t^X\}_{t=1}^{23}$ and $\{S_t^Y\}_{t=1}^{23}$. This makes intuitive sense as the dependence among financial asset returns is usually nonlinear in nature and thus cannot be fully characterized by the usual dCov and HSIC in the high dimensional setup.

TABLE 9
Empirical size and power for the different tests of independence when $p/n = 0.1$.

		I		II		III		IV		V		VI	
		10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
Ex 6.4	(1)	0.110	0.067	0.104	0.049	0.110	0.064	0.109	0.069	0.107	0.060	0.106	0.064
	(2)	0.124	0.068	0.115	0.072	0.126	0.079	0.102	0.068	0.113	0.076	0.112	0.064
Ex 6.5	(1)	1	1	1	1	1	1	0.988	0.967	1	1	0.998	0.987
	(2)	1	1	1	1	1	1	0.806	0.698	0.996	0.993	0.816	0.700
	(3)	1	1	1	1	1	1	1	1	1	1	1	1

TABLE 10
p-values corresponding to the different tests for cross-sector independence of stock returns data considering the natural grouping based on countries.

I	II	III	IV	V	VI	VII
0.0008	0.0011	0.0004	0.1106	0.1129	0.4848	0.1120

Table 11 below shows the p-values corresponding to the different tests for independence when we disregard the natural grouping and consider $d_i = 1$ and $g_j = 1$ for all $1 \leq i \leq p$ and $1 \leq j \leq q$. Considering unit group sizes makes our proposed statistics essentially equivalent to the marginal aggregation approach proposed by [57]. In this case the proposed tests have higher p-values than when we consider the natural grouping, indicating that grouping or clustering might improve the power of testing as they are capable of detecting a wider range of dependencies.

TABLE 11
p-values corresponding to the different tests for cross-sector independence of stock returns data considering unit group sizes.

I	II	III	IV	V	VI	VII
0.0067	0.0532	0.0796	0.1106	0.1129	0.4848	0.1120

Appendix D: Technical Appendix

Proof of Proposition 3.1. To prove (1), note that if d is a metric on a space $\mathcal{X}$, then so is $d^{1/2}$. It is easy to see that K^2 is a metric on $\mathbb{R}^{\tilde{p}}$. To prove (2), note that $(\mathbb{R}^{d_i}, \rho_i)$ has strong negative type for $1 \leq i \leq p$. The rest follows from Corollary 3.20 in [26]. ◇

Proof of Proposition A.1. It is easy to verify that $\mathcal{E}_{n,m}$ is an unbiased estimator of $\mathcal{E}$ and is a two-sample U-statistic with the kernel h. ◇

Proof of Theorem A.2. The first part of the proof follows from Theorem 1 in [40] and the observation that $\mathbb{E}\left[|h| \log^+ |h|\right] \leq \mathbb{E}[h^2]$. The power mean inequality says that for $a_i \in \mathbb{R}$, $1 \leq i \leq n$, $n \geq 2$ and $r > 1$,

$$\left| \sum_{i=1}^n a_i \right|^r \leq n^{r-1} \sum_{i=1}^n |a_i|^r . \tag{D.1}$$

Using the power mean inequality, it is easy to see that the assumptions $\sup_{1 \leq i \leq p} \mathbb{E}\rho_i(X_{(i)}, 0_{d_i}) < \infty$ and $\sup_{1 \leq i \leq p} \mathbb{E}\rho_i(Y_{(i)}, 0_{d_i}) < \infty$ ensure that $\mathbb{E}[h^2] < \infty$. For proving the second part, define $h_{1,0}(X) = \mathbb{E}\left[h(X, X'; Y, Y')|X\right]$ and $h_{0,1}(Y) = \mathbb{E}\left[h(X, X'; Y, Y')|Y\right]$ Clearly, when $X \overset{d}{=} Y$, $h_{1,0}(X)$ and $h_{0,1}(Y)$ are degenerate at 0 almost surely. Following Theorem 1.1 in [30], we have

$$\frac{(m-1)(n-1)}{n+m} \mathcal{E}_{nm}(X, Y) \overset{d}{\longrightarrow} \sum_{k=1}^\infty \sigma_k^2 \left[(a_k U_k + b_k V_k)^2 - (a_k^2 + b_k^2)\right] ,$$

where $\{U_k\}, \{V_k\}$ are two sequences of independent $N(0, 1)$ variables, independent of each other, and (σ_k, a_k, b_k)'s depend on the distribution of (X, Y). The proof can be completed by some simple rearrangement of terms. $\diamond$

Proof of Proposition 4.1. The proof is essentially similar to the proof of Proposition 2.1.1 in [57], replacing the Euclidean distance between, for example, X and X', viz. $\|X - X'\|_{\tilde{p}}$, by the new distance metric $K(X, X')$. To show that $R(X, X') = O_p(L^2(X, X'))$ if $L(X, X') = o_p(1)$, we define $f(x) = \sqrt{1 + x}$. By the definition of the Lagrange's form of the remainder term from Taylor's expansion, we have

$$R(X, X') = \int_0^{L(X,X')} f''(t)\,(\,L(X, X') - t\,)\,dt\,.$$

Using R and L interchangeably with $R(X, X')$ and $L(X, X')$ respectively, we can write

$$\begin{aligned}
|R| \;&\leq\; |L|\left[\int_0^L f''(t)\,\mathbb{1}_{L>0}\,dt \;+\; \int_L^0 f''(t)\,\mathbb{1}_{L<0}\,dt\right] \\
&=\; \frac{|L|}{2}\,\Big|1 - \frac{1}{\sqrt{1 + L}}\Big| \\
&=\; \frac{|L|}{2}\,\frac{|L|}{1 + L + \sqrt{1 + L}} \\
&\leq\; \frac{L^2}{2(1 + L)}\,.
\end{aligned} \tag{D.2}$$

It is clear that $R(X, X') = O_p(L^2(X, X'))$ provided that $L(X, X') = o_p(1)$. $\diamond$

Proof of Theorem 4.1. Observe that $\mathbb{E}\,L(X, Y) = \mathbb{E}\,L(X, X') = \mathbb{E}\,L(Y, Y') = 0$. By Proposition 4.1,

$$\begin{aligned}
\mathcal{E}(X, Y) &= 2\,\mathbb{E}\,[\tau + \tau\,R(X, Y)] \;-\; \mathbb{E}\,[\tau_X + \tau_X\,R(X, X')] \;-\; \mathbb{E}\,[\tau_Y + \tau_Y\,R(Y, Y')] \\
&= 2\tau - \tau_X - \tau_Y \;+\; \mathcal{R}_{\mathcal{E}}\,.
\end{aligned}$$

Clearly $|\mathcal{R}_{\mathcal{E}}| \;\leq\; 2\,\tau\,\mathbb{E}\,[\,|R(X, Y)|\,] + \tau_X\,\mathbb{E}\,[\,|R(X, X')|\,] + \tau_Y\,\mathbb{E}\,[\,|R(Y, Y')|\,]$. By (D.2) and Assumption 4.3, we have

$$\tau|R(X, Y)| \leq \frac{\tau L^2(X, Y)}{2(1 + L(X, Y))} = O(\tau a_p^2) = o_p(1).$$

As $\{\sqrt{p}\,L^2(X, Y)/(1 + L(X, Y))\}$ is uniformly integrable and $\tau \asymp \sqrt{p}$, we must have $\tau\mathbb{E}[|R(X, Y)|] = o(1)$. The other terms can be handled in a similar fashion. $\diamond$

Remark D.1. *Write* $L(X, Y) = \frac{1}{\tau^2}(A_p - \mathbb{E}\,A_p) = \frac{1}{\tau^2}\sum_{i=1}^p (Z_i - \mathbb{E}Z_i)$, *where* $A_p := \sum_{i=1}^p Z_i$ *and* $Z_i := \rho_i(X_i, Y_i)$ *for* $1 \leq i \leq p$. *Assume* $\sup_i \mathbb{E}\rho_i^8(X_i, 0_{d_i}) < \infty$ *and* $\sup_i \mathbb{E}\rho_i^8(X_i, 0_{d_i}) < \infty$, *which imply* $\sup_i \mathbb{E}Z_i^8 < \infty$. *Denote* $L(X, Y)$ *by* L

and $R(X, Y)$ by R for notational simplicities. Further assume that $E \exp(t A_p) = O((1 - \theta_1 t)^{-\theta_2 p})$ for $\theta_1, \theta_2 > 0$ and $\theta_2 p > 4$ uniformly over $t < 0$ (which is clearly satisfied when Z_i's are independent and $\mathbb{E} \exp(t Z_i) \leq a_1 (1 - a_2 t)^{-a_3}$ uniformly over $t < 0$ and $1 \leq i \leq p$ for some $a_1, a_2, a_3 > 0$ with $a_3 p > 4$). Under certain weak dependence assumptions, it can be shown that:

1. *$\{ \sqrt{p} L^2 / (1 + L) \}$ is uniformly integrable;*
2. *$\mathbb{E} R^2 = O(\frac{1}{p^2})$.*

Similar arguments hold for $L(X, X')$ and $R(X, X')$, and, $L(Y, Y')$ and $R(Y, Y')$ as well.

Proof of Remark D.1. To prove the first part, define $L_p := \sqrt{p} L^2 / (1 + L)$. Following Chapter 6 of [34], it suffices to show that $\sup_p \mathbb{E} L_p^2 < \infty$. Towards that end, using Hölder's inequality we observe

$$\mathbb{E} L_p^2 \leq \left(\mathbb{E}(p^2 L^8) \right)^{1/2} \left(\mathbb{E} \left[\frac{1}{(1 + L)^4} \right] \right)^{1/2}. \tag{D.3}$$

With $\sup_i \mathbb{E} Z_i^8 < \infty$ and under certain weak dependence assumptions, it can be shown that $\mathbb{E}(A_p - \mathbb{E} A_p)^8 = O(p^4)$ (see for example Theorem 1 in [13]). Consequently we have $\mathbb{E} L^8 = O(\frac{1}{p^4})$, as $\tau \asymp \sqrt{p}$. Clearly this yields $\mathbb{E}(p^2 L^8) = O(\frac{1}{p^2})$.

Now note that

$$\mathbb{E} \left[\frac{1}{(1 + L)^4} \right] = \tau^8 \mathbb{E} \left(\frac{1}{A_p^4} \right). \tag{D.4}$$

Equation (3) in [9] states that for a non-negative random variable U with moment-generating function $M_U(t) = \mathbb{E} \exp(t U)$, one can write

$$\mathbb{E}(U^{-k}) = (\Gamma(k))^{-1} \int_0^\infty t^{k-1} M_U(-t) \, dt, \tag{D.5}$$

for any positive integer k, provided both the integrals exist. Using equation (D.5), the assumptions stated in Remark D.1 and basic properties of beta integrals, some straightforward calculations yield

$$\mathbb{E} \left(\frac{1}{A_p^4} \right) \leq C_1 \int_0^\infty \frac{t^{4-1}}{(1 + \theta_1 t)^{\theta_2 p}} \, dt = C_2 \frac{\Gamma(\theta_2 p - 4)}{\Gamma(\theta_2 p)}, \tag{D.6}$$

where C_1, C_2 are positive constants, which clearly implies that $\mathbb{E} \left(\frac{1}{A_p^4} \right) = O(\frac{1}{p^4})$. This together with equation (D.4) implies that $\mathbb{E} \left[\frac{1}{(1+L)^4} \right] = O(1)$, as $\tau \asymp \sqrt{p}$.

Combining all the above, we get from (D.3) that $\mathbb{E} L_p^2 = O(\frac{1}{p})$ and therefore $\sup_p \mathbb{E} L_p^2 < \infty$, which completes the proof of the first part.

To prove the second part, note that following the proof of Proposition 4.1 and Hölder's inequality we can write

$$\mathbb{E} R^2 = O \left(\mathbb{E} \left[\frac{L^4}{(1 + L)^2} \right] \right) = O \left(\left(\mathbb{E}(L^8) \right)^{1/2} \left(\mathbb{E} \left[\frac{1}{(1 + L)^4} \right] \right)^{1/2} \right). \tag{D.7}$$

Following the arguments as in the proof of the first part, clearly we have $\mathbb{E} L^8 = O(\frac{1}{p^4})$ and $\mathbb{E}\left[\frac{1}{(1+L)^4}\right] = O(1)$. From this and equation (D.7), it is straightforward to verify that $\mathbb{E} R^2 = O(\frac{1}{p^2})$, which completes the proof of the second part. $\diamond$

Proof of Lemma 4.1. To see (2), first observe that the sufficient part is straightforward from equation (4.3) in the main paper. For the necessary part, denote $a = \operatorname{tr}\Sigma_X$, $b = \operatorname{tr}\Sigma_Y$ and $c = \|\mu_X - \mu_Y\|^2$. Then we have $2\sqrt{a+b+c} = \sqrt{2a} + \sqrt{2b}$. Some straightforward calculations yield $(\sqrt{2a} - \sqrt{2b})^2 + 4c = 0$ which implies the rest.

To see (1), again the sufficient part is straightforward from equation (4.2) in the paper and the form of K given in equation (3.1) in the paper. For the necessary part, first note that as $(\mathbb{R}^{d_i}, \rho_i)$ is a metric space of strong negative type for $1 \leq i \leq p$, there exists a Hilbert space $\mathcal{H}_i$ and an injective map $\phi_i : \mathbb{R}^{d_i} \to \mathcal{H}_i$ such that $\rho_i(z, z') = \|\phi_i(z) - \phi_i(z')\|_{\mathcal{H}_i}^2$, where $\langle \cdot, \cdot \rangle_{\mathcal{H}_i}$ is the inner product defined on $\mathcal{H}_i$ and $\| \cdot \|_{\mathcal{H}_i}$ is the norm induced by the inner product (see Proposition 3 in [39] for detailed discussions). Further, if k_i is a distance-induced kernel induced by the metric ρ_i, then by Proposition 14 in [39], $\mathcal{H}_i$ is the RKHS with the reproducing kernel k_i and $\phi_i(z)$ is essentially the canonical feature map for $\mathcal{H}_i$, viz. $\phi_i(z) : z \mapsto k_i(\cdot, z)$. It is easy to see that

$$
\begin{aligned}
\tau_X^2 &= \mathbb{E} \sum_{i=1}^p \|\phi_i(X_{(i)}) - \phi_i(X'_{(i)})\|_{\mathcal{H}_i}^2 \;=\; 2\,\mathbb{E} \sum_{i=1}^p \|\phi_i(X_{(i)}) - \mathbb{E}\,\phi_i(X_{(i)})\|_{\mathcal{H}_i}^2, \\
\tau_Y^2 &= \mathbb{E} \sum_{i=1}^p \|\phi_i(Y_{(i)}) - \phi_i(Y'_{(i)})\|_{\mathcal{H}_i}^2 \;=\; 2\,\mathbb{E} \sum_{i=1}^p \|\phi_i(Y_{(i)}) - \mathbb{E}\,\phi_i(Y_{(i)})\|_{\mathcal{H}_i}^2, \\
\tau^2 &= \mathbb{E} \sum_{i=1}^p \|\phi_i(X_{(i)}) - \phi_i(Y_{(i)})\|_{\mathcal{H}_i}^2 \;=\; \tau_X^2/2 + \tau_Y^2/2 + \zeta^2,
\end{aligned}
$$

where $\zeta^2 = \sum_{i=1}^p \|\mathbb{E}\,\phi(X_{(i)}) - \mathbb{E}\,\phi(Y_{(i)})\|_{\mathcal{H}_i}^2$. Thus $2\tau - \tau_X - \tau_Y = 0$ is equivalent to

$$
4(\tau_X^2/2 + \tau_Y^2/2 + \zeta^2) = (\tau_X + \tau_Y)^2 = \tau_X^2 + \tau_Y^2 + 2\tau_X\tau_Y.
$$

which implies that

$$
4\zeta^2 + (\tau_X - \tau_Y)^2 = 0.
$$

Therefore, $2\tau - \tau_X - \tau_Y = 0$ holds if and only if (1) $\zeta = 0$, i.e., $\mathbb{E}\,\phi_i(X_{(i)}) = \mathbb{E}\,\phi_i(Y_{(i)})$ for all $1 \leq i \leq p$, and, (2) $\tau_X = \tau_Y$, i.e.,

$$
\mathbb{E} \sum_{i=1}^p \|\phi_i(X_{(i)}) - \mathbb{E}\,\phi_i(X_{(i)})\|_{\mathcal{H}_i}^2 \;=\; \mathbb{E} \sum_{i=1}^p \|\phi_i(Y_{(i)}) - \mathbb{E}\,\phi_i(Y_{(i)})\|_{\mathcal{H}_i}^2.
$$

Now if $X \sim P$ and $Y \sim Q$, then note that

$$
\mathbb{E}\,\phi_i(X_{(i)}) \;=\; \int_{\mathbb{R}^{d_i}} k_i(\cdot, z)\,dP_i(z) \;=\; \Pi_i(P_i)
$$

$$\text{and} \quad \mathbb{E}\,\phi_i(Y_{(i)}) \;=\; \int_{\mathbb{R}^{d_i}} k_i(\cdot, z)\, dQ_i(z) \;=\; \Pi_i(Q_i)\,,$$

where Π_i is the mean embedding function (associated with the distance induced kernel k_i) defined in Section 2.1, P_i and Q_i are the distributions of $X_{(i)}$ and $Y_{(i)}$, respectively. As ρ_i is a metric of strong negative type on $\mathbb{R}^{d_i}$, the induced kernel k_i is characteristic to $\mathcal{M}_1(\mathbb{R}^{d_i})$ and hence the mean embedding function Π_i is injective. Therefore condition (1) above implies $X_{(i)} \overset{d}{=} Y_{(i)}$. $\diamondsuit$

Now we introduce some notation before presenting the proof of Theorem 4.2. The key of our analysis is to study the variance of the leading term of $\mathcal{E}_{n,m}(X, Y)$ in the HDLSS setup, propose the variance estimator and study the asymptotic behavior of the variance estimator. It will be shown later (in the proof of Theorem 4.2) that the leading term in the Taylor's expansion of $\mathcal{E}_{n,m}(X, Y) - (2\tau - \tau_X - \tau_Y)$ can be written as $L_1 + L_2$, where

$$
\begin{aligned}
L_1 \;&:=\; \frac{1}{nm\tau} \sum_{k=1}^{n}\sum_{l=1}^{m}\sum_{i=1}^{p} d_{kl}(i) - \frac{1}{n(n-1)\tau_X} \sum_{k<l}\sum_{i=1}^{p} d_{kl}^X(i) \\
&\qquad - \frac{1}{m(m-1)\tau_Y} \sum_{k<l}\sum_{i=1}^{p} d_{kl}^Y(i) \\
&:=\; L_1^1 - L_1^2 - L_1^3\,,
\end{aligned}
\tag{D.8}
$$

where L_1^i's are defined accordingly and

$$
\begin{aligned}
L_2 \;:=\;& \frac{1}{nm\tau} \sum_{k=1}^{n}\sum_{l=1}^{m}\sum_{i=1}^{p} \Big(\mathbb{E}\left[\rho_i(X_{k(i)}, Y_{l(i)})|X_{k(i)}\right] + \left[\rho_i(X_{k(i)}, Y_{l(i)})|Y_{l(i)}\right] \\
&\hspace{5cm} - 2\,\mathbb{E}\,\rho_i(X_{k(i)}, Y_{l(i)}) \Big) \\
&- \frac{1}{n(n-1)\tau_X} \sum_{k<l}\sum_{i=1}^{p} \Big(\mathbb{E}\left[\rho_i(X_{k(i)}, X_{l(i)})|X_{k(i)}\right] + \left[\rho_i(X_{k(i)}, X_{l(i)})|X_{l(i)}\right] \\
&\hspace{5cm} - 2\,\mathbb{E}\,\rho_i(X_{k(i)}, X_{l(i)}) \Big) \\
&- \frac{1}{m(m-1)\tau_Y} \sum_{k<l}\sum_{i=1}^{p} \Big(\mathbb{E}\left[\rho_i(Y_{k(i)}, Y_{l(i)})|Y_{k(i)}\right] + \left[\rho_i(Y_{k(i)}, Y_{l(i)})|Y_{l(i)}\right] \\
&\hspace{5cm} - 2\,\mathbb{E}\,\rho_i(Y_{k(i)}, Y_{l(i)}) \Big)\,.
\end{aligned}
\tag{D.9}
$$

By the double-centering properties, it is easy to see that L_1^i for $1 \leq i \leq 3$ are uncorrelated. Define

$$
V := \frac{1}{nm\tau^2} \sum_{i,i'=1}^{p} \mathbb{E}\left[d_{kl}(i)\, d_{kl}(i')\right] + \frac{1}{2n(n-1)\tau_X^2} \sum_{i,i'=1}^{p} \mathbb{E}\left[d_{kl}^X(i)\, d_{kl}^X(i')\right]
$$

$$+ \frac{1}{2m(m-1)\tau_Y^2} \sum_{i,i'=1}^{p} \mathbb{E}\left[d_{kl}^Y(i)\, d_{kl}^Y(i')\right] \tag{D.10}$$

$$:= V_1 \,+\, V_2 \,+\, V_3,$$

where V_i's are defined accordingly. Further let

$$\widetilde{V_1} \,:=\, nmV_1\,, \quad \widetilde{V_2} \,:=\, 2n(n-1)V_2\,, \quad \widetilde{V_3} \,:=\, 2m(m-1)V_3\,. \tag{D.11}$$

It can be verified that

$$\mathbb{E}\left[d_{kl}^X(i)\, d_{kl}^X(i')\right] \,=\, D_{\rho_i,\rho_{i'}}^2\left(X_{(i)}, X_{(i')}\right).$$

Thus we have

$$\widetilde{V_2} \,=\, \frac{1}{\tau_X^2} \sum_{i,i'=1}^{p} D_{\rho_i,\rho_{i'}}^2\left(X_{(i)}, X_{(i')}\right) \quad \text{and} \quad \widetilde{V_3} \,=\, \frac{1}{\tau_Y^2} \sum_{i,i'=1}^{p} D_{\rho_i,\rho_{i'}}^2\left(Y_{(i)}, Y_{(i')}\right). \tag{D.12}$$

We study the variances of L_1^i for $1 \le i \le 3$ and propose some suitable estimators. The variance for L_1^2 is given by

$$var(L_1^2) \,=\, \frac{1}{n^2(n-1)^2\tau_X^2} \sum_{i,i'=1}^{p} \sum_{k<l} \mathbb{E}\left[d_{kl}^X(i)\, d_{kl}^X(i')\right] \,=\, V_2\,.$$

Clearly

$$\frac{n(n-1)V_2}{2} \,=\, \frac{1}{4\tau_X^2} \sum_{i,i'=1}^{p} D_{\rho_i,\rho_j}^2\left(X_{(i)}, X_{(i')}\right).$$

From Theorem 5.3 in Section 5.1, we know that for fixed n and growing p, $\widetilde{\mathcal{D}_n^2}(X, X)$ is asymptotically equivalent to $\frac{1}{4\tau_X^2} \sum_{i,i'=1}^{p} \widetilde{D_n^2}_{\,;\,\rho_i,\rho_j}\left(X_{(i)}, X_{(i')}\right)$. Therefore an estimator of $\widetilde{V_2}$ is given by $4\,\widetilde{\mathcal{D}_n^2}(X, X)$. Note that the computational cost of $\widetilde{\mathcal{D}_n^2}(X, X)$ is linear in p while direct calculation of its leading term $\frac{1}{4\tau_X^2} \sum_{i,i'=1}^{p} \widetilde{D_n^2}_{\,;\,\rho_i,\rho_j}\left(X_{(i)}, X_{(i')}\right)$ requires computation in the quadratic order of p. Similarly it can be shown that the variance of L_1^3 is V_3 and $\widetilde{V_3}$ can be estimated by $4\,\widetilde{\mathcal{D}_m^2}(Y, Y)$. Likewise some easy calculations show that the variance of L_1^1 is V_1. Define

$$\hat{\rho}_i(X_{k(i)}, Y_{l(i)}) := \rho_i(X_{k(i)}, Y_{l(i)}) \,-\, \frac{1}{n} \sum_{a=1}^{n} \rho_i(X_{a(i)}, Y_{l(i)}) \,-\, \frac{1}{m} \sum_{b=1}^{m} \rho_i(X_{k(i)}, Y_{b(i)})$$

$$+ \frac{1}{nm} \sum_{a=1}^{n} \sum_{b=1}^{m} \rho_i(X_{a(i)}, Y_{b(i)})\,, \tag{D.13}$$

and

$$
\begin{aligned}
\hat{R}(X_k, Y_l) \ := \ & R(X_k, Y_l) - \frac{1}{n}\sum_{a=1}^{n} R(X_a, Y_l) - \frac{1}{m}\sum_{b=1}^{m} R(X_k, Y_b) \\
& + \frac{1}{nm}\sum_{a=1}^{n}\sum_{b=1}^{m} R(X_a, Y_b) \,.
\end{aligned} \tag{D.14}
$$

It can be verified that

$$
\hat{\rho}_i(X_{k(i)}, Y_{l(i)}) \ = \ d_{kl}(i) - \frac{1}{n}\sum_{a=1}^{n} d_{al}(i) - \frac{1}{m}\sum_{b=1}^{m} d_{kb}(i) + \frac{1}{nm}\sum_{a=1}^{n}\sum_{b=1}^{m} d_{ab}(i).
$$

Observe that

$$
\mathbb{E}\left[\hat{\rho}_i(X_{k(i)}, Y_{l(i)})\rho_{i'}(X_{k(i')}, Y_{l(i')})\right] \ = \ (1 - 1/n)(1 - 1/m)\,\mathbb{E}\left[d_{kl}(i)\,d_{kl}(i')\right]. \tag{D.15}
$$

Let $\hat{\mathbf{A}}_i = (\hat{\rho}_i(X_{k(i)}, Y_{l(i)}))_{k,l}$, $\mathbf{A}_i = (\rho_i(X_{k(i)}, Y_{l(i)}))_{k,l} \in \mathbb{R}^{n \times m}$. Note that

$$
\begin{aligned}
& \frac{1}{(n-1)(m-1)}\,\mathbb{E}\sum_{k=1}^{n}\sum_{l=1}^{m} \hat{\rho}_i(X_{k(i)}, Y_{l(i)})\hat{\rho}_i(X_{k(i')}, Y_{l(i')}) \\
&= \frac{1}{(n-1)(m-1)}\,\mathbb{E}\,\mathrm{tr}(\hat{\mathbf{A}}_i\hat{\mathbf{A}}_{i'}^{\top}) \\
&= \frac{1}{(n-1)(m-1)}\,\mathbb{E}\,\mathrm{tr}(\hat{\mathbf{A}}_i\mathbf{A}_{i'}^{\top}) \\
&= \frac{1}{(n-1)(m-1)}\,\mathbb{E}\sum_{k=1}^{n}\sum_{l=1}^{m} \rho_i(X_{k(i')}, Y_{l(i')})\,\hat{\rho}_i(X_{k(i)}, Y_{l(i)}) \\
&= \mathbb{E}\left[d_{kl}(i)\,d_{kl}(i')\right],
\end{aligned} \tag{D.16}
$$

which suggests that

$$
\check{V}_1 = \frac{1}{nm\tau^2}\sum_{i,i'=1}^{p} \frac{1}{(n-1)(m-1)}\sum_{k=1}^{n}\sum_{l=1}^{m} \hat{\rho}_i(X_{k(i)}, Y_{l(i)})\,\hat{\rho}_i(X_{k(i')}, Y_{l(i')})
$$

is an unbiased estimator for V_1. However, the computational cost for $\check{V}_1$ is linear in p^2 which is prohibitive for large p. We aim to find a joint metric whose computational cost is linear in p whose leading term is proportional to $\check{V}_1$. It can be verified that $cdCov_{n,m}^2(X, Y)$ is asymptotically equivalent to

$$
\frac{1}{4\tau^2}\sum_{i,i'=1}^{p} \frac{1}{(n-1)(m-1)}\sum_{k=1}^{n}\sum_{l=1}^{m} \hat{\rho}_i(X_{k(i)}, Y_{l(i)})\hat{\rho}_i(X_{k(i')}, Y_{l(i')})\,.
$$

This can be seen from the observation that

$$4\,cdCov^2_{n,m}(X,Y)$$

$$= \frac{1}{\tau^2} \sum_{i,i'=1}^{p} \frac{1}{(n-1)(m-1)} \sum_{k=1}^{n} \sum_{l=1}^{m} \hat{\rho}_i(X_{k(i)}, Y_{l(i)})\, \hat{\rho}_{i'}(X_{k(i')}, Y_{l(i')})$$

$$+ \frac{\tau^2}{(n-1)(m-1)} \sum_{k=1}^{n} \sum_{l=1}^{m} \hat{R}^2(X_k, Y_l) \tag{D.17}$$

$$+ \frac{1}{(n-1)(m-1)} \sum_{k=1}^{n} \sum_{l=1}^{m} \frac{1}{\tau} \sum_{i=1}^{p} \hat{\rho}_i(X_{k(i)}, Y_{(li)})\, \tau \hat{R}(X_k, Y_l).$$

Using the Hölder's inequality as well as the fact that $\tau^2\, \hat{R}^2(X_k, Y_l)$ is $O_p(\tau^2 a_p^4) = o_p(1)$ under Assumption 4.3. Therefore, we can estimate $\widetilde{V}_1$ by $4cdCov^2_{n,m}(X,Y)$. Thus the variance of L_1 is V which can be estimated by

$$\hat{V} := \frac{1}{nm}\, 4\,cdCov^2_{n,m}(X,Y) + \frac{1}{2n(n-1)}\, 4\widetilde{\mathcal{D}^2_n}(X,X)$$

$$+ \frac{1}{2m(m-1)}\, 4\widetilde{\mathcal{D}^2_m}(Y,Y) \tag{D.18}$$

$$:= \hat{V}_1 + \hat{V}_2 + \hat{V}_3\,.$$

Proof of Theorem 4.2. Using Proposition 4.1, some algebraic calculations yield

$$\mathcal{E}_{nm}(X,Y) - (2\tau - \tau_X - \tau_Y)$$

$$= \frac{\tau}{nm} \sum_{k=1}^{n} \sum_{l=1}^{m} L(X_k, Y_l) - \frac{\tau_X}{2n(n-1)} \sum_{k \neq l}^{n} L(X_k, X_l) - \frac{\tau_Y}{2m(m-1)} \sum_{k \neq l}^{m} L(Y_k, Y_l)$$

$$+ R_{n,m}$$

$$= \frac{1}{nm\tau} \sum_{k=1}^{n} \sum_{l=1}^{m} \sum_{i=1}^{p} \big(\rho_i(X_{k(i)}, Y_{l(i)}) - \mathbb{E}\, \rho_i(X_{k(i)}, Y_{l(i)})\big)$$

$$- \frac{1}{2n(n-1)\tau_X} \sum_{k \neq l}^{n} \sum_{i=1}^{p} \big(\rho_i(X_{k(i)}, X_{l(i)}) - \mathbb{E}\, \rho_i(X_{k(i)}, X_{l(i)})\big)$$

$$- \frac{1}{2m(m-1)\tau_Y} \sum_{k \neq l}^{m} \sum_{i=1}^{p} \big(\rho_i(Y_{k(i)}, Y_{l(i)}) - \mathbb{E}\, \rho_i(Y_{k(i)}, Y_{l(i)})\big) + R_{n,m},$$

where

$$R_{n,m} = \frac{2\tau}{nm} \sum_{k=1}^{n} \sum_{l=1}^{m} R(X_k, Y_l) - \frac{\tau_X}{n(n-1)} \sum_{k \neq l}^{n} R(X_k, X_l)$$

$$- \frac{\tau_Y}{m(m-1)} \sum_{k \neq l}^{m} R(Y_k, Y_l)\,. \tag{D.19}$$

By Assumption 4.3, $R_{n,m} = O_p(\tau a_p^2 + \tau_X b_p^2 + \tau_Y c_p^2) = o_p(1)$ as $p \to \infty$. Denote the leading term above by L. We can rewrite L as $L_1 + L_2$, where L_1 and L_2 are defined in equations (D.8) and (D.9), respectively. Some calculations yield that

$$
\begin{aligned}
L_2 &= \frac{1}{n} \sum_{k=1}^{n} \left[\frac{1}{\tau} \sum_{i=1}^{p} \mathbb{E}\left[\rho_i(X_{k(i)}, Y_{(i)}) | X_{k(i)} \right] - \frac{1}{\tau_X} \sum_{i=1}^{p} \mathbb{E}\left[\rho_i(X_{k(i)}, X'_{(i)}) | X_{k(i)} \right] \right] \\
&\quad - (\tau - \tau_X) \\
&\quad + \frac{1}{m} \sum_{l=1}^{m} \left[\frac{1}{\tau} \sum_{i=1}^{p} \mathbb{E}\left[\rho_i(X_{(i)}, Y_{l(i)}) | Y_{l(i)} \right] - \frac{1}{\tau_Y} \sum_{i=1}^{p} \mathbb{E}\left[\rho_i(Y_{l(i)}, Y'_{(i)}) | Y_{l(i)} \right] \right] \\
&\quad - (\tau - \tau_Y) \\
&= \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\left[\tau L(X_k, Y) - \tau_X L(X_k, X') \mid X_k \right] \\
&\quad + \frac{1}{m} \sum_{l=1}^{m} \mathbb{E}\left[\tau L(X, Y_l) - \tau_X L(Y_l, Y') \mid Y_l \right].
\end{aligned}
$$

(D.20)

For $(P_X, P_Y) \in \mathcal{P}$, we have $L_2 = o_p(1)$.

Under Assumption 4.4, the asymptotic distribution of L_1 as $p \to \infty$ is given by

$$
L_1 \xrightarrow{d} N\left(0, \frac{\sigma^2}{nm} + \frac{\sigma_X^2}{2n(n-1)} + \frac{\sigma_Y^2}{2m(m-1)} \right).
$$

Define the vector $d_{\text{vec}} := \left(\frac{1}{\tau} \sum_{i=1}^{p} d_{kl}(i) \right)_{1 \le k \le n, \, 1 \le l \le m}$. It can be verified that

$$
4(n-1)(m-1)\, cdCov_{n,m}^2(X,Y) = d_{\text{vec}}^\top A \, d_{\text{vec}}
$$

(D.21)

where $A = A_1 + A_2 + A_3 + A_4$ with $A_1 = I_n \otimes I_m$, $A_2 = -I_n \otimes \frac{1}{m} 1_m 1_m^\top$, $A_3 = -\frac{1}{n} 1_n 1_n^\top \otimes I_m$ and $A_4 = \frac{1}{nm} 1_{nm} 1_{nm}^\top$. Here $\otimes$ denotes the Kronecker product. It is not hard to see that $A^2 = A$ and $\text{rank}(A) = (n-1)(m-1)$. Therefore by Assumption 4.4, we have as $p \to \infty$,

$$
4(n-1)(m-1)\, cdCov_{n,m}^2(X,Y) \xrightarrow{d} \sigma^2 \chi_{(n-1)(m-1)}^2.
$$

By Theorem 5.5, we have as $p \to \infty$,

$$
4\widetilde{\mathcal{D}_n^2}(X,X) \xrightarrow{d} \frac{\sigma_X^2}{v_n} \chi_{v_n}^2, \quad \text{i.e.,} \quad 4 v_n \widetilde{\mathcal{D}_n^2}(X,X) \xrightarrow{d} \sigma_X^2 \chi_{v_n}^2,
$$

and similarly

$$
4 v_m \widetilde{\mathcal{D}_m^2}(Y,Y) \xrightarrow{d} \sigma_Y^2 \chi_{v_m}^2.
$$

By Assumption 4.4, $\chi_{(n-1)(m-1)}^2$, $\chi_{v_n}^2$ and $\chi_{v_m}^2$ are mutually independent. The proof can be completed by combining all the arguments above and using the continuous mapping theorem. $\diamond$

Proof of Proposition 4.2. Note that as $n, m \to \infty$,

$$\mathbb{E}\left[(M - m_0)^2\right] = \frac{2(n-1)(m-1)\sigma^4 + 2v_n \sigma_X^4 + 2v_m \sigma_Y^4}{\left\{(n-1)(m-1) + v_n + v_m\right\}^2} = o(1),$$

where $m_0 = \mathbb{E}[M]$. Therefore by Chebyshev's inequality, $M - m_0 = o_p(1)$ as $n, m \to \infty$. As a consequence, we have $M \xrightarrow{p} m_0^*$ as $n, m \to \infty$. Observing that Φ is a bounded function, the rest follows from Lebesgue's Dominated Convergence Theorem. $\Diamond$

Under H_0, without any loss of generality define $U_1 = X_1, \ldots, U_n = X_n, U_{n+1} := Y_1, \ldots, U_{n+m} := Y_m$. Further define

$$\phi_{i_1 i_2} := \phi(U_{i_1}, U_{i_2})$$

$$= \begin{cases} -\frac{1}{n(n-1)} H(U_{i_1}, U_{i_2}) & \text{if } i_1, i_2 \in \{1, \ldots, n\}, \\ \frac{1}{nm} H(U_{i_1}, U_{i_2}) & \text{if } i_1 \in \{1, \ldots, n\}, i_2 \in \{n+1, \ldots, n+m\}, \\ -\frac{1}{m(m-1)} H(U_{i_1}, U_{i_2}) & \text{if } i_1, i_2 \in \{n+1, \ldots, n+m\}. \end{cases}$$

(D.22)

It can be verified that $\operatorname{cov}(\phi_{i_1 i_2}, \phi_{i_1' i_2'}) = 0$ if the cardinality of the set $\{i_1, i_2\} \cap \{i_1', i_2'\}$ is less than 2. Define

$$\breve{T}_{n,m} = \frac{\mathcal{E}_{n,m}(X, Y)}{\sqrt{V}}.$$

Lemma D.1. *Under H_0 and Assumptions 4.5, B.1 and B.2, as n, m and $p \to \infty$, we have*

$$\breve{T}_{n,m} \xrightarrow{d} N(0, 1).$$

Proof of Lemma D.1. Set $N = n + m$. Define $V_{Nj} := \sum_{i=1}^{j-1} \phi_{ij}$ for $2 \le j \le N$, $S_{Nr} := \sum_{j=2}^r V_{Nj} = \sum_{j=2}^r \sum_{i=1}^{j-1} \phi_{ij}$ for $2 \le r \le N$, and $\mathcal{F}_{N,r} := \sigma(X_1, \ldots, X_r)$. Then the leading term of $\mathcal{E}_{nm}(X, Y)$, viz., L_1 (see equation (D.8)) can be expressed as

$$L_1 = S_{NN} = \sum_{j=2}^N V_{Nj} = \sum_{j=2}^N \sum_{i=1}^{j-1} \phi_{ij}$$

$$= \sum_{1 \le i_1 < i_2 \le n} \phi_{i_1 i_2} + \sum_{i_1=1}^n \sum_{i_2=n+1}^N \phi_{i_1 i_2} + \sum_{n+1 \le i_1 < i_2 \le N} \phi_{i_1 i_2}.$$

By Corollary 3.1 of [20], it suffices to show the following:

1. For each N, $\{S_{Nr}, \mathcal{F}_{N,r}\}_{r=1}^N$ is a sequence of zero mean and square integrable martingales,

2. $\frac{1}{V} \sum_{j=2}^N \mathbb{E}\left[V_{Nj}^2 \mid \mathcal{F}_{N,j-1}\right] \xrightarrow{P} 1$,

3. $\frac{1}{V} \sum_{j=2}^{N} \mathbb{E}\left[V_{Nj}^2 \, \mathbb{1}(|V_{Nj}| > \epsilon\sqrt{V}) \,|\, \mathcal{F}_{N,j-1} \right] \xrightarrow{P} 0, \quad \forall \, \epsilon > 0.$

To show (1), it is easy to see that S_{Nr} is square integrable, $\mathbb{E}(S_{Nr}) = \sum_{j=2}^{r} \sum_{i=1}^{j-1} \mathbb{E}(\phi_{ij}) = 0$, and, $\mathcal{F}_{N,1} \subseteq \mathcal{F}_{N,2} \subseteq \ldots \subseteq \mathcal{F}_{N,N}$. We only need to show $\mathbb{E}(S_{Nq}\,|\,\mathcal{F}_{N,r}) = S_{Nr}$ for $q > r$. Now $\mathbb{E}(S_{Nq}\,|\,\mathcal{F}_{N,r}) = \sum_{j=2}^{q} \sum_{i=1}^{j-1} \mathbb{E}(\phi_{ij}\,|\,\mathcal{F}_{N,r})$. If $j \le r < q$ and $i < j$, then $\mathbb{E}(\phi_{ij}\,|\,\mathcal{F}_{N,r}) = \phi_{ij}$. If $r < j \le q$, then:

   (i) if $r < i < j \le q$, then $\mathbb{E}(\phi_{ij}\,|\,\mathcal{F}_{N,r}) = \mathbb{E}(\phi_{ij}) = 0$,

  (ii) if $i \le r < j \le q$, then $\mathbb{E}(\phi_{ij}\,|\,\mathcal{F}_{N,r}) = 0$ (due to $\mathcal{U}$-centering).

Therefore $\mathbb{E}(S_{Nq}\,|\,\mathcal{F}_{N,r}) = S_{Nr}$ for $q > r$. This completes the proof of (1).

To show (2), define $L_j(i,k) := \mathbb{E}\left[\phi_{ij}\,\phi_{kj}\,|\,\mathcal{F}_{N,j-1}\right]$ for $i,k < j \le N$, and

$$\eta_N := \sum_{j=2}^{N} \mathbb{E}\left[V_{Nj}^2 \,|\, \mathcal{F}_{N,j-1} \right] = \sum_{j=2}^{N} \sum_{i,k=1}^{j-1} \mathbb{E}[\phi_{ij}\,\phi_{kj}\,|\,\mathcal{F}_{N,j-1}] = \sum_{j=2}^{N} \sum_{i,k=1}^{j-1} L_j(i,k).$$

Note that $\mathbb{E}\left[L_j(i,k)\right] = 0$ for $i \ne k$. Clearly

$$\mathbb{E}[\eta_N] = \sum_{j=2}^{N} \mathbb{E}[V_{Nj}^2] = \sum_{j=2}^{N} \sum_{i,k=1}^{j-1} \mathbb{E}[\phi_{ij}\,\phi_{kj}] = \sum_{j=2}^{N} \sum_{i=1}^{j-1} \mathbb{E}[\phi_{ij}^2] = V. \quad \text{(D.23)}$$

By virtue of Chebyshev's inequality, it will suffice to show $\text{var}(\frac{\eta_N}{V}) = o(1)$. Note that

$$\mathbb{E}\left[L_j(i,k)\,L_{j'}(i',k')\right]$$
$$= \begin{cases} \mathbb{E}\left[\phi^2(U_i,U_j)\phi^2(U_i,U_{j'}')\right] & i = k = i' = k', \\ \mathbb{E}\left[\phi(U_i,U_j)\phi(U_k,U_j)\phi(U_i,U_{j'}')\phi(U_k,U_{j'}')\right] & i = i' \ne k = k' \\ & \text{or } i = k' \ne k = i', \\ \mathbb{E}\left[\phi^2(U_i,U_j)\right]\mathbb{E}\left[\phi^2(U_{i'},U_{j'})\right] & i = k \ne i' = k'. \end{cases} \quad \text{(D.24)}$$

In view of equation (D.22), it can be verified that the above expression for $\mathbb{E}\,L_j(i,k)\,L_{j'}(i',k')$ holds true for $j = j'$ as well. Therefore

$$\text{var}\left(\eta_N^2\right) = \sum_{j,j'=2}^{N} \sum_{i,k=1}^{j-1} \sum_{i',k'=1}^{j'-1} \text{cov}\left(L_j(i,k),\,L_{j'}(i',k')\right)$$
$$= \sum_{j=j'} \left\{ \sum_{i=1}^{j-1} \text{cov}\left(\phi^2(U_i,U_j),\phi^2(U_i,U_j')\right) \right.$$
$$\left. + 2\sum_{\substack{i \ne k}}^{j-1} \mathbb{E}\left[\phi(U_i,U_j)\phi(U_k,U_j)\phi(U_i,U_j')\phi(U_k,U_j')\right] \right\}$$

$$+ \ 2 \sum_{2 \leq j < j' \leq N} \left\{ \sum_{i=1}^{j-1} \text{cov} \left(\phi^2(U_i, U_j), \phi^2(U_i, U'_{j'}) \right) \right.$$

$$\left. + \ 2 \sum_{i \neq k}^{j-1} \mathbb{E} \left[\phi(U_i, U_j) \phi(U_k, U_j) \phi(U_i, U'_{j'}) \phi(U_k, U'_{j'}) \right] \right\}.$$

Under Assumption 4.5 and H_0, it can be verified that

$$\text{var}(\eta_N) \ = \ O\Big(\frac{1}{N^5} \mathbb{E} \left[H^2(X, X'') H^2(X', X'') \right] \tag{D.25}$$
$$+ \frac{1}{N^4} \mathbb{E} \left[H(X, X'') H(X', X'') H(X, X''') H(X', X''') \right] \Big),$$

and

$$V^2 \ \asymp \ \frac{1}{N^4} \left(\mathbb{E} \left[H^2(X, X') \right] \right)^2 . \tag{D.26}$$

Therefore under Assumption B.1 and H_0, we have

$$\text{var} \left(\frac{\eta_N}{V} \right) = o(1),$$

which completes the proof of (2). To show (3), note that it suffices to show

$$\frac{1}{V^2} \sum_{j=2}^{N} \mathbb{E} \left[V_{Nj}^4 \mid \mathcal{F}_{N, j-1} \right] \ \xrightarrow{P} \ 0 .$$

Observe that

$$\sum_{j=2}^{N} \mathbb{E} \left[V_{Nj}^4 \right] \ = \ \sum_{j=2}^{N} \mathbb{E} \left(\sum_{i=1}^{j-1} \phi_{ij} \right)^4$$

$$= \ \sum_{j=2}^{N} \sum_{i=1}^{j-1} \mathbb{E}[\phi^4(U_i, U_j)] \ + \ 3 \sum_{j=2}^{N} \sum_{i_1 \neq i_2}^{j-1} \mathbb{E}[\phi^2(U_{i_1}, U_j) \, \phi^2(U_{i_2}, U_j)].$$

Under Assumption 4.5, we have

$$\sum_{j=2}^{N} \mathbb{E} \left[V_{Nj}^4 \right] \ = \ O \left(\frac{1}{N^6} \mathbb{E} \left[H^4(X, X') \right] \ + \ \frac{1}{N^5} \mathbb{E} \left[H^2(X, X'') H^2(X', X'') \right] \right).$$

This along with the observation from equation (D.25) and Assumption B.1 complete the proof of (3).

Finally to see that $\frac{R_{n,m}}{\sqrt{V}} = o_p(1)$, note that from equation (D.19) we can derive using power mean inequality that $\mathbb{E} \, R_{n,m}^2 \leq C \, \tau^2 \, \mathbb{E} \left[R^2(X, X') \right]$ for some

positive constant C. Using this, equation (D.26), Chebyshev's inequality and Hölder's inequality, we have for any $\epsilon > 0$

$$
\begin{aligned}
P\left(\left|\frac{R_{n,m}}{\sqrt{V}}\right| > \epsilon\right) \;\leq\; & \frac{\mathbb{E}\,R_{n,m}^2}{\epsilon^2\,V} \;\leq\; C'\,\frac{N^2\,\tau^2\,\mathbb{E}\left[R^2(X,X')\right]}{\epsilon^2\,\mathbb{E}\left[H^2(X,X')\right]} \\
\leq\; & \frac{C'}{\epsilon^2}\left(\frac{N^4\,\tau^4\,\mathbb{E}\left[R^4(X,X')\right]}{\left(\mathbb{E}\left[H^2(X,X')\right]\right)^2}\right)^{1/2},
\end{aligned}
\tag{D.27}
$$

for some positive constant C'. From this and Assumptions 4.5 and B.2, we get $\frac{R_{n,m}}{\sqrt{V}} = o_p(1)$, as $N \asymp n$. This completes the proof of the lemma. $\diamond$

Lemma D.2. *Under H_0 and Assumptions 4.5 and B.2, as n, m and $p \to \infty$, we have*

$$
\frac{\left|\mathbb{E}\left[\hat{V}_i\right] - V_i\right|}{V_i} = o(1) \;,\quad 1 \leq i \leq 3\,,
$$

where V_i and $\hat{V}_i$, $1 \leq i \leq 3$ are defined in equations (D.10) and (D.18), respectively in the appendix.

Proof of Lemma D.2. We first deal with $\hat{V}_2$. Note that

$$
\widetilde{\mathcal{D}}_n^2(X,X) \;=\; \frac{1}{n(n-3)}\sum_{k\neq l}\left(\widetilde{D}_{kl}^X\right)^2,
$$

where

$$
\begin{aligned}
\widetilde{D}_{kl}^X \;=\; & K(X_k, X_l) \;-\; \frac{1}{n-2}\sum_{b=1}^{n} K(X_k, X_b) \;-\; \frac{1}{n-2}\sum_{a=1}^{n} K(X_a, X_l) \\
& +\; \frac{1}{(n-1)(n-2)}\sum_{a,b=1}^{n} K(X_a, X_b) \\
\;=\; & \frac{1}{2\tau}\sum_{i=1}^{p}\widetilde{\rho}_i(X_{k(i)}, X_{l(i)}) \;+\; \tau\widetilde{R}(X_k, X_l)\,,
\end{aligned}
\tag{D.28}
$$

using Proposition 4.1. As a consequence, we can write

$$
\begin{aligned}
\widetilde{\mathcal{D}}_n^2(X,X) \;=\; & \frac{1}{4\tau^2}\sum_{i,i'=1}^{p}\widetilde{D}_{n\,;\,\rho_i,\rho_{i'}}^2(X_{(i)}, X_{(i')}) \;+\; \frac{\tau^2}{n(n-3)}\sum_{k\neq l}\widetilde{R}^2(X_k, X_l) \\
& +\; \frac{1}{n(n-3)}\sum_{k\neq l}\frac{1}{\tau}\sum_{i=1}^{p}\widetilde{\rho}_i(X_{k(i)}, X_{(li)})\,\tau\widetilde{R}(X_k, X_l)\,.
\end{aligned}
\tag{D.29}
$$

Note that following Step 3 in Section 1.6 in the supplementary material of [54], we can write

$$
\widetilde{R}(X_k, X_l) \;=\; \frac{n-3}{n-1}\bar{R}(X_k, X_l) \;-\; \frac{n-3}{(n-1)(n-2)}\sum_{b\notin\{k,l\}}\bar{R}(X_k, X_b)
$$

$$- \frac{n-3}{(n-1)(n-2)} \sum_{a \notin \{k,l\}} \bar{R}(X_a, X_l)$$

$$+ \frac{1}{(n-1)(n-2)} \sum_{a,b \notin \{k,l\}} \bar{R}(X_a, X_b),$$

where $\bar{R}(X, X') = R(X, X') - E[R(X, X')|X] - E[R(X, X')|X'] + E[R(X, X')]$. Using the power mean inequality, it can be verified that $\mathbb{E}\left[\widetilde{R}^2(X_k, X_l)\right] \leq C\,\mathbb{E}\left[\bar{R}^2(X_k, X_l)\right]$ for some positive constant C. Using this and the Hölder's inequality, the expectation of the third term in the summation in equation (D.29) can be bounded as follows

$$\left| \mathbb{E}\left[\frac{1}{n(n-3)} \sum_{k \neq l} \frac{1}{\tau} \sum_{i=1}^{p} \widetilde{\rho}_i(X_{k(i)}, X_{l(i)}) \, \tau \widetilde{R}(X_k, X_l) \right] \right|$$

$$\leq \frac{1}{n(n-3)} \sum_{k \neq l} \left(\mathbb{E}\left[\left(\frac{1}{\tau} \sum_{i=1}^{p} \widetilde{\rho}_i(X_{k(i)}, X_{l(i)}) \right)^2 \right] \tau^2 \, \mathbb{E}\left[\bar{R}^2(X_k, X_l) \right] \right)^{1/2}$$

$$\leq C' \left(\left(\frac{1}{\tau^2} \sum_{i,i'=1}^{p} D_{\rho_i, \rho_{i'}}^2(X_{(i)}, X_{(i')}) \right) \tau^2 \, \mathbb{E}\left[\bar{R}^2(X, X') \right] \right)^{1/2}$$

for some positive constant C'. Combining all the above, we get

$$|\mathbb{E}(\hat{V}_2) - V_2| \leq \frac{C_1}{n(n-1)} \tau^2 \, \mathbb{E}\,\bar{R}^2(X, X')$$

$$+ \frac{C_2}{n(n-1)} \left(\left(\frac{1}{\tau^2} \sum_{i,i'=1}^{p} D_{\rho_i, \rho_{i'}}^2(X_{(i)}, X_{(i')}) \right) \tau^2 \, \mathbb{E}\left[\bar{R}^2(X, X') \right] \right)^{1/2},$$

for some positive constants C_1 and C_2. As $V_2 = \frac{1}{2n(n-1)} E[H^2(X, X')]$,

$$\frac{\left| \mathbb{E}[\hat{V}_2] - V_2 \right|}{V_2} = o(1) \quad \text{is satisfied if} \quad \frac{\tau^2 \, \mathbb{E}\left[\bar{R}^2(X, X') \right]}{\mathbb{E}[H^2(X, X')]} = o(1).$$

Using power mean inequality and Jensen's inequality, it is not hard to verify that $\mathbb{E}\left[\bar{R}^4(X, X')\right] = O\left(\mathbb{E}\left[R^4(X, X')\right]\right)$. Using this and Hölder's inequality, we have

$$\frac{\tau^2 \, \mathbb{E}\left[\bar{R}^2(X, X') \right]}{\mathbb{E}[H^2(X, X')]} = O\left(\left(\frac{\tau^4 \, \mathbb{E}\left[R^4(X, X') \right]}{(\mathbb{E}\left[H^2(X, X') \right])^2} \right)^{1/2} \right).$$

Clearly Assumption B.2 implies $\frac{\tau^4 \, \mathbb{E}[R^4(X,X')]}{(\mathbb{E}[H^2(X,X')])^2} = o(1)$, which in turn implies

$$\frac{\tau^2 \, \mathbb{E}\left[\bar{R}^2(X, X') \right]}{\mathbb{E}[H^2(X, X')]} = o(1).$$

Similar expressions can be derived for $\hat{V}_3$ as well. For the term involving $\hat{V}_1$, in the similar fashion, we can write

$$
\begin{aligned}
\mathbb{E}&\left[4\,cdCov_{n,m}^2(X,Y)\right]\\
=\ &\frac{1}{\tau^2}\sum_{i,i'=1}^{p}\frac{1}{(n-1)(m-1)}\sum_{k=1}^{n}\sum_{l=1}^{m}\mathbb{E}\left[\hat{\rho}_i(X_{k(i)},Y_{l(i)})\,\hat{\rho}_{i'}(X_{k(i')},Y_{l(i')})\right]\\
&+\ \tau^2\frac{1}{(n-1)(m-1)}\sum_{k=1}^{n}\sum_{l=1}^{m}\mathbb{E}\left[\hat{R}^2(X_k,Y_l)\right]\\
&+\ \frac{1}{(n-1)(m-1)}\sum_{k=1}^{n}\sum_{l=1}^{m}\frac{1}{\tau}\sum_{i=1}^{p}\mathbb{E}\left[\hat{\rho}_i(X_{k(i)},Y_{(li)})\,\tau\hat{R}(X_k,Y_l)\right],
\end{aligned}
\tag{D.30}
$$

where the expression for $\hat{R}(X_k,Y_l)$ is given in equation (D.14). Following equation (D.16) we can write

$$
\begin{aligned}
&\frac{1}{\tau^2}\sum_{i,i'=1}^{p}\frac{1}{(n-1)(m-1)}\sum_{k=1}^{n}\sum_{l=1}^{m}\mathbb{E}\left[\hat{\rho}_i(X_{k(i)},Y_{l(i)})\,\hat{\rho}_{i'}(X_{k(i')},Y_{l(i')})\right]\\
&=\ \mathbb{E}\left[H^2(X,Y)\right].
\end{aligned}
$$

Therefore in view of equations (D.10), (D.15) and (D.18), using the power mean inequality we can write

$$
\begin{aligned}
&|\mathbb{E}(\hat{V}_1)-V_1|\\
&\leq\frac{C_1'}{nm}\tau^2\,\mathbb{E}\,\bar{R}^2(X,Y)+\frac{C_2'}{nm}\left(\left(\frac{1}{\tau^2}\sum_{i,i'=1}^{p}\mathbb{E}\left[d_{kl}(i)d_{kl}(i')\right]\right)\tau^2\,\mathbb{E}\left[\bar{R}^2(X,Y)\right]\right)^{1/2},
\end{aligned}
$$

for some positive constants C_1' and C_2'. Then under H_0 and Assumptions 4.5 and B.2, we have

$$
\frac{\left|\mathbb{E}(\hat{V}_1)-V_1\right|}{V_1}=o(1).\diamondsuit
$$

Lemma D.3. *Under H_0 and Assumptions 4.5, B.1 and B.2, as n,m and $p\to\infty$, we have*

$$
\frac{var(\hat{V}_i)}{V_i^2}=o(1),\ \ 1\leq i\leq 3.
$$

Proof of Lemma D.3. Again we deal with $\hat{V}_2$ first. To simplify the notations,

denote $A_{ij} = K(X_i, X_j)$ and $\widetilde{A}_{ij} = \widetilde{D}_{ij}^X$ for $1 \leq i \neq j \leq n$. Observe that

$$
\text{var}\left(\widetilde{\mathcal{D}}_n^2(X, X)\right) = \text{var}\left(\frac{1}{n(n-3)} \sum_{i \neq j} \widetilde{A}_{ij}^2\right)
$$

$$
\asymp \frac{1}{n^4} \left[\sum_{i<j} \text{var}(\widetilde{A}_{ij}^2) + \sum_{i<j<j'} \text{cov}(\widetilde{A}_{ij}^2, \widetilde{A}_{jj'}^2) + \sum_{\substack{i<j, i'<j' \\ \{i,j\} \cap \{i',j'\}=\phi}} \text{cov}(\widetilde{A}_{ij}^2, \widetilde{A}_{i'j'}^2)\right]. \tag{D.31}
$$

As in the proof of Lemma D.2, we can write

$$
\widetilde{A}_{ij} = \frac{n-3}{n-1} \bar{A}_{ij} - \frac{n-3}{(n-1)(n-2)} \sum_{l \notin \{i,j\}} \bar{A}_{il} - \frac{n-3}{(n-1)(n-2)} \sum_{k \notin \{i,j\}} \bar{A}_{kj}
$$

$$
+ \frac{1}{(n-1)(n-2)} \sum_{k,l \notin \{i,j\}} \bar{A}_{kl}, \tag{D.32}
$$

where the four summands are uncorrelated with each other. Using the power mean inequality, it can be shown that

$$
\mathbb{E}(\widetilde{A}_{ij}^4) \leq C \mathbb{E}(\bar{A}_{ij}^4) = C \mathbb{E}\left[\bar{K}^4(X, X')\right],
$$

for some positive constant C, where $\bar{K}(X, X') = K(X, X') - E[K(X, X')|X] - E[K(X, X')|X'] + E[K(X, X')]$ (similarly define $\bar{L}(X, X')$). Therefore the first summand in equation (D.31) scaled by $\widetilde{V}_2^2$ is $o(1)$ as $n, p \to \infty$, provided

$$
\frac{1}{n^2} \frac{\mathbb{E}\left[\bar{K}^4(X, X')\right]}{\widetilde{V}_2^2} = o(1),
$$

where $\widetilde{V}_2$ is defined in equations (D.11) and (D.12). Note that

$$
\bar{K}(X, X') = \frac{\tau_X}{2} \bar{L}(X, X') + \tau_X \bar{R}(X, X').
$$

Using the power mean inequality we can write

$$
\frac{1}{n^2} \frac{\mathbb{E}\left[\bar{K}^4(X, X')\right]}{(\mathbb{E}\left[H^2(X, X')\right])^2} \leq C_0 \frac{1}{n^2} \frac{\tau_X^4 \mathbb{E}\left[\bar{L}^4(X, X')\right]}{(\mathbb{E}\left[H^2(X, X')\right])^2} + C_0' \frac{1}{n^2} \frac{\tau_X^4 \mathbb{E}\left[\bar{R}^4(X, X')\right]}{(\mathbb{E}\left[H^2(X, X')\right])^2}
$$

for some positive constants C_0 and C_0'. It is easy to see that

$$
\bar{L}(X_k, X_l) = \frac{1}{\tau_X^2} \bar{K}^2(X_k, X_l) = \frac{1}{\tau_X^2} \sum_{i=1}^p d_{kl}^X(i) = \frac{1}{\tau_X} H(X_k, X_l). \tag{D.33}
$$

From equation (D.33) it is easy to see that the condition

$$\frac{1}{n^2}\frac{\tau_X^4\,\mathbb{E}\left[\bar{L}^4(X,X')\right]}{\left(\mathbb{E}\left[H^2(X,X')\right]\right)^2}=o(1)\quad\text{is equivalent to}\quad\frac{1}{n^2}\frac{\mathbb{E}\left[H^4(X,X')\right]}{\left(\mathbb{E}\left[H^2(X,X')\right]\right)^2}=o(1).$$

For the third summand in equation (D.31), observe that

$$
\begin{aligned}
\widetilde{A}_{ij}^2 &= O(1)\bar{A}_{ij}^2+O\left(\frac{1}{n^2}\right)\sum_{l,l'\notin\{i,j\}}\bar{A}_{il}\bar{A}_{il'}+O\left(\frac{1}{n^2}\right)\sum_{k,k'\notin\{i,j\}}\bar{A}_{kj}\bar{A}_{k'j}\\
&\quad+O\left(\frac{1}{n^4}\right)\sum_{k,k',l,l'\notin\{i,j\}}\bar{A}_{kl}\bar{A}_{k'l'}+O\left(\frac{1}{n}\right)\bar{A}_{ij}\sum_{l\notin\{i,j\}}\bar{A}_{il}\\
&\quad+O\left(\frac{1}{n}\right)\bar{A}_{ij}\sum_{k\notin\{i,j\}}\bar{A}_{kj}+O\left(\frac{1}{n^2}\right)\bar{A}_{ij}\sum_{k,l\notin\{i,j\}}\bar{A}_{kl}\\
&\quad+O\left(\frac{1}{n^2}\right)\sum_{k,l\notin\{i,j\}}\bar{A}_{il}\bar{A}_{kj}+O\left(\frac{1}{n^3}\right)\sum_{k,l,l'\notin\{i,j\}}\bar{A}_{il}\bar{A}_{kl'}\\
&\quad+O\left(\frac{1}{n^3}\right)\sum_{k,k',l\notin\{i,j\}}\bar{A}_{kl}\bar{A}_{k'j}\,.
\end{aligned}
$$

$$(D.34)$$

Likewise $\widetilde{A}_{i'j'}^2$ admits a similar expression as in equation (D.34). We claim that when $\{i,j\}\cap\{i',j'\}=\phi$, the leading term of $\mathrm{cov}(\widetilde{A}_{ij}^2,\widetilde{A}_{i'j'}^2)$ is $O\left(\frac{1}{n^2}\,\mathbb{E}\,(\bar{A}_{ij}^4)\right)$. To see this first note that $\bar{A}_{ij}$ is independent of $\bar{A}_{i'j'}$ when $\{i,j\}\cap\{i',j'\}=\phi$. Using the double-centering properties, it can be verified that

$$
\begin{aligned}
\mathrm{cov}\left(\bar{A}_{i'j'}^2,\bar{A}_{ij}\sum_{l\notin\{i,j\}}\bar{A}_{il}\right) &= \mathrm{cov}\left(\bar{A}_{i'j'}^2,\bar{A}_{ij}\sum_{k\notin\{i,j\}}\bar{A}_{kj}\right)\\
&= \mathrm{cov}\left(\bar{A}_{i'j'}^2,\bar{A}_{ij}\sum_{k,l\notin\{i,j\}}\bar{A}_{kl}\right)=0.
\end{aligned}
$$

To compute the quantity $\mathrm{cov}\left(\bar{A}_{i'j'}^2\,,\,O\left(\frac{1}{n^2}\right)\sum_{l,l'\notin\{i,j\}}\bar{A}_{il}\bar{A}_{il'}\right)$, consider the following cases:

Case 1. When $l=l'=i'$ or $l=l'=j'$ or $l=i',l'=j'$, $\mathrm{cov}\left(\bar{A}_{i'j'}^2,\bar{A}_{il}\bar{A}_{il'}\right)$ boils down to $\mathrm{cov}(\bar{A}_{i'j'}^2,\bar{A}_{ii'}^2)$ or $\mathrm{cov}(\bar{A}_{i'j'}^2,\bar{A}_{ij'}^2)$ or $\mathrm{cov}(\bar{A}_{i'j'}^2,\bar{A}_{ii'}\bar{A}_{ij'})$.

Case 2. When $l=i,l'\notin\{i,j,i',j'\}$ or $l=j',l'\notin\{i,j,i',j'\}$, $\mathrm{cov}\left(\bar{A}_{i'j'}^2,\bar{A}_{il}\bar{A}_{il'}\right)$ boils down to $\mathrm{cov}(\bar{A}_{i'j'}^2,\bar{A}_{ii'}\bar{A}_{il'})$ or $\mathrm{cov}(\bar{A}_{i'j'}^2,\bar{A}_{ij'}\bar{A}_{il'})$, which can be easily verified to be zero.

Case 3. When $\{l,l'\}\cap\{i',j'\}=\phi$, $\mathrm{cov}\left(\bar{A}_{i'j'}^2,\bar{A}_{il}\bar{A}_{il'}\right)$ is again zero.

Similar arguments can be made about

$$\mathrm{cov}\left(\bar{A}_{i'j'}^2\,,\,O\left(\frac{1}{n^2}\right)\sum_{k,k'\notin\{i,j\}}\bar{A}_{kj}\bar{A}_{k'j}\right)$$

and

$$\mathrm{cov}\left(\bar{A}_{i'j'}^2\,,\,O\left(\frac{1}{n^2}\right)\sum_{k,l\notin\{i,j\}}\bar{A}_{il}\bar{A}_{kj}\right).$$

With this and using Hölder's inequality, it can be verified that when $\{i,j\}\cap\{i',j'\}=\phi$, the leading term of $\mathrm{cov}(\widetilde{A}_{ij}^2,\widetilde{A}_{i'j'}^2)$ is $O\left(\frac{1}{n^2}\,\mathbb{E}\,(\bar{A}_{ij}^4)\right)$. Therefore the third summand in equation (D.31) scaled by $\widetilde{V_2}^2$ can be argued to be $o(1)$ in similar lines of the argument for the first summand in equation (D.31).

For the second summand in equation (D.31), in the similar line we can argue that the leading term of $\mathrm{cov}(\widetilde{A}_{ij}^2,\widetilde{A}_{jj'}^2)$ is

$$O\left(\frac{1}{n}\right)\mathbb{E}\left[\bar{A}_{ij}^4\right]\,+\,O(1)\,\mathbb{E}\left[\bar{A}_{ij}^2\bar{A}_{jj'}^2\right].$$

Therefore the leading term of $\frac{1}{n^4}\sum_{i<j<j'}\mathrm{cov}(\widetilde{A}_{ij}^2,\widetilde{A}_{jj'}^2)$ is

$$O\left(\frac{1}{n^2}\right)\mathbb{E}\left[\bar{A}_{ij}^4\right]\,+\,O\left(\frac{1}{n}\right)\mathbb{E}\left[\bar{A}_{ij}^2\bar{A}_{jj'}^2\right].$$

For the second term above, using the power mean inequality we can write

$$\frac{1}{n}\frac{\mathbb{E}\left[\bar{A}_{ij}^2\,\bar{A}_{jj'}^2\right]}{(\mathbb{E}\left[H^2(X,X')\right])^2}\;\le\;C_3\frac{1}{n}\frac{\tau^4\,\mathbb{E}\left[\bar{L}^2(X,X')\,\bar{L}^2(X',X'')\right]}{(\mathbb{E}\left[H^2(X,X')\right])^2}$$
$$+\,C_3'\frac{1}{n}\frac{\tau^4\,\mathbb{E}\left[\bar{L}^2(X,X')\,\bar{R}^2(X',X'')\right]}{(\mathbb{E}\left[H^2(X,X')\right])^2}+C_3''\frac{1}{n}\frac{\tau^4\,\mathbb{E}\left[\bar{R}^2(X,X')\,\bar{R}^2(X',X'')\right]}{(\mathbb{E}\left[H^2(X,X')\right])^2}$$
$$=\,C_3\frac{1}{n}\frac{\mathbb{E}\left[H^2(X,X')\,H^2(X',X'')\right]}{(\mathbb{E}\left[H^2(X,X')\right])^2}$$
$$+\,C_3'\frac{1}{n}\frac{\tau^2\,\mathbb{E}\left[H^2(X,X')\,\bar{R}^2(X',X'')\right]}{(\mathbb{E}\left[H^2(X,X')\right])^2}+C_3''\frac{1}{n}\frac{\tau^4\,\mathbb{E}\left[\bar{R}^2(X,X')\,\bar{R}^2(X',X'')\right]}{(\mathbb{E}\left[H^2(X,X')\right])^2}$$

for some positive constants C_3,C_3' and C_3''. Using Hölder's inequality it can be seen that the second summand in equation (D.31) scaled by $\widetilde{V_2}^2$ is $o(1)$ as $n,p\to\infty$ under Assumptions B.1 and B.2. This completes the proof that

$$\frac{\mathrm{var}(\hat{V}_2)}{V_2^2}=o(1).$$

A similar line of argument and the simple observation that

$$
\begin{aligned}
\hat{K}(X_k, Y_l) &= K(X_k, Y_l) - \frac{1}{n} \sum_{a=1}^{n} K(X_a, Y_l) - \frac{1}{m} \sum_{b=1}^{m} K(X_k, Y_b) \\
&\quad + \frac{1}{nm} \sum_{a=1}^{n} \sum_{b=1}^{m} K(X_a, Y_b) \\
&= \bar{K}(X_k, Y_l) - \frac{1}{n} \sum_{a=1}^{n} \bar{K}(X_a, Y_l) - \frac{1}{m} \sum_{b=1}^{m} \bar{K}(X_k, Y_b) \\
&\quad + \frac{1}{nm} \sum_{a=1}^{n} \sum_{b=1}^{m} \bar{K}(X_a, Y_b)
\end{aligned}
$$

will show that under Assumptions 4.5, B.1 and B.2,

$$
\frac{\text{var}(\hat{V}_1)}{V_1^2} = o(1) \qquad \text{and} \qquad \frac{\text{var}(\hat{V}_3)}{V_3^2} = o(1) \,. \Diamond
$$

Lemma D.4. *Under H_0 and Assumptions 4.5, B.1 and B.2, as n, m and $p \to \infty$, we have $\hat{V}/V \xrightarrow{P} 1$.*

Proof. It is enough to show that

$$
\mathbb{E} \left[\left(\frac{\hat{V}}{V} - 1 \right)^2 \right] = o(1) \,, \quad \text{i.e.,} \quad \frac{\text{var}(\hat{V}) + \left(\mathbb{E}\,[\hat{V}] - V \right)^2}{V^2} = o(1) \,.
$$

It suffices to show the following

$$
\frac{\text{var}(\hat{V}_i)}{V_i^2} = o(1) \quad \text{and} \quad \frac{\left(\mathbb{E}\,[\hat{V}_i] - V_i \right)^2}{V_i^2} = o(1), \quad 1 \le i \le 3.
$$

The proof can be completed using Lemmas D.2 and D.3. $\Diamond$

Proof of Theorem B.1. The proof essentially follows from Lemma D.1 and D.4. $\Diamond$

Proof of Proposition A.2. The proof of the first part follows similar lines of the proof of Proposition 1 in [49], replacing the Euclidean distance between X and X', viz. $\|X - X'\|_{\tilde{p}}$, by $K(X, X')$. The second part of the proposition has a proof similar to Lemma 2.1 in [52] and Section 1.1 in the Supplement of [52]. $\Diamond$

Proof of Theorem A.4. The first two parts of the theorem immediately follow from Proposition 2.6 and Theorem 2.7 in [26], respectively and the parallel U-statistics theory (see for example [41]). The third part follows from the first part and the fact that $\mathcal{D}$ is non-zero for two dependent random vectors. $\Diamond$

Proof of Theorem 5.1. Following the definition of $\mathcal{D}(X,Y)$ and applying Proposition 4.1, we can write

$$
\frac{1}{\tau_{XY}}\mathcal{D}^2(X,Y)
$$

$$
=\mathbb{E}\,\frac{K(X,X')}{\tau_X}\,\frac{K(Y,Y')}{\tau_Y} \;+\; \mathbb{E}\,\frac{K(X,X')}{\tau_X}\,\mathbb{E}\,\frac{K(Y,Y')}{\tau_Y} \;-\; 2\,\mathbb{E}\,\frac{K(X,X')}{\tau_X}\,\frac{K(Y,Y'')}{\tau_Y}
$$

$$
=\mathbb{E}\left(1+\frac{1}{2}L(X,X')+R(X,X')\right)\left(1+\frac{1}{2}L(Y,Y')+R(Y,Y')\right)
$$

$$
+\;\mathbb{E}\left(1+\frac{1}{2}L(X,X')+R(X,X')\right)\mathbb{E}\left(1+\frac{1}{2}L(Y,Y')+R(Y,Y')\right)
$$

$$
-\;2\,\mathbb{E}\left(1+\frac{1}{2}L(X,X')+R(X,X')\right)\left(1+\frac{1}{2}L(Y,Y'')+R(Y,Y'')\right)
$$

$$
=L\;+\;R,
$$

where

$$
L \;=\; \frac{1}{4}\Big[\,\mathbb{E}\,L(X,X')L(Y,Y') \;+\; \mathbb{E}\,L(X,X')\,\mathbb{E}\,L(Y,Y')
$$
$$
-\;2\,\mathbb{E}\,L(X,X')L(Y,Y'')\,\Big],
$$

and

$$
R \;=\; \mathbb{E}\left[\frac{1}{2}L(X,X')R(Y,Y')\;+\;\frac{1}{2}R(X,X')L(Y,Y')\;+\;R(X,X')R(Y,Y')\right]
$$
$$
-\;2\,\mathbb{E}\left[\frac{1}{2}L(X,X')R(Y,Y'')\;+\;\frac{1}{2}R(X,X')L(Y,Y'')\right.
$$
$$
\left.+\;R(X,X')R(Y,Y'')\right]
$$
$$
+\;\mathbb{E}\,R(X,X')\,\mathbb{E}\,R(Y,Y').
$$

Some simple calculations yield

$$
L \;=\; \frac{1}{4\tau_{XY}^2}\Big\{\mathbb{E}\left[K^2(X,X')K^2(Y,Y')\right]\;+\;\mathbb{E}\left[K^2(X,X')\right]\mathbb{E}\left[K^2(Y,Y')\right]
$$
$$
-\;2\,\mathbb{E}\left[K^2(X,X')K^2(Y,Y'')\right]\Big\}
$$
$$
=\;\frac{1}{4\tau_{XY}^2}\sum_{i=1}^{p}\sum_{j=1}^{q}\Big\{\mathbb{E}\left[\rho_i(X_{(i)},X'_{(i)})\,\rho_j(Y_{(j)},Y'_{(j)})\right]
$$
$$
+\;\mathbb{E}\left[\rho_i(X_{(i)},X'_{(i)})\right]\mathbb{E}\left[\rho_j(Y_{(j)},Y'_{(j)})\right]
$$
$$
-\;2\,\mathbb{E}\left[\rho_i(X_{(i)},X'_{(i)})\,\rho_j(Y_{(j)},Y''_{(j)})\right]\Big\}
$$
$$
=\;\frac{1}{4\tau_{XY}^2}\sum_{i=1}^{p}\sum_{j=1}^{q}D^2_{\rho_i,\rho_j}(X_{(i)},Y_{(j)}).
$$

To observe that the remainder term is negligible, note that under Assumption 5.2,

$$\mathbb{E}\left[L(X,X')R(Y,Y')\right] \leq \left(\mathbb{E}\left[L(X,X')^2\right]\mathbb{E}\left[R(Y,Y')^2\right]\right)^{1/2} = O(a_p' b_q'^2),$$

$$\mathbb{E}\left[R(X,X')L(Y,Y')\right] \leq \left(\mathbb{E}\left[R(X,X')^2\right]\mathbb{E}\left[L(Y,Y')^2\right]\right)^{1/2} = O(a_p'^2 b_q'),$$

$$\mathbb{E}\left[R(X,X')R(Y,Y')\right] \leq \left(\mathbb{E}\left[R(X,X')^2\right]\mathbb{E}\left[R(Y,Y')^2\right]\right)^{1/2} = O(a_p'^2 b_q'^2),$$

Clearly, $\mathcal{R} = \tau_{XY}R = O(\tau_{XY}\, a_p'^2 b_q' + \tau_{XY}\, a_p' b_q'^2)$. $\qquad\diamond$

Proof of Theorem 5.2. The proof is essentially similar to the proof of Theorem 5.1. Note that using Proposition 4.1, we can write

$$\frac{1}{\tau_Y}\mathcal{D}^2(X,Y)$$

$$= \mathbb{E}\,K(X,X')\frac{K(Y,Y')}{\tau_Y} + \mathbb{E}\,K(X,X')\,\mathbb{E}\,\frac{K(Y,Y')}{\tau_Y} - 2\,\mathbb{E}\,K(X,X')\frac{K(Y,Y'')}{\tau_Y}$$

$$= \mathbb{E}\,K(X,X')\left(1 + \frac{1}{2}L(Y,Y') + R(Y,Y')\right)$$

$$\quad + \mathbb{E}\,K(X,X')\,\mathbb{E}\left(1 + \frac{1}{2}L(Y,Y') + R(Y,Y')\right)$$

$$\quad - 2\,\mathbb{E}\,K(X,X')\left(1 + \frac{1}{2}L(Y,Y'') + R(Y,Y'')\right)$$

$$= L + R,$$

where

$$L = \frac{1}{2\tau_Y^2}\sum_{j=1}^q \left\{ \mathbb{E}\left[K(X,X')\,\rho_j(Y_{(j)},Y'_{(j)})\right] + \mathbb{E}\left[K(X,X')\,\mathbb{E}\left[\rho_j(Y_{(j)},Y'_{(j)})\right]\right.$$

$$\left. - 2\,\mathbb{E}\left[K(X,X')\,\rho_j(Y_{(j)},Y''_{(j)})\right]\right\}$$

$$= \frac{1}{2\tau_Y^2}\sum_{j=1}^q D^2_{K,\,\rho_j}(X,Y_{(j)}),$$

and

$$R = \mathbb{E}\left[K(X,X')R(Y,Y')\right] + \mathbb{E}\left[K(X,X')\right]\mathbb{E}\left[R(Y,Y')\right]$$

$$\quad - 2\,\mathbb{E}\left[K(X,X')R(Y,Y'')\right].$$

Under the assumption that $\mathbb{E}\left[R^2(Y,Y')\right] = O(b_q'^4)$, using Hölder's inequality it is easy to see that $\tau_Y R = O(\tau_Y\, b_q'^2) = o(1)$. $\qquad\diamond$

Proof of Theorem 5.3. Following equation (D.28), we have for $1 \leq k \neq l \leq n$

$$\widetilde{D}_{kl}^X = \frac{\tau_X}{2}\widetilde{L}(X_k,X_l) + \tau_X\widetilde{R}(X_k,X_l)$$

$$= \frac{1}{2\tau_X} \sum_{i=1}^{p} \widetilde{\rho}_i(X_{k(i)}, X_{l(i)}) \; + \; \tau_X \widetilde{R}(X_k, X_l),$$

$$\widetilde{D}_{kl}^Y \;=\; \frac{\tau_Y}{2} \widetilde{L}(Y_k, Y_l) \;+\; \tau_Y \widetilde{R}(Y_k, Y_l) \;=\; \frac{1}{2\tau_Y} \sum_{j=1}^{q} \widetilde{\rho}_i(Y_{k(j)}, Y_{l(j)}) \;+\; \tau_Y \widetilde{R}(Y_k, Y_l).$$

From equation (2.14) in the main paper it is easy to check that

$$\widetilde{\mathcal{D}}_n^2(X, Y)$$

$$= \frac{1}{4\tau_{XY}} \sum_{i=1}^{p} \sum_{j=1}^{q} \widetilde{D}_{n\,;\,\rho_i,\rho_j}^2(X_{(i)}, Y_{(j)}) \;+\; \frac{\tau_{XY}}{2n(n-3)} \sum_{k \neq l} \widetilde{L}(X_k, X_l) \widetilde{R}(Y_k, Y_l)$$

$$+ \; \frac{\tau_{XY}}{2n(n-3)} \sum_{k \neq l} \widetilde{L}(Y_k, Y_l) \widetilde{R}(X_k, X_l) \;+\; \frac{\tau_{XY}}{n(n-3)} \sum_{k \neq l} \widetilde{R}(X_k, X_l) \widetilde{R}(Y_k, Y_l).$$

Under Assumption 5.3, using Hölder's inequality and power mean inequality, it can be verified that

$$\sum_{k \neq l} \widetilde{L}(X_k, X_l) \widetilde{R}(Y_k, Y_l) \;\le\; \left(\sum_{k \neq l} \widetilde{L}(X_k, X_l)^2 \sum_{k \neq l} \widetilde{R}(Y_k, Y_l)^2 \right)^{1/2} \;=\; O_p(a_p b_q^2),$$

$$\sum_{k \neq l} \widetilde{L}(Y_k, Y_l) \widetilde{R}(X_k, X_l) \;\le\; \left(\sum_{k \neq l} \widetilde{L}(Y_k, Y_l)^2 \sum_{k \neq l} \widetilde{R}(X_k, X_l)^2 \right)^{1/2} \;=\; O_p(a_p^2 b_q),$$

$$\sum_{k \neq l} \widetilde{R}(X_k, X_l) \widetilde{R}(Y_k, Y_l) \;\le\; \left(\sum_{k \neq l} \widetilde{R}(X_k, X_l)^2 \sum_{k \neq l} \widetilde{R}(Y_k, Y_l)^2 \right)^{1/2} \;=\; O_p(a_p^2 b_q^2).$$

This completes the proof of the theorem. $\diamond$

Proof of Theorem 5.4. Following equation (D.28), we have for $1 \le k \neq l \le n$

$$\widetilde{D}_{kl}^Y \;=\; \frac{1}{2\tau_Y} \sum_{j=1}^{q} \widetilde{\rho}_j(Y_{k(j)}, Y_{l(j)}) \;+\; \tau_Y \widetilde{R}(Y_k, Y_l),$$

and therefore

$$\widetilde{\mathcal{D}}_n^2(X, Y) \;=\; \frac{1}{2\tau_Y} \sum_{j=1}^{q} \widetilde{D}_{n\,;\,K,\rho_j}^2(X, Y_{(j)}) \;+\; \frac{\tau_Y}{n(n-3)} \sum_{k \neq l} \widetilde{K}(X_k, X_l) \widetilde{R}(Y_k, Y_l).$$

Using power mean inequality, it can be verified that $\sum_{k \neq l} \widetilde{K}(X_k, X_l) \widetilde{R}(Y_k, Y_l) = O_p(b_q^2)$. This completes the proof of the theorem. $\diamond$

Proof of Theorem 5.5. The proof follows similar lines of the proof Theorem 2.2.1 in [57], with the distance metric being the one from the class of metrics we proposed in equation (3.1). $\diamond$

Proof of Theorem 5.6. The proof of the theorem follows similar lines of the proof of Proposition 2.2.2 in [57]. ◇

Proof of Theorem B.2. The decomposition into the leading term follows the similar lines of the proof of Theorem 5.3. The negligibility of the remainder term can be shown by mimicking the proof of Theorem 3.1.1 in [57]. ◇

Proof of Theorem B.3. It essentially follows similar lines of Proposition 3.2.1 in [57]. ◇

Acknowledgments

References

[1] BARINGHAUS, L. and FRANZ, C. (2004). On a new multivariate two-sample test. *Journal of Multivariate Analysis*, *88*(1), 190-206. MR2021870

[2] BERGSMA, W. and DASSIOS, A. (2014). A consistent test of independence based on a sign covariance related to Kendall's tau. *Bernoulli*, *20*(2) 1006-1028. MR3178526

[3] BERRETT, T.B., and SAMWORTH, R.J. (2019). Nonparametric independence testing via mutual information. *Biometrika*, *106*(3) 547–566. MR3992389

[4] BICKEL, P. J. (1969). A Distribution Free Version of the Smirnov Two Sample Test in the p-Variate Case. *The Annals of Mathematical Statistics*, *40*(1) 1-23. MR0256519

[5] BÖTTCHER, B. (2017). Dependence structures - estimation and visualization using distance multivariance. arxiv:1712.06532.
Conditions. A Survey and Some Open Questions. *Probability Surveys*, *2* 107-144.

[6] CHAKRABORTY, S. and ZHANG, X. (2019). Distance Metrics for Measuring Joint Dependence with Application to Causal Inference. *Journal of the American Statistical Association*, *114*(528), 1638-1650. MR4047289

[7] CHATTERJEE, S.(2019). A new coefficient of correlation. *Journal of the American Statistical Association*, to appear. MR0772649

[8] CHEN, H. and FRIEDMAN, J. H. (2017). A New Graph-Based Two-Sample Test for Multivariate and Object Data. *Journal of the American Statistical Association*, *112*(517), 397-409. MR3646580

[9] CRESSIE, N., DAVIS, A.S., FOLKS, J.L. and POLICELLO II, G.E. (1981). The Moment-Generating Function and Negative Integer Moments. *The American Statistician*, *35*(3), 148-150. MR0632425

[10] DARLING, D. A. (1957). The Kolmogorov-Smirnov, Cramer-von Mises Tests. *The Annals of Mathematical Statistics*, *29*(3) 842-851. MR0093870

[11] Dau, H. A., Keogh, E., Kamgar, K., Yeh, C. C. M., Zhu, Y., Gharghabi, S., Ratanamahatana, C. A., Chen, Y., Hu, B., Begum, N., Bagnall, A., Mueen, A. and Batista, G. (2018). The UCR Time Series Classification Archive. URL https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

[12] David, H. T. (1958). A Three-Sample Kolmogorov-Smirnov Test. *The Annals of Mathematical Statistics*, *28*(4) 823-838. MR0093871

[13] Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, *84*(2), 313-342. MR1719345

[14] Doukhan, P. and Neumann, M.H. (2008). The notion of ψ-weak dependence and its applications to bootstrapping time series. *Stochastic Processes and their Applications*, *5*, 146-168. MR2426177

[15] Edelmann, D., Fokianos, K. and Pitsillou, M. (2018). An Updated Literature Review of Distance Correlation and its Applications to Time Series. arxiv:1710.01146. MR3994758

[16] Friedman, J. H. and Rafsky, L. C. (1979). Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-Sample Tests. *Annals of Statistics*, *7*(4) 697-717. MR0532236

[17] Gretton, A., Bousquet, O., Smola, A. and Schölkopf, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. *Algorithmic Learning Theory, Springer-Verlag*, 63-77. MR2255909

[18] Gretton, A., Fukumizu, C. H. Teo., Song, L., Schölkopf, B. and Smola, A. (2007). A kernel statistical test of independence. *Advances in Neural Information Processing Systems*, *20* 585-592.

[19] Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B. and Smola, A. (2012). A Kernel Two-Sample Test. *Journal of Machine Learning Research*, *13* 723-773. MR2913716

[20] Hall, P. and Heyde, C. C. (1980). Martingale Limit Theory and Its Applications. *Academic press*. MR0624435

[21] Huo, X. and Székely, G. J. (2016). Fast computing for distance covariance. *Technometrics*, *58*(4) 435-446. MR3556612

[22] Jin, Z. and Matteson, D. S. (2017). Generalizing Distance Covariance to Measure and Test Multivariate Mutual Dependence. https://arxiv.org/abs/1709.02532. MR3858367

[23] Josse, J. and Holmes, S. (2014). Tests of independence and Beyond. arxiv:1307.7383.

[24] Kim, I., Balakrishnan, S. and Wasserman, L. (2018). Robust multivariate nonparametric tests via projection-pursuit. arXiv:1803.00715. MR4185814

[25] Li, J. (2018). Asymptotic normality of interpoint distances for high-dimensional data with applications to the two-sample problem. *Biometrika*, 105(3), 529-546. MR3842883

[26] Lyons, R. (2013). Distance covariance in metric spaces. *Annals of Probability*, *41*(5) 3284-3305. MR3127883

[27] Ma, L. and Mao, J. (2019). Fisher Exact Scanning for Dependency. *Jour-*

nal of the American Statistical Association, *114*(525), 245-258. MR3941252

[28] MAA, J.-F., PEARL, D. K. and BARTOSZYŃSKI, R. (1996). Reducing multidimensional two-sample data to one-dimensional interpoint comparisons. *Annals of Statistics*, *24*(3) 1069-1074. MR1401837

[29] MATTESON, D. S. and TSAY, R. S. (2017). Independent component analysis via distance covariance. *Journal of the American Statistical Association*, *112*(518), 623-637. MR3671757

[30] NEUHAUS, G.(1977). Functional Limit Theorems for U-Statistics in the Degenerate Case. *Journal of Multivariate Analysis*, *7*, 424-439. MR0455084

[31] PHILLIPS, P.C.B. and MOON, H.R. (2000). Nonstationary panel data analysis: an overview of some recent developments. *Econometric Reviews*, *19*(3), 263-286. MR1791963

[32] RAMDAS, A., REDDI, S. J., POCZOS, B., SINGH, A. and WASSERMAN, L.. Adaptivity and Computation-Statistics Tradeoffs for Kernel and Distance based High Dimensional Two Sample Testing. arXiv:1508.00655.

[33] RAMDAS, A., REDDI, S. J., POCZOS, B., SINGH, A. and WASSERMAN, L.. On the Decreasing Power of Kernel and Distance Based Nonparametric Hypothesis Tests in High Dimensions. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

[34] RESNICK, S. I. (1999). *A Probability Path*. Springer. MR1664717

[35] RIO, E. (1993). Covariance inequalities for strongly mixing processes. *Annales de l'I. H. P., Section B*, *29*(4) 587-597. MR1251142

[36] ROY, A., GOSWAMI, A. and MURTHY, C.A. (2018). Multivariate dependency measure based on copula and Gaussian kernel. arXiv:1708.07485v2.

[37] SARKAR, S. and GHOSH, A.K. (2018). Some multivariate tests of independence based on ranks of nearest neighbors. *Technometrics*, *60*(1) 101-111. MR3768054

[38] SCHILLING, M. F. (1986). Multivariate Two-Sample Tests Based on Nearest Neighbors. *Journal of the American Statistical Association*, *81*(395) 799-806. MR0860514

[39] SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Annals of Statistics*, *41*(5) 2263-2291. MR3127866

[40] SEN, P. K. (1977). Almost Sure Convergence of Generalized U-Statistics. *Annals of Probability*, *5*(2) 287-290. MR0436444

[41] SERFLING, R. J. (1980). Approximation Theorems of Mathematical Statistics. *Wiley*, New York. MR0595165

[42] SRIPERUMBUDUR, B., GRETTON, A., FUKUMIZU, K., SCHÖLKOPF, B. and LANCKRIET, G.R.G (2010). Hilbert Space Embeddings and Metrics on Probability Measures. *Journal of Machine Learning Research*, *11* 1517-1561. MR2645460

[43] SHAO, X. and ZHANG, J. (2014). Martingale Difference Correlation and Its Use in High-Dimensional Variable Screening. *Journal of the American Statistical Association*, *109*(507) 1302-1318. MR3265698

[44] SZÉKELY, G. J. (2002). E-Statistics: the Energy of Statistical Samples. Technical report.

[45] Székely, G. J. and Rizzo, M. L. (2004). Testing for equal distributions in high dimension. *InterStat*, *5*.

[46] Székely, G. J. and Rizzo, M. L. (2005). Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. *Journal of Classification*, *22*(2) 151-183. MR2231170

[47] Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing independence by correlation of distances. *Annals of Statistics*, *35*(6) 2769-2794. MR2382665

[48] Székely, G. J. and Rizzo, M. L. (2013). The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis*, *117* 193-213. MR3053543

[49] Székely, G. J. and Rizzo, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Annals of Statistics*, *42*(6) 2382-2412. MR3269983

[50] Wang, X., Wenliang, P., Hu, W., Tian, Y. and Zhang, H. (2015). Conditional distance correlation. *Journal of the American Statistical Association*, *110*(512) 1726-1734. MR3449068

[51] Weihs, L., Drton, M., and Meinshausen, N. (2018). Symmetric rank covariances: a generalized framework for nonparametric measures of dependence. *Biometrika*, *105*(3) 547–562. MR3842884

[52] Yao, S., Zhang, X. and Shao, X. (2018). Testing Mutual Independence in High Dimension via Distance Covariance. *Journal of the Royal Statistical Society, Series B*, *80*(3) 455-480. MR3798874

[53] Zhang, K. (2019). BET on Independence. *Journal of the American Statistical Association*, *114*(528), 1620–1637. MR4047288

[54] Zhang, X., Yao, S. and Shao, X. (2018). Conditional Mean and Quantile Dependence Testing in High Dimension. *Annals of Statistics*, *46*(1) 219-246. MR3766951

[55] Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis*, *33*(3), 438-457. MR2915095

[56] Zhu, L., Xu, K., Li, R., and Zhong, W. (2017). Projection correlation between two random vectors. *Biometrika*, *104*(4) 829-843. MR3737307

[57] Zhu, C., Yao, S., Zhang, X. and Shao, X. (2020). Distance-based and RKHS-based Dependence Metrics in High-dimension. *Annals of Statistics*, *48*(6) 3366-3394. MR4185812