

PARTIAL DISTANCE CORRELATION WITH METHODS FOR DISSIMILARITIES

GÁBOR J. SZÉKELY AND MARIA L. RIZZO

ABSTRACT. Distance covariance and distance correlation are scalar coefficients that characterize independence of random vectors in arbitrary dimension. Properties, extensions, and applications of distance correlation have been discussed in the recent literature, but the problem of defining the partial distance correlation has remained an open question of considerable interest. The problem of partial distance correlation is more complex than partial correlation partly because the squared distance covariance is not an inner product in the usual linear space. For the definition of partial distance correlation we introduce a new Hilbert space where the squared distance covariance is the inner product. We define the partial distance correlation statistics with the help of this Hilbert space, and develop and implement a test for zero partial distance correlation. Our intermediate results provide an unbiased estimator of squared distance covariance, and a neat solution to the problem of distance correlation for dissimilarities rather than distances.

[2000] Primary: 62H20, 62H15; Secondary: 62Gxx

independence, multivariate, partial distance correlation, dissimilarity, energy statistics

1. INTRODUCTION

Distance covariance (dCov) and distance correlation characterize multivariate independence for random vectors in arbitrary, not necessarily equal dimension. In this work we focus on the open problem of *partial distance correlation*. Our intermediate results include methods for applying distance correlation to dissimilarity matrices.

The distance covariance, denoted $\mathcal{V}(X, Y)$, of two random vectors X and Y characterizes independence; that is,

$$\mathcal{V}(X, Y) \geq 0$$

with equality to zero if and only if X and Y are independent. This coefficient is defined by a weighted L_2 norm measuring the distance between the joint characteristic function (c.f.) $\phi_{X, Y}$ of X and Y , and the product $\phi_X \phi_Y$ of the marginal c.f.'s of X and Y . If X and Y take values in \mathbb{R}^p and \mathbb{R}^q , respectively, $\mathcal{V}(X, Y)$ is

Date: March 22, 2014.

Gábor J. Székely is Program Director in Statistics at the National Science Foundation, 4201 Wilson Blvd. #1025, Arlington, VA 22230, email: gszekely@nsf.gov.

Maria L. Rizzo is Associate Professor in the Dept. of Mathematics and Statistics at Bowling Green State University, Bowling Green, OH 43403, email: mrizzo@bgsu.edu.

Research of the first author was supported by the National Science Foundation, while working at the Foundation.

the non-negative square root of

$$\begin{aligned} \mathcal{V}^2(X, Y) &= \|\phi_{X,Y}(t, s) - \phi_X(t)\phi_Y(s)\|_w^2 \\ &:= \int_{R^{p+q}} |\phi_{X,Y}(t, s) - \phi_X(t)\phi_Y(s)|^2 w(t, s) dt ds, \end{aligned} \quad (1.1)$$

where $w(t, s) := (|t|_p^{1+p}|s|_q^{1+q})^{-1}$. The integral exists provided that X and Y have finite first moments. Note that Feuerverger (1993) [5] proposed a bivariate test based on this idea and applied the same weight function w , where it may have first appeared.

The following identity is established in Székely and Rizzo [32, Theorem 8, p. 1250]. Let (X, Y) , (X', Y') , and (X'', Y'') be independent and identically distributed (iid), each with joint distribution (X, Y) . Then

$$\begin{aligned} \mathcal{V}^2(X, Y) &= E|X - X'| |Y - Y'| + E|X - X'| E|Y - Y'| \\ &\quad - E|X - X'| |Y - Y''| - E|X - X''| |Y - Y'|, \end{aligned} \quad (1.2)$$

provided that X and Y have finite first moments. In Section 4 an alternate version of (1.2) is defined for X and Y taking values in a separable Hilbert space. That definition and intermediate results lead to the definition of partial distance covariance.

We summarize a few key properties and computing formulas below for easy reference. The distance correlation (dCor) $\mathcal{R}(X, Y)$ is a standardized coefficient, $0 \leq \mathcal{R}(X, Y) \leq 1$, that also characterizes independence:

$$\mathcal{R}^2(X, Y) = \begin{cases} \frac{\mathcal{V}(X, Y)}{\sqrt{\mathcal{V}(X, X)\mathcal{V}(Y, Y)}}, & \mathcal{V}(X, X)\mathcal{V}(Y, Y) > 0; \\ 0, & \mathcal{V}(X, X)\mathcal{V}(Y, Y) = 0. \end{cases}$$

For more details see [31] and [32]. Properties, extensions, and applications of distance correlation have been discussed in the recent literature; see e.g. Dueck et al. [4], Lyons [16], Kong et al. [12], and Li, Zhong, and Zhu [15]. However, there is considerable interest among researchers and statisticians on the open problem of a suitable definition and supporting theory of partial distance correlation. Among the many potential application areas of partial distance correlation are variable selection (see Example 7) and graphical models; see e.g. Wermuth and Cox [37] for an example of work that motivated the question in that context.

In this work we introduce the definition of partial distance covariance (pdCov) and partial distance correlation (pdCor) statistics and population coefficients. First, let us see why it is not straightforward to generalize distance correlation to partial distance correlation in a meaningful way that preserves the essential properties one would require, and allows for interpretation and inference.

One could try to follow the definitions of the classical partial covariance and partial correlation that are based on orthogonal projections in a Euclidean space, but there is a serious difficulty. Orthogonality in case of partial distance covariance and partial distance correlation means independence, but when we compute the orthogonal projection of a random variable onto the condition variable, the “remainder” in the difference is typically not independent of the condition.

Alternately, the form of sample distance covariance (Definition 1) may suggest an inner product, so one might think of working in the Hilbert space of double centered distance matrices (2.1), where the inner product is the squared distance covariance statistic (2.2). Here we are facing another problem: what would the

projections represent? The difference D of double centered distance matrices is typically not a double centered distance matrix of any sample. This does not affect formal computations, but if we cannot interpret our formulas in terms of samples then inference becomes impossible.

To overcome these difficulties while preserving the essential properties of distance covariance, we finally arrived at an elegant solution which starts with defining an alternate type of double centering called “ \mathcal{U} -centering” (see Definition 2 and Proposition 1 below). The corresponding inner product is an unbiased estimator of squared population distance covariance. In the Hilbert space of “ \mathcal{U} -centered” matrices, all linear combinations, and in particular projections, are zero diagonal \mathcal{U} -centered matrices. We prove a representation theorem that connects the orthogonal projections to random samples in Euclidean space. The corresponding population coefficients are defined by extension to certain separable Hilbert spaces. Methods for inference are outlined and implemented, including methods for non-Euclidean dissimilarities.

Definitions and background for dCov and dCor are summarized in Section 2. The partial distance correlation statistic is introduced in Section 3, and Section 4 covers the population partial distance covariance and inference for a test of the hypothesis of zero partial distance correlation. Examples and Applications are given in Section 5, followed by a Summary in Section 6. Appendix A contains proofs of statements.

2. BACKGROUND

In this section we summarize the definitions for distance covariance and distance correlation statistics.

2.1. Notation. Random vectors are denoted with upper case letters and their sample realizations with lowercase letters. In certain contexts we need to work with a data matrix, which is denoted in boldface, such as \mathbf{U} or \mathbf{V} . For example, the sample distance correlation of a realization from the joint distribution of (X, Y) is denoted $\mathcal{R}_n(\mathbf{X}, \mathbf{Y})$. The norm $|\cdot|_d$ denotes the Euclidean norm when its argument is in \mathbb{R}^d , and we omit the subscript if the meaning is clear in context. A primed random variable denotes an independent and identically distributed (iid) copy of the unprimed symbol; so X, X' are iid, etc. The pair x, x' denotes two realizations of the random variable X . The transpose of x is denoted by x^T . Other notation will be introduced within the section where it is needed.

In this paper a *random sample* refers to independent and identically distributed (iid) realizations from the underlying joint or marginal distribution.

2.2. Sample dCov and dCor. The distance covariance and distance correlation statistics are functions of the double centered distance matrices of the samples. For an observed random sample $\{(x_i, y_i) : i = 1, \dots, n\}$ from the joint distribution of random vectors X and Y , compute the Euclidean distance matrices $(a_{ij}) = (|x_i - x_j|_p)$ and $(b_{ij}) = (|y_i - y_j|_q)$. Define

$$\hat{A}_{ij} = a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}, \quad i, j = 1, \dots, n, \quad (2.1)$$

where

$$\bar{a}_{i.} = \frac{1}{n} \sum_{j=1}^n a_{ij}, \quad \bar{a}_{.j} = \frac{1}{n} \sum_{i=1}^n a_{ij}, \quad \bar{a}_{..} = \frac{1}{n^2} \sum_{i,j=1}^n a_{ij}.$$

Similarly, define $\widehat{B}_{ij} = b_{ij} - \bar{b}_{i.} - \bar{b}_{.j} + \bar{b}_{..}$, for $i, j = 1, \dots, n$.

remark 1. In previous work, we denoted the double centered distance matrix of a sample by a plain capital letter, so that \widehat{A} and \widehat{B} above were denoted by A and B . In this paper we work with two different methods of centering, so we have introduced notation to clearly distinguish the two methods. In the first centering above the centered versions \widehat{A}_{ij} have the property that all rows and columns sum to zero. Another type of centering, which we will call unbiased or \mathcal{U} -centering and denote by \widetilde{A}_{ij} in Definition 2 below, has the additional property that all expectations are zero: that is, $E[\widetilde{A}_{ij}] = 0$ for all i, j . Although this centering has a somewhat more complicated formula (3.1), we will see in this paper the advantages of the new centering.

Definition 1. *The sample distance covariance $\mathcal{V}_n(\mathbf{X}, \mathbf{Y})$ and sample distance correlation $\mathcal{R}_n(\mathbf{X}, \mathbf{Y})$ are defined by*

$$\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{i,j=1}^n \widehat{A}_{ij} \widehat{B}_{ij}. \quad (2.2)$$

and

$$\mathcal{R}_n^2(\mathbf{X}, \mathbf{Y}) = \begin{cases} \frac{\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y})}{\sqrt{\mathcal{V}_n^2(\mathbf{X})\mathcal{V}_n^2(\mathbf{Y})}}, & \mathcal{V}_n^2(\mathbf{X})\mathcal{V}_n^2(\mathbf{Y}) > 0; \\ 0, & \mathcal{V}_n^2(\mathbf{X})\mathcal{V}_n^2(\mathbf{Y}) = 0. \end{cases} \quad (2.3)$$

respectively, where the squared sample distance variance is defined by

$$\mathcal{V}_n^2(\mathbf{X}) = \mathcal{V}_n^2(\mathbf{X}, \mathbf{X}) = \frac{1}{n^2} \sum_{i,j=1}^n \widehat{A}_{ij}^2. \quad (2.4)$$

remark 2. Although it is obvious, it is perhaps worth noting that $\mathcal{V}_n(\mathbf{X}, \mathbf{Y})$ and $\mathcal{R}_n(\mathbf{X}, \mathbf{Y})$ are rigid motion invariant, and they can be defined entirely in terms of the distance matrices, or equivalently in terms of the double centered distance matrices. This fact is important for what follows, because as we will see, one can compute partial distance covariance by operating on certain transformations of the distance matrices, without reference to the original data that generated the matrices. (These definitions are not invariant to monotone transformation of coordinates, such as ranks.)

If X and Y have finite first moments, the population distance covariance coefficient $\mathcal{V}^2(X, Y)$ exists and equals zero if and only if the random vectors X and Y are independent. Some of the properties of distance covariance and distance correlation include:

- (i) $\mathcal{V}_n(\mathbf{X}, \mathbf{Y})$ and $\mathcal{R}_n(\mathbf{X}, \mathbf{Y})$ converge almost surely to $\mathcal{V}(X, Y)$ and $\mathcal{R}(X, Y)$, as $n \rightarrow \infty$.
- (ii) $\mathcal{V}_n(\mathbf{X}, \mathbf{Y}) \geq 0$ and $\mathcal{V}_n(\mathbf{X}) = 0$ if and only if every sample observation is identical.
- (iii) $0 \leq \mathcal{R}_n(\mathbf{X}, \mathbf{Y}) \leq 1$.
- (iv) If $\mathcal{R}_n(\mathbf{X}, \mathbf{Y}) = 1$ then there exists a vector a , a non-zero real number b and an orthogonal matrix R such that $\mathbf{Y} = a + b\mathbf{X}R$, for the data matrices \mathbf{X} and \mathbf{Y} .

A consistent test of multivariate independence is based on the sample distance covariance: large values of $n\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y})$ support the alternative hypothesis that X

and Y are dependent (see [31, 32]). A high-dimensional distance correlation t -test of independence was introduced by Székely and Rizzo [34]; the tests are implemented as `dcov.test` and `dcor.ttest`, respectively, in the *energy* package (Rizzo and Székely, [25]) for R (R Core Team [22]).

The definitions of sample distance covariance and sample distance correlation can be extended to random samples taking values in any, possibly different, metric spaces. For defining the population coefficient we need to suppose more (see Lyons [16]). If X and Y take values in possibly different separable Hilbert spaces and both X and Y have finite expectations then it remains true that $\mathcal{V}^2(X, Y) \geq 0$, and equals zero if and only if X and Y are independent. This implies that the results of this paper can be extended to separable Hilbert space valued variables. Extensions and population coefficients `pdCov` and `pdCor` are discussed in Section 4. Extending Lyons [16], Sejdinovic et al. [28] discuss the equivalence of distance based and RKHS-based statistics for testing dependence.

For theory, background, and further properties of the population and sample coefficients, see Székely, et al. [31], Székely and Rizzo [32], and Lyons [16], and the references therein. On the weight function applied see also Feuerverger [5] and Székely and Rizzo [33]. For an overview of recent methods for measuring dependence and testing independence of random vectors, including distance covariance, readers are referred to Josse and Holmes [10].

3. PARTIAL DISTANCE CORRELATION

We introduce partial distance correlation by first considering the sample coefficient.

3.1. The Hilbert space of centered distance matrices.

Definition 2. Let $A = (a_{ij})$ be a symmetric, real valued $n \times n$ matrix with zero diagonal, $n > 2$. Define the \mathcal{U} -centered matrix \tilde{A} as follows. Let the (i, j) -th entry of \tilde{A} be defined by

$$\tilde{A}_{i,j} = \begin{cases} a_{i,j} - \frac{1}{n-2} \sum_{\ell=1}^n a_{i,\ell} - \frac{1}{n-2} \sum_{k=1}^n a_{k,j} + \frac{1}{(n-1)(n-2)} \sum_{k,\ell=1}^n a_{k,\ell}, & i \neq j; \\ 0, & i = j. \end{cases} \quad (3.1)$$

Here “ \mathcal{U} -centered” is so named because as shown below, the corresponding inner product (3.3) defines an unbiased estimator of squared distance covariance.

Proposition 1. Let (x_i, y_i) , $i = 1, \dots, n$ denote a sample of observations from the joint distribution (X, Y) of random vectors X and Y . Let $A = (a_{ij})$ be the Euclidean distance matrix of the sample x_1, \dots, x_n from the distribution of X , and $B = (b_{ij})$ be the Euclidean distance matrix of the sample y_1, \dots, y_n from the distribution of Y . Then if $E(|X| + |Y|) < \infty$, for $n > 3$,

$$(\tilde{A} \cdot \tilde{B}) := \frac{1}{n(n-3)} \sum_{i \neq j} \tilde{A}_{i,j} \tilde{B}_{i,j} \quad (3.2)$$

is an unbiased estimator of squared population distance covariance $\mathcal{V}^2(X, Y)$.

Proposition 1 is proved in Appendix (A.1). It is obvious that $\tilde{A} = 0$ if all of the sample observations are identical. More generally, $\tilde{A} = 0$ if and only if the n sample

observations are equally distant or at least $n - 1$ of the n sample observations are identical.

For a fixed $n \geq 4$, we define a Hilbert space \mathcal{H}_n generated by Euclidean distance matrices of arbitrary sets (samples) of n points in a Euclidean space \mathbb{R}^p , $p \geq 1$. Consider the linear span \mathcal{S}_n of all $n \times n$ distance matrices of samples $\{x_1, \dots, x_n\}$. Let $A = (a_{ij})$ be an arbitrary element in \mathcal{S}_n . Then A is a real valued, symmetric matrix with zero diagonal.

Let $\mathcal{H}_n = \{\tilde{A} : A \in \mathcal{S}_n\}$ and for each pair of elements $C = (C_{i,j})$, $D = (D_{i,j})$ in the linear span of \mathcal{H}_n define their inner product

$$(C \cdot D) = \frac{1}{n(n-3)} \sum_{i \neq j} C_{ij} D_{ij}. \quad (3.3)$$

If $(C \cdot C) = 0$ then $(C \cdot D) = 0$ for any $D \in \mathcal{H}_n$.

Below in Theorem 2 it is shown that every matrix $C \in \mathcal{H}_n$ is the \mathcal{U} -centered distance matrix of a configuration of n points in a Euclidean space \mathbb{R}^p , where $p \leq n - 2$.

Theorem 1. *The linear span of all $n \times n$ matrices $\mathcal{H}_n = \{\tilde{A} : A \in \mathcal{S}_n\}$ is a Hilbert space with inner product defined by (3.3).*

Proof. Let \mathcal{H}_n denote the linear span of \mathcal{H}_n . If $\tilde{A}, \tilde{B} \in \mathcal{H}_n$ and $c, d \in \mathbb{R}$, then $(c\tilde{A} + d\tilde{B})_{i,j} = (c\tilde{A} + d\tilde{B})_{i,j}$, so $(c\tilde{A} + d\tilde{B}) \in \mathcal{H}_n$. It is also true that $C \in \mathcal{H}_n$ implies that $-C \in \mathcal{H}_n$, and the zero element is the $n \times n$ zero matrix.

Then since $(c\tilde{A} + d\tilde{B}) = (c\tilde{A} + d\tilde{B})$, for the inner product, we only need to prove that for $\tilde{A}, \tilde{B}, \tilde{C} \in \mathcal{H}_n$ and real constants c , the following statements hold:

- (i) $(\tilde{A} \cdot \tilde{A}) \geq 0$.
- (ii) $(\tilde{A} \cdot \tilde{A}) = 0$ only if $\tilde{A} = 0$.
- (iii) $((c\tilde{A}) \cdot \tilde{B}) = c(\tilde{A} \cdot \tilde{B})$
- (iv) $((\tilde{A} + \tilde{B}) \cdot \tilde{C}) = (\tilde{A} \cdot \tilde{C}) + (\tilde{B} \cdot \tilde{C})$.

Statements (i) and (ii) hold because $(\tilde{A} \cdot \tilde{A})$ is proportional to a sum of squares. Statements (iii) and (iv) follow easily from the definition of \mathcal{H}_n , \tilde{A} , and (3.3). \square

The space \mathcal{H}_n is finite dimensional because it is a subspace of the space of all symmetric, zero-diagonal $n \times n$ matrices.

In what follows, \mathcal{H}_n denotes the Hilbert space of Theorem 1 with inner product (3.3), and $|\tilde{A}| = (\tilde{A}, \tilde{A})^{1/2}$ is the norm of \tilde{A} .

3.2. Sample pdCov and pdCor. From Theorem 1 a projection operator (3.4) can be defined in the Hilbert space \mathcal{H}_n , $n \geq 4$, and applied to define partial distance covariance and partial distance correlation for random vectors in Euclidean spaces. Let \tilde{A}, \tilde{B} , and \tilde{C} be elements of \mathcal{H}_n corresponding to samples x, y , and z , respectively, and let

$$P_{z^\perp}(x) = \tilde{A} - \frac{(\tilde{A} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})} \tilde{C}, \quad P_{z^\perp}(y) = \tilde{B} - \frac{(\tilde{B} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})} \tilde{C}, \quad (3.4)$$

denote the orthogonal projection of $\tilde{A}(x)$ onto $(\tilde{C}(z))^\perp$ and the orthogonal projection of $\tilde{B}(y)$ onto $(\tilde{C}(z))^\perp$, respectively. In case $(\tilde{C} \cdot \tilde{C}) = 0$ the projections are defined $P_{z^\perp}(x) = \tilde{A}$ and $P_{z^\perp}(y) = \tilde{B}$. Clearly $P_{z^\perp}(x)$ and $P_{z^\perp}(y)$ are elements

of \mathcal{H}_n , their dot product is defined by (3.3), and we can define an estimator of $\text{pdCov}(X, Y; Z)$ via projections.

Definition 3 (Partial distance covariance). *Let (x, y, z) be a random sample observed from the joint distribution of (X, Y, Z) . The sample partial distance covariance (pdCov) is defined by*

$$\text{pdCov}(x, y; z) = (P_{z^\perp}(x) \cdot P_{z^\perp}(y)), \quad (3.5)$$

where $P_{z^\perp}(x)$, and $P_{z^\perp}(y)$ are defined by (3.4), and

$$(P_{z^\perp}(x) \cdot P_{z^\perp}(y)) = \frac{1}{n(n-3)} \sum_{i \neq j} (P_{z^\perp}(x))_{i,j} (P_{z^\perp}(y))_{i,j}. \quad (3.6)$$

Definition 4 (Partial distance correlation). *Let (x, y, z) be a random sample observed from the joint distribution of (X, Y, Z) . Then sample partial distance correlation is defined as the cosine of the angle θ between the ‘vectors’ $P_{z^\perp}(x)$ and $P_{z^\perp}(y)$ in the Hilbert space \mathcal{H}_n :*

$$R^*(x, y; z) := \cos \theta = \frac{(P_{z^\perp}(x) \cdot P_{z^\perp}(y))}{|P_{z^\perp}(x)| |P_{z^\perp}(y)|}, \quad |P_{z^\perp}(x)| |P_{z^\perp}(y)| \neq 0, \quad (3.7)$$

and otherwise $R^*(x, y; z) := 0$.

3.3. Representation in Euclidean space. Sample pdCov and pdCor have been defined via projections in the Hilbert space generated by \mathcal{U} -centered distance matrices. In this section, we start to consider the interpretation of sample pdCov and sample pdCor.

Since pdCov is defined as the inner product (3.6) of two \mathcal{U} -centered matrices, and (unbiased squared) distance covariance (3.2) is computed as inner product, a natural question is the following. Are matrices $P_{z^\perp}(x)$ and $P_{z^\perp}(y)$ the \mathcal{U} -centered *Euclidean* distance matrices of samples of points \mathbf{U} and \mathbf{V} , respectively? If so, then the sample partial distance covariance (3.6) is distance covariance of \mathbf{U} and \mathbf{V} , as defined by (3.2).

For every sample of points $\mathbf{X} = [x_1, \dots, x_n]$, $x_i \in \mathbb{R}^p$, there is a \mathcal{U} -centered matrix $\tilde{A} = \tilde{A}(x)$ in \mathcal{H}_n . Conversely, given an arbitrary element H of \mathcal{H}_n , does there exist a configuration of points $\mathbf{U} = [u_1, \dots, u_n]$ in some Euclidean space \mathbb{R}^q , for some $q \geq 1$, such that the \mathcal{U} -centered Euclidean distance matrix of sample \mathbf{U} is exactly equal to the matrix H ? In this section we prove that the answer is yes: $P_{z^\perp}(x)$, $P_{z^\perp}(y)$ of (3.6), and in general every element in \mathcal{H}_n , is the \mathcal{U} -centered distance matrix of some sample of n points in a Euclidean space.

First a few properties of centered distance matrices are established in the following lemma.

Lemma 1. *Let \tilde{A} be a \mathcal{U} -centered distance matrix. Then*

- (i) *Rows and columns of \tilde{A} sum to zero.*
- (ii) *$\widetilde{(\tilde{A})} = \tilde{A}$. That is, if B is the matrix obtained by \mathcal{U} -centering an element $\tilde{A} \in \mathcal{H}_n$, $B = \tilde{A}$.*
- (iii) *\tilde{A} is invariant to double centering. That is, if B is the matrix obtained by double centering the matrix \tilde{A} , then $B = \tilde{A}$.*
- (iv) *If c is a constant and B denotes the matrix obtained by adding c to the off-diagonal elements of \tilde{A} , then $\tilde{B} = \tilde{A}$.*

See Appendix A.3 for proof of Lemma 1.

As Lemma 1(iv) is essential for our results, it becomes clear that we cannot apply double centering as in the original (biased) definition of distance covariance here. The invariance with respect to the constant c in (iv) holds for \mathcal{U} -centered matrices but it does not hold for double centered matrices.

An $n \times n$ matrix D is called *Euclidean* if there exist points v_1, \dots, v_n in a Euclidean space such that their Euclidean distance matrix is exactly D ; that is, $d_{ij}^2 = |v_i - v_j|^2 = (v_i - v_j)^T(v_i - v_j)$, $i, j = 1, \dots, n$. Necessary and sufficient conditions that D is Euclidean are well known results of classical (metric) MDS. With the help of Lemma 1, and certain results from the theory of MDS, we can find for each element $H \in \mathcal{H}_n$ a configuration of points v_1, \dots, v_n in Euclidean space such that their \mathcal{U} -centered distance matrix is exactly equal to H .

A solution to the underlying system of equations to solve for the points v is found in Schoenberg [27] and Young and Householder [38]. It is a classical result that is well known in (metric) multidimensional scaling. Mardia, Kent, and Bibby [19] summarize the result in Theorem 14.2.1, and provide a proof. For an overview of the methodology see also Cox and Cox [3], Gower [7], Mardia [18], and Torgerson [36]. We apply the converse (b) of Theorem 14.2.1 as stated in Mardia, Kent, and Bibby [19], summarized below.

Let (d_{ij}) be a dissimilarity matrix and define $a_{ij} = -\frac{1}{2}d_{ij}^2$. Form the double centered matrix $\widehat{A} = (a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..})$. If \widehat{A} is positive semi-definite (p.s.d.) of rank p , then a configuration of points corresponding to \widehat{D} can be constructed as follows. Let $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ be the positive eigenvalues of \widehat{A} , with corresponding normalized eigenvectors v_1, \dots, v_p , such that $v_k^T v_k = \lambda_k$, $k = 1, \dots, p$. Then if V is the $n \times p$ matrix of eigenvectors, the rows of V are p -dimensional vectors that have interpoint distances equal to (d_{ij}) , and $\widehat{A} = VV^T$ is the inner product matrix of this set of points. The solution is constrained such that the centroid of the points is the origin. There is at least one zero eigenvalue so $p \leq n - 1$.

When the matrix \widehat{A} is not positive semi-definite, this leads us to the *additive constant problem*, which refers to the problem of finding a constant c such that by adding the constant to all off-diagonal entries of (d_{ij}) to obtain a dissimilarity matrix D_c , the resulting double centered matrix is p.s.d. Let $\widehat{A}_c(d_{ij}^2)$ denote the double centered matrix obtained by double centering $-\frac{1}{2}(d_{ij}^2 + c(1 - \delta^{ij}))$, where δ^{ij} is the Kronecker delta. Let $\widehat{A}_c(d_{ij})$ denote the matrix obtained by double centering $-\frac{1}{2}(d_{ij} + c(1 - \delta^{ij}))$. The smallest value of c that makes $\widehat{A}_c(d_{ij}^2)$ p.s.d. is $c^* = -2\lambda_n$, where λ_n is the smallest eigenvalue of $\widehat{A}_0(d_{ij}^2)$. Then $\widehat{A}_c(d_{ij}^2)$ is p.s.d. for every $c \geq c^*$. The number of positive eigenvalues of the p.s.d. double centered matrix $\widehat{A}_c(d_{ij}^2)$ determines the dimension required for the representation in Euclidean space.

However, we require a constant to be added to the elements d_{ij} rather than d_{ij}^2 . That is, we require a constant c , such that the dissimilarities $d_{ij}^{(c)} = d_{ij} + c(1 - \delta^{ij})$ are Euclidean. The solution by Cailliez [2] is c^* , where c^* is the largest eigenvalue of a $2n \times 2n$ block matrix

$$\begin{bmatrix} 0 & \widehat{A}_0(d_{ij}^2) \\ I & \widehat{A}_0(d_{ij}) \end{bmatrix},$$

where 0 is the zero matrix and I is the identity matrix of size n (see Cailliez [2] or Cox and Cox [3, Sec. 2.2.8] for details). This result guarantees that there exists a constant c^* such that the adjusted dissimilarities $d_{ij}^{(c)}$ are Euclidean. In this case at most $n - 2$ dimensions are required (Cailliez [2, Theorem 1]).

Finally, given an arbitrary element H of \mathcal{H}_n , the problem is to find a configuration of points $\mathbf{V} = [v_1, \dots, v_n]$ such that the \mathcal{U} -centered distance matrix of \mathbf{V} is exactly equal to the element H . Thus, if $H = (h_{ij})$ we are able to find points \mathbf{V} such that the Euclidean distance matrix of \mathbf{V} equals $H_c = (h_{ij} + c(1 - \delta^{ij}))$, and we need $\tilde{H}_c = H$. Now since $(c\tilde{A} + d\tilde{B}) = (c\tilde{A} + d\tilde{B})$, we can apply Lemma 1 to H , and $\tilde{H}_c = H$ follows from Lemma 1(ii) and Lemma 1(iv). Hence, by applying classical MDS with the additive constant theorem, and Lemma 1 (ii) and (iv), we obtain the configuration of points \mathbf{V} such that their \mathcal{U} -centered distances are exactly equal to the element $H \in \mathcal{H}_n$. Lemma 1(iv) also shows that the inner product is invariant to the constant c .

This establishes our theorem on representation in Euclidean space.

Theorem 2. *Let H be an arbitrary element of the Hilbert space \mathcal{H}_n of \mathcal{U} -centered distance matrices. Then there exists a sample v_1, \dots, v_n in a Euclidean space of dimension at most $n - 2$, such that the \mathcal{U} -centered distance matrix of v_1, \dots, v_n is exactly equal to H .*

remark 3. The above details also serve to illustrate why a Hilbert space of double centered matrices (as applied in the original, biased statistic \mathcal{V}_n^2) is not applicable for a meaningful definition of partial distance covariance. The diagonals of double centered distance matrices are not zero, so we cannot get an exact solution via MDS, and the inner product would depend on c . Another problem is that while \mathcal{V}_n^2 is always non-negative, the inner product of projections could easily be negative.

Methods for Dissimilarities. In community ecology and other fields of application, it is often the case that only the (non-Euclidean) dissimilarity matrices are available (see e.g. the genetic distances of Table 3). Suppose that the dissimilarity matrices are symmetric with zero diagonal. An application of Theorem 2 provides methods for this class of non-Euclidean dissimilarities. In this case, Theorem 2 provides that there exists samples in Euclidean space such that their \mathcal{U} -centered Euclidean distance matrices are equal to the dissimilarity matrices. Thus, to apply distance correlation methods to this type of problem, one only needs to obtain the Euclidean representation. Existing software implementations of classical MDS can be applied to obtain the representation derived above. For example, classical MDS based on the method outlined in Mardia [18] is implemented in the R function `cmdscale`, which is in the `stats` package for R. The `cmdscale` function includes options to apply the additive constant of Cailliez [2] and to specify the dimension. The matrix of points \mathbf{V} is returned in the component `points`. For an exact representation, we can specify the dimension argument equal to $n - 2$.

Example 1. To illustrate application of Theorem 2 for non-Euclidean dissimilarities, we computed the Bray-Curtis dissimilarity matrix of the `iris setosa` data, a four-dimensional data set available in R. The Bray-Curtis dissimilarity defined in Cox and Cox [3, Table 1.1] as

$$\delta_{ij} = \frac{1}{p} \frac{\sum_k |x_{ik} - x_{jk}|}{\sum_k (x_{ik} + x_{jk})}, \quad x_i, x_j \in \mathbb{R}^p,$$

is not a distance since it does not satisfy the triangle inequality. A Bray-Curtis method is available in the `distance` function of the `ecodist` package for R [6]. We find a configuration of 50 points in \mathbb{R}^{48} that have \mathcal{U} -centered distances equal to the \mathcal{U} -centered dissimilarities. The MDS computations are handled by the R function `cmdscale`. Function `Ucenter`, which implements \mathcal{U} -centering, is in the R package `pdcor` [26].

```
> x <- iris[1:50, 1:4]
> iris.d <- distance(x, method="bray-curtis")
> AU <- Ucenter(iris.d)
> v <- cmdscale(as.dist(AU), k=48, add=TRUE)$points
```

The points `v` are a 50×48 data matrix, of 50 points in \mathbb{R}^{48} . Next we compare the \mathcal{U} -centered distance matrix of the points `v` with the original object `AU` from \mathcal{H}_n :

```
> VU <- Ucenter(v)
> all.equal(AU, VU)
[1] TRUE
```

The last line of output shows that the points `v` returned by `cmdscale` have \mathcal{U} -centered distance matrix `VU` equal to our original element `AU` of the Hilbert space.

◇

Example 1 shows that the sample distance covariance can be defined for dissimilarities via the inner product in \mathcal{H}_n . Alternately one can compute $\mathcal{V}_n^2(\mathbf{U}, \mathbf{V})$, where \mathbf{U}, \mathbf{V} are the Euclidean representations corresponding to the two \mathcal{U} -centered dissimilarity matrices that exist by Theorem 2. Using the corresponding definitions of distance variance, sample distance correlation for dissimilarities is well defined by (3.8) or $\mathcal{R}_n^2(\mathbf{U}, \mathbf{V})$. Similarly one can define `pdCov` and `pdCor` when one or more of the dissimilarity matrices of the samples is not Euclidean distance. However, as in the case of Euclidean distance, we need Theorems 3-4 in Section 4 to define the corresponding population coefficients, and develop a test of independence.

3.4. Simplified computing formula for `pdCor`. Let

$$R_{x,y}^* := \begin{cases} \frac{(\tilde{A} \cdot \tilde{B})}{|\tilde{A}| |\tilde{B}|}, & |\tilde{A}| |\tilde{B}| \neq 0; \\ 0, & |\tilde{A}| |\tilde{B}| = 0, \end{cases} \quad (3.8)$$

where $\tilde{A} = \tilde{A}(x), \tilde{B} = \tilde{B}(y)$ are the \mathcal{U} -centered distance matrices of the samples x and y , and $|\tilde{A}| = (\tilde{A} \cdot \tilde{A})^{1/2}$. The statistics $R_{x,y}^*$ and $R^*(x, y; z)$ take values in $[-1, 1]$, but they are measured in units comparable to the squared distance correlation $\mathcal{R}_n^2(\mathbf{X}, \mathbf{Y})$.

Proposition 2. *If $(1 - (R_{x,z}^*)^2)(1 - (R_{y,z}^*)^2) \neq 0$, a computing formula for $R_{x,y;z}^*$ in Definition (3.7) is*

$$R_{x,y;z}^* = \frac{R_{x,y}^* - R_{x,z}^* R_{y,z}^*}{\sqrt{1 - (R_{x,z}^*)^2} \sqrt{1 - (R_{y,z}^*)^2}}. \quad (3.9)$$

See Appendix A.2 for a proof.

Equation (3.9) provides a simple and familiar form of computing formula for the partial distance correlation. The computational algorithm is easily implemented, as summarized below.

Algorithm to compute partial distance correlation $R_{x,y;z}^$ from Euclidean distance matrices $A = (|x_i - x_j|)$, $B = (|y_i - y_j|)$, and $C = (|z_i - z_j|)$:*

(i) Compute the \mathcal{U} -centered matrices \tilde{A} , \tilde{B} , and \tilde{C} , using

$$\tilde{A}_{i,j} = a_{i,j} - \frac{a_{i.}}{n-2} - \frac{a_{.j}}{n-2} + \frac{a_{..}}{(n-1)(n-2)}, \quad i \neq j,$$

and $\tilde{A}_{i,i} = 0$.

(ii) Compute inner products and norms using

$$(\tilde{A} \cdot \tilde{B}) = \frac{1}{n(n-3)} \sum_{i \neq j} \tilde{A}_{i,j} \tilde{B}_{i,j}, \quad |\tilde{A}| = (\tilde{A} \cdot \tilde{A})^{1/2}$$

and $R_{x,y}^*$, $R_{x,z}^*$, and $R_{y,z}^*$ using $R_{x,y}^* = \frac{(\tilde{A} \cdot \tilde{B})}{|\tilde{A}| |\tilde{B}|}$.

(iii) If $R_{x,z}^2 \neq 1$ and $R_{y,z}^2 \neq 1$

$$R_{x,y;z}^* = \frac{R_{x,y}^* - R_{x,z}^* R_{y,z}^*}{\sqrt{1 - (R_{x,z}^*)^2} \sqrt{1 - (R_{y,z}^*)^2}},$$

otherwise apply the definition (3.7).

Note that it is not necessary to explicitly compute the projections, when (3.9) is applied. The above algorithm has a straightforward translation to code; see e.g. the *pdcor* package [26] for an implementation in R.

4. POPULATION COEFFICIENTS AND INFERENCE

4.1. Population coefficients. The population distance covariance has been defined in terms of the joint and marginal characteristic functions of the random vectors. Here we give an equivalent definition following Lyons [16], who generalizes distance correlation to separable Hilbert spaces. Instead of starting with the distance matrices $(a_{ij}) = (|x_i - x_j|_p)$ and $(b_{ij}) = (|y_i - y_j|_q)$, the starting point of the population definition are the bivariate distance functions $a(x, x') := |x - x'|_p$ and $b(y, y') = |y - y'|_q$, where x, x' are realizations of the random variables X and y, y' are realizations of the random variable Y .

We can also consider the random versions. Let $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ be random variables with finite expectations. The random distance functions are $a(X, X') := |X - X'|_p$ and $b(Y, Y') = |Y - Y'|_q$. Here the primed random variable X denotes an independent and identically distributed (iid) copy of the variable X , and similarly Y, Y' are iid.

The population operations of double centering involves expected values with respect to the underlying population random variable. For a given random variable X with cdf F_X , we define the corresponding *double centering function with respect to X* as

$$\begin{aligned} \hat{A}_X(x, x') &:= a(x, x') - \int_{\mathbb{R}^p} a(x, x') dF_X(x') - \int_{\mathbb{R}^p} a(x, x') dF_X(x) \\ &\quad + \int_{\mathbb{R}^p} \int_{\mathbb{R}^p} a(x, x') dF_X(x') dF_X(x), \end{aligned} \quad (4.1)$$

provided the integrals exist.

Here $\hat{A}_X(x, x')$ is a real valued function of two realizations of X , and the subscript X references the underlying random variable. Similarly for X, X' iid with cdf F_X ,

we define the random variable \widehat{A}_X as an abbreviation for $\widehat{A}_X(X, X')$, which is a random function of (X, X') . Similarly we define

$$\begin{aligned} \widehat{B}_Y(y, y') &:= b(y, y') - \int_{\mathbb{R}^q} b(y, y') dF_Y(y') - \int_{\mathbb{R}^q} b(y, y') dF_Y(y) \\ &\quad + \int_{\mathbb{R}^q} \int_{\mathbb{R}^q} b(y, y') dF_Y(y') dF_Y(y), \end{aligned}$$

and the random function $\widehat{B}_Y := \widehat{B}_Y(Y, Y')$.

Now for X, X' iid, and Y, Y' iid, such that X, Y have finite expectations, the population distance covariance $\mathcal{V}(X, Y)$ is defined by

$$\mathcal{V}^2(X, Y) := E[\widehat{A}_X \widehat{B}_Y]. \quad (4.2)$$

The definition (4.2) of $\mathcal{V}^2(X, Y)$ is equivalent to the original definition (1.1). However, as we will see in the next sections, (4.2) is an appropriate starting point to develop the corresponding definition of pdCov and pdCor population coefficients.

More generally, we can consider *dissimilarity functions* $a(x, x')$. In this paper, a dissimilarity function is a symmetric function $a(x, x') : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ with $a(x, x) = 0$. The corresponding random dissimilarity functions $a(X, X')$ are random variables such that $a(X, X') = a(X', X)$, and $a(X, X) = 0$.

Double centered dissimilarities are formally defined by the same equations as double centered distances in (4.1).

We also define the population counterpart of \mathcal{U} -centering. If $x = x'$, or $Pr(X = x) = 1$, or $Pr(X = x') = 1$, or $Pr(X = X') = 1$ then $\widetilde{A}_X(x, x') := 0$. Otherwise

$$\begin{aligned} \widetilde{A}_X(x, x') &:= a(x, x') - \frac{\int a(x, x') dF_X(x')}{Pr(X' \neq x)} \\ &\quad - \frac{\int a(x, x') dF_X(x)}{Pr(X \neq x')} + \frac{\iint a(x, x') dF_X(x') dF_X(x)}{Pr(X \neq X')}. \end{aligned} \quad (4.3)$$

We denote by \widetilde{A}_X the random variable: $\widetilde{A}_X := \widetilde{A}_X(X, X')$.

Note that for continuous X , the population definition (4.3) of \mathcal{U} -centering is the same as population double centering (4.1) with the diagonal set to zero.

For Theorem 3 in Section 4.2, we need the invariance of \mathcal{U} -centering with respect to the addition of a constant to all but the self-dissimilarities. Let $a^{(c)}(x, x')$ denote the modified dissimilarity function obtained by adding the constant c to all but the self-dissimilarities of $a(x, x')$. Similarly $\widetilde{A}_X^{(c)}(x, x')$ and $\widetilde{A}_X^{(c)} = \widetilde{A}_X^{(c)}(X, X')$ are the modified \mathcal{U} -centered functions with respect to X , where population \mathcal{U} -centering is defined by (4.3).

Lemma 2 (Invariance). *Let $X \in \mathbb{R}^p$ with finite expectation and let $a(x, x') : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ be a dissimilarity function such that $a(x, x') = O(|x| + |x'|)$. Then for all constants $c \in \mathbb{R}$,*

$$\widetilde{A}_X^{(c)}(x, x') = \widetilde{A}_X(x, x'),$$

for all $x \in \mathbb{R}^p, x' \in \mathbb{R}^p$.

See the appendix (A.4) for a proof of Lemma 2.

The following lemma establishes that linear combinations of \mathcal{U} -centered dissimilarities are \mathcal{U} -centered dissimilarities.

Lemma 3. *Suppose that $X \in \mathbb{R}^p$, $Y \in \mathbb{R}^q$, $a(x, x')$ is a dissimilarity on $\mathbb{R}^p \times \mathbb{R}^p$, and $b(y, y')$ is a dissimilarity on $\mathbb{R}^q \times \mathbb{R}^q$. Let $\tilde{A}_X(x, x')$ and $\tilde{B}_Y(y, y')$ denote the dissimilarity obtained by \mathcal{U} -centering $a(x, x')$ and $b(y, y')$, respectively. Then if c_1 and c_2 are real scalars,*

$$c_1 \tilde{A}_X(x, x') + c_2 \tilde{B}_Y(y, y') = \tilde{D}_T(t, t'),$$

where $T = [X, Y] \in \mathbb{R}^p \times \mathbb{R}^q$, and $\tilde{D}_T(t, t')$ is the result of \mathcal{U} -centering $d(t, t') = c_1 a(x, x') + c_2 b(y, y')$.

See Appendix A.5 for the proof.

The linear span of \mathcal{U} -centered distance functions $\tilde{A}_X(x, x')$ is a subspace of the space of \mathcal{U} -centered dissimilarity functions with the property $a(x, x') = O(|x| + |x'|)$. In this case all integrals in (4.3) and \tilde{A}_X are finite if X has finite expectation. The linear span of the random functions \tilde{A}_X for random vectors X with $E|X| < \infty$ is clearly a linear space, such that the linear extension of (4.2)

$$E[\tilde{A}_X \tilde{B}_Y] = \mathcal{V}^2(X, Y)$$

to the linear span is an inner product space or pre-Hilbert space; its completion with respect to the metric arising from its inner product

$$(\tilde{A}_X \cdot \tilde{B}_Y) := E[\tilde{A}_X \tilde{B}_Y] \tag{4.4}$$

(and norm) is a Hilbert space which we denote by \mathcal{H} .

Once we have a Hilbert space we can apply similar projections as in Section 2 to arrive at the definition of population partial distance covariance, which is formally the same as (3.9).

4.2. Representation theorems for the population case. Let $X \in \mathbb{R}^p$, be a random variable with cdf F_X , and denote by x, x' two realizations of X . In the following, ℓ_2 consists of all square summable sequences $\{x_i\}_{i=1}^\infty$, together with inner product $(x \cdot y) = \sum_{i=1}^\infty x_i y_i$. This defines a separable Hilbert space, which is a measure space with respect to the sigma algebra of Borel sets.

Theorem 3. *Suppose that $X \in \mathbb{R}^p$ is a random variable with finite fourth moment, $a(x, x') : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ is a dissimilarity function with $a(x, x') = O(|x| + |x'|)$. Then for all x in \mathbb{R}^p , there exists a*

$$\Psi(x) = (\Psi_1(x), \Psi_2(x), \dots) \in \ell_2, \quad \Psi_k(x) \in L_2(\mathbb{R}^p), \quad k = 1, 2, \dots,$$

such that the \mathcal{U} -centered function of the distance $|\Psi(x) - \Psi(x')|$ with respect to X at (X, X') equals $\tilde{A}_X(X, X')$ with probability one.

Proof. Let $d(x, x') = -\frac{1}{2}a^2(x, x')$ and let $\hat{D}_X(x, x')$ denote the result of double centering $d(x, x')$ as defined in (4.1). Let $\Psi_k(x) \in L_2(\mathbb{R}^p)$, $k = 1, 2, \dots$, be an eigenfunction of the integral operator with kernel $\hat{D}_X(x, x')$. That is, there exist $\lambda_k \in \mathbb{R}$, $k = 1, 2, \dots$, such that

$$\int \hat{D}_X(x, x') \Psi(x) dF_X(x) = \lambda_k \Psi_k(x').$$

If the operator with kernel $\hat{D}_X(x, x')$ is positive semi-definite, all eigenvalues λ_k , $k = 1, 2, \dots$, are nonnegative. In this case let Ψ_k , $k = 1, 2, \dots$ be the eigenfunctions

that have positive eigenvalues $\lambda_k > 0$. Normalize Ψ_k such that its squared norm is $\int \Psi_k^2(x) dF_X(x) = \lambda_k$. Then by extension of MDS if $a(x, x') \geq 0$,

$$a(x, x') = |\Psi(x) - \Psi(x')| = \left\{ \sum_k (\Psi_k(x) - \Psi_k(x'))^2 \right\}^{1/2},$$

which is the distance between the ℓ_2 vectors $\Psi(x)$ and $\Psi(x')$.

If $\widehat{D}_X(x, x')$ is not positive semi-definite, we need to solve a version of the additive constant problem, restated for the population case. We just need to suppose that $a(x, x')$ is a *nonnegative* dissimilarity function such that $a(x, x') = O(|x| + |x'|)$. Nonnegativity of $a(x, x')$ can be replaced by *bounded from below* because a sufficiently large c can be chosen to make the modified distances $a^{(c)}(x, x')$ nonnegative. Although the function $a^{(c)}(x, x')$ depends on c , as in the finite (sample) case, the \mathcal{U} -centered distance function $\widetilde{A}_X^{(c)}(x, x')$ does not depend on c by our Lemma 2.

To prove the existence of c^* in the population additive constant problem, we need to check that the eigenvalue sequence λ_k is bounded. This property of the eigenvalue sequence follows from the theory of Hilbert–Schmidt integral operators if we can guarantee that the integral operator is compact (see e.g. Riesz and Sz.-Nagy [24, Chapter VI]). This compactness is guaranteed if $E[(\widehat{D}_X(X, X'))^2] < \infty$, which is true if $E[|X|^4] < \infty$ because $\widehat{D}_X(x, x')$ is essentially a quadratic function. From the spectral theory of compact integral operators we know that

$$\widehat{D}_X(x, x') = \sum_k \Psi_k(x) \Psi_k(x')$$

in the L_2 convergence sense and thus

$$E[\widehat{D}_X(X, X)] = \sum_k E[\Psi_k(X)^2] = \sum_k \lambda_k < \infty.$$

It is also true that $E[\Psi(X)] = 0$, because $E[\widehat{D}_X(x, X)] = 0$ a.s.

The kernel $\widehat{A}_X(x, x')$ also defines a compact integral operator and when $a(x, x')$ is Euclidean distance, $\widehat{A}_X(x, x')$ is actually the same as the negative of the kernel $h(x, x')$ in Székely and Rizzo [35, Eq.8.1]. There we have also seen that the sequence of eigenvalues is not only bounded, the sum of the nonnegative eigenvalues is also finite. As in the finite (sample) case if $a(x, x') \geq 0$ the boundedness of the eigenvalues of $\widehat{D}_X(x, x')$ and $\widehat{A}_X(x, x')$ imply that c^* is finite.

After this what might not be clear is the following. If the function $a(x, x')$ can take negative values and if $a(x, x')$ is not bounded from below, then we cannot simply add a big enough constant to all off-diagonal values to make all of them nonnegative. In this case we can use the randomness of $a(X, X')$ and the fact that $P(|X| > c) \rightarrow 0$ as $c \rightarrow +\infty$. If we add a constant c to all $\{a(x, x') : x \neq x'\}$ then the corresponding $\Psi^{(c)}(X)$ is stochastically bounded; that is, $P(|\Psi^{(c)}(X)| > C) \rightarrow 0$ as $C \rightarrow \infty$. This implies the existence of an accumulation point $\Psi(X)$, in the sense that there exists a sequence $C_k \rightarrow \infty$ as $k \rightarrow \infty$ such that $P(|\Psi(X) - \Psi^{(C_k)}(X)| > \epsilon) \rightarrow 0$ for arbitrary small $\epsilon > 0$. \square

We can now formulate the definition of population partial distance correlation and an important theorem on the population version of partial distance covariance. Assume that X, Y, Z satisfy the hypotheses of Theorem 3.

Definition 5 (Population partial distance covariance). *Introduce the scalar coefficients*

$$\alpha := \frac{\mathcal{V}^2(X, Z)}{\mathcal{V}^2(Z, Z)}, \quad \beta := \frac{\mathcal{V}^2(Y, Z)}{\mathcal{V}^2(Z, Z)}.$$

If $\mathcal{V}^2(Z, Z) = 0$ define $\alpha = \beta = 0$. The \mathcal{U} -centered projections of \tilde{A}_X and \tilde{B}_Y onto the orthogonal complement of \tilde{C}_Z in our Hilbert space \mathcal{H} are defined

$$P_{Z^\perp}(X) := \tilde{A}_X(X, X') - \alpha \tilde{C}_Z(Z, Z'), \quad P_{Z^\perp}(Y) := \tilde{B}_Y(Y, Y') - \beta \tilde{C}_Z(Z, Z'),$$

or in short $P_{Z^\perp}(X) = \tilde{A}_X - \alpha \tilde{C}_Z$ and $P_{Z^\perp}(Y) = \tilde{B}_Y - \beta \tilde{C}_Z$, where \tilde{C}_Z denotes \mathcal{U} -centered with respect to the random variable Z . These formulas are the population counterparts of (3.4).

The population partial distance covariance is defined by the inner product

$$(P_{Z^\perp}(X) \cdot P_{Z^\perp}(Y)) := E[(\tilde{A}_X - \alpha \tilde{C}_Z) \cdot (\tilde{B}_Y - \beta \tilde{C}_Z)].$$

Definition 6 (Population pdCor). *Population partial distance correlation is defined*

$$\mathcal{R}^*(X, Y; Z) := \frac{(P_{Z^\perp}(X) \cdot P_{Z^\perp}(Y))}{|P_{Z^\perp}(X)| |P_{Z^\perp}(Y)|},$$

where $|P_{Z^\perp}(X)| = (P_{Z^\perp}(X) \cdot P_{Z^\perp}(X))^{1/2}$. If $|P_{Z^\perp}(X)| |P_{Z^\perp}(Y)| = 0$ we define $\mathcal{R}^*(X, Y; Z) = 0$. If $\alpha = \beta = 0$, we have $(P_{Z^\perp}(X) \cdot P_{Z^\perp}(Y)) = E[\tilde{A}_X \cdot \tilde{B}_Y] = \mathcal{V}^2(X, Y)$, and $\mathcal{R}^*(X, Y; Z) = \mathcal{R}^*(X, Y)$.

Since orthogonality in the space of \mathcal{U} -centered distance functions means independence, we have the following representation theorem for the population.

Theorem 4. *Suppose that X, Y, Z are arbitrary random vectors taking values in not-necessarily identical Euclidean spaces, and X, Y, Z have finite fourth moments. Let α , and β be the scalars introduced in Definition 5. Then there exist ℓ_2 -valued random variables U and V with finite second moments, such that:*

- (i) *The \mathcal{U} -centered distance function of U with respect to (X, Z) at $((X, Z), (X', Z'))$ equals $\tilde{A}_X - \alpha \tilde{C}_Z$ with probability one.*
- (ii) *The \mathcal{U} -centered distance function of V with respect to (Y, Z) at $((Y, Z), (Y', Z'))$ equals $\tilde{B}_Y - \beta \tilde{C}_Z$ with probability one.*
- (iii) *The random variables U and V are independent of Z .*
- (iv) *$\text{pdCov}(X, Y; Z) = (U \cdot V)$ and $\text{pdCor}(X, Y; Z) = 0$ if and only if U and V are independent.*

Proof. By Theorem 3 and Lemma 3 there exists a Borel function $\Psi^* : \mathbb{R}^{p+r} \rightarrow \ell_2$ whose \mathcal{U} -centered distance function with respect to (X, Z) at $((X, Z), (X', Z'))$ is equal to $\tilde{A}_X - \alpha \tilde{C}_Z$ with probability one. Introduce $U := \Psi^*(X, Z)$. Similarly define V . The random variables U and V have finite second moments and they are independent of Z , because $((\tilde{A}_X - \alpha \tilde{C}_Z) \cdot \tilde{C}_Z) = 0$ and $((\tilde{B}_Y - \beta \tilde{C}_Z) \cdot \tilde{C}_Z) = 0$. Finally,

$$((\tilde{A}_X - \alpha \tilde{C}_Z) \cdot (\tilde{B}_Y - \beta \tilde{C}_Z)) = 0$$

if and only if U and V are independent. □

More generally, as discussed above, Theorem 4 holds for X, Y, Z in separable Hilbert spaces.

4.3. Discussion. Let us rephrase Theorem 4 for easier interpretation. We can say that a random variable U *represents* a dissimilarity function d if the \mathcal{U} -centered distance function of U with respect to some random variable W at (W, W') is the same as the \mathcal{U} -centered d with respect to W at (W, W') . Using this language, Theorem 4 means the following. Under the conditions of the theorem one can always find two random variables, U and V , such that they represent the dissimilarity functions $\tilde{A}_X(x, x') - \alpha\tilde{C}_Z(z, z')$ and $\tilde{B}_X(y, y') - \beta\tilde{C}_Z(z, z')$, respectively, both U and V are independent of Z , and $\text{pdCor}(X, Y; Z) = 0$ if and only if U and V are independent. Thus if we measure the dependence of X and Y with Z “removed” by considering the residuals $\tilde{A}_X(x, x') - \alpha\tilde{C}_Z(z, z')$ and $\tilde{B}_Y(y, y') - \beta\tilde{C}_Z(z, z')$, they can be represented by random variables U and V , respectively, such that U, V are independent if and only if $\text{pdCor}(X, Y; Z) = 0$.

One does not always need the ℓ_2 valued random variables U and V for a representation, however. For example, if X, Y , and Z are real valued random variables that are jointly Gaussian, then we know that the “residuals” are also jointly Gaussian; therefore instead of ℓ_2 valued random variables U and V it is sufficient to work with certain linear combinations of X, Y, Z and they are also jointly Gaussian random variables.

Theorem 4 is a generalization of a classical result on partial correlation (pcor) for multivariate Gaussian distributions: $\text{pcor}(X, Y; Z) = 0$ if and only if X and Y are conditionally independent given Z . See Baba, Shibata and Sibuya (2004) [1]. As this property does not hold for the general case, Theorem 4 is an important result for non-Gaussian distributions.

Partial distance correlation is a scalar quantity that captures dependence, while conditional inference (conditional distance correlation) is a more complex notion as it is a function of the condition. Although one might hope that partial distance correlation $\mathcal{R}^*(X, Y; Z) = 0$ if and only if X , and Y are conditionally independent given Z , this is, however, not the case. Both notions capture *overlapping* aspects of dependence, but mathematically they are not equivalent. To see this observe that $\mathcal{R}^*(X, Y; Z) = 0$ if and only if $\mathcal{R}^*(X, Y) = \mathcal{R}^*(X, Z)\mathcal{R}^*(Y, Z)$. With the help of this equation one can construct examples that show $\mathcal{R}^*(X, Y; Z) = 0$ is not equivalent to conditional independence of X and Y given Z .

4.4. Inference. We have defined partial distance correlation, which is bias corrected and has an informal interpretation, but we also want to determine whether $R^*(x, y; z)$ is significantly different from zero, which is not a trivial problem. Although theory and implementation for a consistent test of multivariate independence based on distance covariance is available [31, 25], clearly the implementation of the test by randomization is not directly applicable to the problem of testing for zero partial distance correlation.

Since $\text{pdCor}(X, Y; Z) = 0$ if and only if $\text{pdCov}(X, Y; Z) = 0$, we develop a test for $H_0 : \text{pdCov}(X, Y; Z) = 0$ vs $H_1 : \text{pdCov}(X, Y; Z) \neq 0$ based on the inner product $(P_{z^\perp}(x) \cdot P_{z^\perp}(y))$. Unlike the problem of testing independence of X and Y , however, we do not have the distance matrices or samples corresponding to the projections $P_{z^\perp}(x)$ and $P_{z^\perp}(y)$. For a test we need to consider that $P_{z^\perp}(x)$ and $P_{z^\perp}(y)$ are arbitrary elements of \mathcal{H}_n .

We have proved (Theorem 2) that $P_z(x)$ can be represented as the \mathcal{U} -centered distance matrix of some configuration of n points \mathbf{U} in a Euclidean space \mathbb{R}^p , $p \leq n - 2$, and that $P_z(y)$ has such a representation \mathbf{V} in \mathbb{R}^q , $q \leq n - 2$. Hence,

a test for $\text{pdCov}(X, Y; Z) = 0$ can be defined by applying the distance covariance test statistic $\mathcal{V}_n^2(\mathbf{U}, \mathbf{V})$. The resulting test procedure is practical to apply, with similar computational complexity as the original dCov test of independence. One can apply the test of independence implemented by permutation bootstrap in the `dcov.test` function of the energy package [25] for R. Alternately a test based on the inner product (3.5) is implemented in the `pdcor` [26] package.

5. EXAMPLES AND APPLICATIONS

In this section we summarize simulation results for tests of the null hypothesis of zero partial distance correlation and two applications. We compared our simulation results with two types of tests based on linear correlation. Methods for dissimilarities are illustrated using data on genetic distances for samples of maize in Example 6, and variable selection is discussed in Example 7.

5.1. Partial distance covariance test. The test for zero partial distance correlation is a test of whether the inner product ($P_{Z^\perp}(X) \cdot P_{Z^\perp}(Y)$) of the projections is zero. We obtain the corresponding samples \mathbf{U}, \mathbf{V} such that the \mathcal{U} -centered distances of \mathbf{U} and \mathbf{V} are identical to $P_z(x)$ and $P_z(y)$, respectively, using classical (metric) MDS. We used the `cmdscale` function in R to obtain the samples \mathbf{U} and \mathbf{V} in both cases. The `pdcov` test applies the inner product of the double centered distance matrices of \mathbf{U} and \mathbf{V} as the test statistic. Alternately, one could apply the original `dcov` test, to the joint sample (\mathbf{U}, \mathbf{V}) . These two test statistics are essentially equivalent, but the `pdcov` test applies an unbiased statistic defined by (3.3), while the `dcov` test applies a biased statistic, $n\mathcal{V}_n^2(\mathbf{U}, \mathbf{V})$ (2.2). Both tests are implemented as permutation tests, where for a test of $\text{pdCor}(X, Y; Z) = 0$ the sample indices of the X sample are randomized for each replicate to obtain the sampling distribution of the test statistic under the null hypothesis.

The term ‘permutation test’ is sometimes restricted to refer to the exact test, which is implemented by generating all possible permutations. Except for very small samples, it is not computationally feasible to generate all permutations, so a large number of randomly generated permutations are used. This approach, which we have implemented, is sometimes called a randomization test to distinguish it from an exact permutation test.

In our simulations the `pdcov` and `dcov` tests for zero partial distance correlation were equivalent under null or alternative hypotheses in the sense that the type 1 error rates and estimated power agreed to within one standard error. In power comparisons, therefore, we reported only the `pdcov` result.

5.2. Partial correlation. The linear partial correlation $r(x, y; z)$ measures the partial correlation between one dimensional data vectors x and y with z removed (or controlling for z). The sample partial correlation coefficient is

$$r(x, y; z) = \frac{r(x, y) - r(x, z)r(y, z)}{\sqrt{(1 - r(x, z)^2)(1 - r(y, z)^2)}}, \quad (5.1)$$

where $r(x, y)$ denotes the linear (Pearson) sample correlation. The partial correlation test is usually implemented as a t test (see e.g. Legendre [13, p. 40]). In examples where x, y, z are one dimensional, we have included the partial correlation t -test in comparisons. However, for small samples and some non-Gaussian

data, the type 1 error rate is inflated. In cases where type 1 error rate of *pcor* was not controlled we did not report power.

5.3. The partial Mantel test. The partial Mantel test is a test of the hypothesis that there is a linear association between the distances of X and Y , controlling for Z . This extension of the Mantel test [17] was proposed by Smouse, et al. [29] for a partial correlation analysis on three distance matrices. The Mantel and partial Mantel tests are commonly applied in community ecology (see e.g. Legendre and Legendre [14]), population genetics, sociology, etc.

Let U_1, U_2, U_3 denote the upper triangles of the $n \times n$ distance matrices of the X , Y and Z samples, respectively. Then the partial Mantel test statistic is the sample linear partial correlation between the $n(n-1)/2$ elements of U_1 and U_2 controlling for U_3 . Let u_1, u_2, u_3 be the corresponding data vectors obtained by representing U_1, U_2, U_3 as vectors. The partial Mantel statistic is $r(u_1, u_2; u_3)$, computed using formula (5.1).

Since u_i are not iid samples, the usual t test is not applicable, so the partial Mantel test is usually implemented as a permutation (randomization) test. See Legendre [13] for a detailed algorithm and simulation study comparing different methods of computing a partial Mantel test statistic. Based on the results reported by Legendre, we implemented the method of permutation of the raw data. The algorithm is given in detail on page 44 by Legendre [13], and it is very similar to the algorithm we have applied for the *energy* tests (*pdcov* and *dcov* tests for zero *pdCor*). This method (permutation of raw data) for the partial Mantel test is implemented in the *ecodist* package [6] and also the *vegan* package [20] for R. For simulations, the `mantel` function in the *ecodist* package, which is implemented mainly in compiled code, is much faster than the `mantel` or `mantel.partial` functions in the *vegan* package, which are implemented in R.

remark 4. Both the *pdcov* and partial Mantel (*Mantel*) tests are based on distances. One may ask “is distance covariance different or more general than covariance of distances?” The answer is yes; it can be shown that

$$\text{dCov}^2(X, Y) = \text{Cov}(|X - X'|, |Y - Y'|) - 2 \text{Cov}(|X - X'|, |Y - Y''|),$$

where $(X, Y), (X', Y')$, and (X'', Y'') are iid. The *dcov* tests are tests of independence of X and Y ($\text{dCov}^2(X, Y) = 0$), while the Mantel test is a test of the hypothesis $\text{Cov}(|X - X'|, |Y - Y'|) = 0$. An example of dependent data such that their distances are uncorrelated but $\text{dCov}^2(X, Y) > 0$ is given e.g. by Lyons [16]. Thus, distance covariance tests are more general than Mantel tests, in the sense that distance covariance measures all types of departures from independence.

5.4. Simulation design. In each permutation test $R = 999$ replicates are generated and the estimated p -value is computed as

$$\hat{p} = \frac{1 + \sum_{k=1}^R I(T^{(k)} \geq T_0)}{1 + R},$$

where $I(\cdot)$ is the indicator function, T_0 is the observed value of the test statistic, and $T^{(k)}$ is the statistic for the k -th sample. The test is rejected at significance level α if $\hat{p} \leq \alpha$. The partial correlation test (*pcor*) is also included for comparison. It is implemented as a t -test [11].

Type 1 error rates and estimated power are determined by a simulation size of 10,000 tests in each case; for $n = 10$ the number of tests is 100,000. The standard error is at most 0.005 (0.0016 for $n = 10$).

Example 2. This example is a comparison of type 1 error on uncorrelated standard normal data. The vectors X , Y , and Z are each iid standard normal. Results summarized in Table 1 show that type 1 error rates for *pdcov*, *dcov*, and *partial Mantel* tests are within two se of the nominal significance level, while type 1 error rates for *pcor* are inflated for $n = 10, 20$.

TABLE 1. Example 2: Type 1 error rates at nominal significance level α for uncorrelated standard trivariate normal data.

n	α	pdcov	dcov	Mantel	pcor	α	pdcov	dcov	Mantel	pcor
10	0.05	0.051	0.051	0.050	0.090	0.10	0.101	0.102	0.100	0.143
20	0.05	0.048	0.048	0.051	0.064	0.10	0.098	0.100	0.100	0.120
30	0.05	0.052	0.051	0.047	0.059	0.10	0.095	0.097	0.098	0.107
50	0.05	0.051	0.051	0.051	0.058	0.10	0.104	0.104	0.106	0.105
100	0.05	0.049	0.050	0.048	0.057	0.10	0.100	0.100	0.096	0.105

Example 3. This example is the same as Example 2 except that the X variable is standard lognormal rather than standard normal. Results summarized in Table 2 demonstrate that the type 1 error rates are controlled at their nominal significance level for *pdcov*, *dcov*, and *partial Mantel* tests (all implemented as permutation tests), while the *pcor t* test has inflated type 1 error rates for $n \leq 30$.

Results for type 1 error in Examples 2 and 3 are also summarized in Figures 1(a) and 1(b), respectively, for significance level $\alpha = 0.10$.

TABLE 2. Example 3: Type 1 error rates at nominal significance level α for uncorrelated non-normal data. The partial distance correlation and other statistics are computed for standard lognormal X and standard normal Y , removing standard normal Z .

n	α	pdcov	dcov	Mantel	pcor	α	pdcov	dcov	Mantel	pcor
10	0.05	0.052	0.053	0.050	0.090	0.10	0.103	0.103	0.099	0.142
20	0.05	0.049	0.050	0.054	0.071	0.10	0.101	0.102	0.102	0.122
30	0.05	0.049	0.049	0.047	0.062	0.10	0.104	0.104	0.099	0.112
50	0.05	0.050	0.051	0.047	0.058	0.10	0.104	0.104	0.097	0.108
100	0.05	0.051	0.052	0.048	0.054	0.10	0.101	0.101	0.097	0.106

Example 4. In this example, power of tests is compared for correlated trivariate normal data with standard normal marginal distributions. It is a modification of Example 2 such that the variables X , Y , and Z are each *correlated* standard normal. The pairwise correlations are $\rho(X, Y) = \rho(X, Z) = \rho(Y, Z) = 0.5$. The power comparison summarized in Figure 2(a) shows that *pdcov* has higher power than *pcor* or *partial Mantel* tests. The simulation parameters are identical to those described for type 1 error simulations.

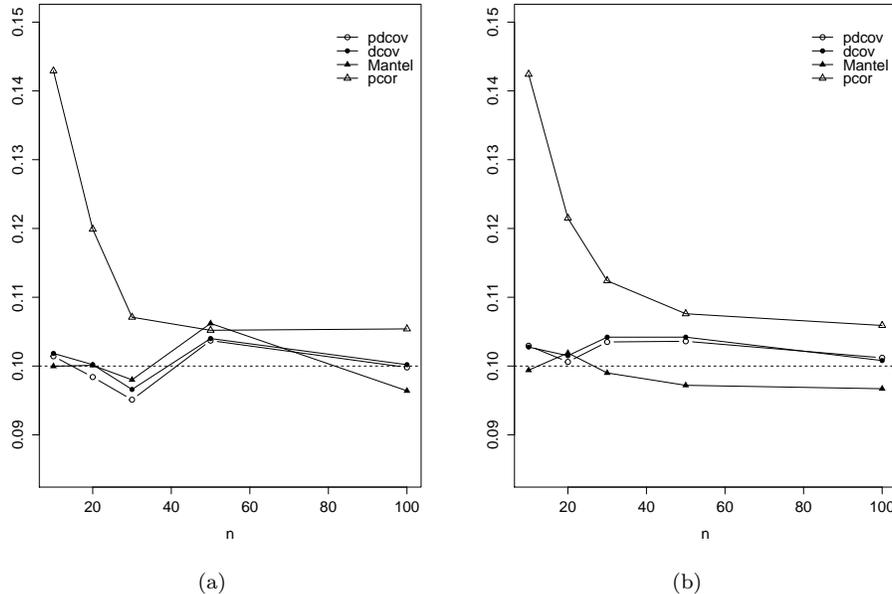


FIGURE 1. Type 1 error rates for partial distance covariance, inner product (dcov) test, partial Mantel test, and partial correlation test, at significance level $\alpha = 0.10$. Figure (a) summarizes Example 2 (uncorrelated standard normal data). Figure (b) summarizes Example 3 (uncorrelated non-normal data).

Example 5. This example presents a power comparison for correlated non-normal data. It is a modification of Example 3 such that the variables X , Y , and Z are each *correlated*, X is standard lognormal, while Y and Z are each standard normal. The pairwise correlations are $\rho(\log X, Y) = \rho(\log X, Z) = \rho(Y, Z) = 0.5$. The power comparison summarized in Figure 2(b) shows that *pdcor* has higher power than *pcor* or *partial Mantel* tests. The simulation parameters are identical to those described for type 1 error simulations. Again the *pdcov* test has superior power performance. The relative performance of the *pcor* and *partial Mantel* tests are reversed (partial Mantel test with lowest power) in this example compared with Example 4.

Example 6. In this example, we compute distance correlation for dissimilarities and apply the distance covariance inner product test. The data are two sets of dissimilarities for seven maize populations from Piepho [21], originally from Reif et al. [23]. The two dissimilarity arrays are given in Table 3, with genetic distance by simple sequence repeats, modified Roger's distances below the diagonal and mid-parent heterosis for days to silking of crosses above the diagonal. Note: comparing with Reif et al. [23, Tables 2, 4], there is an error in row 2, column 5 of Piepho [21, Table 2], where the coefficient for (Pop21, Pop29) should be 0.4, not -0.4.

Piepho [21] compared two tests for association between the matrices; the two-sided t -test based on Pearson's correlation of distances, and the permutation Mantel

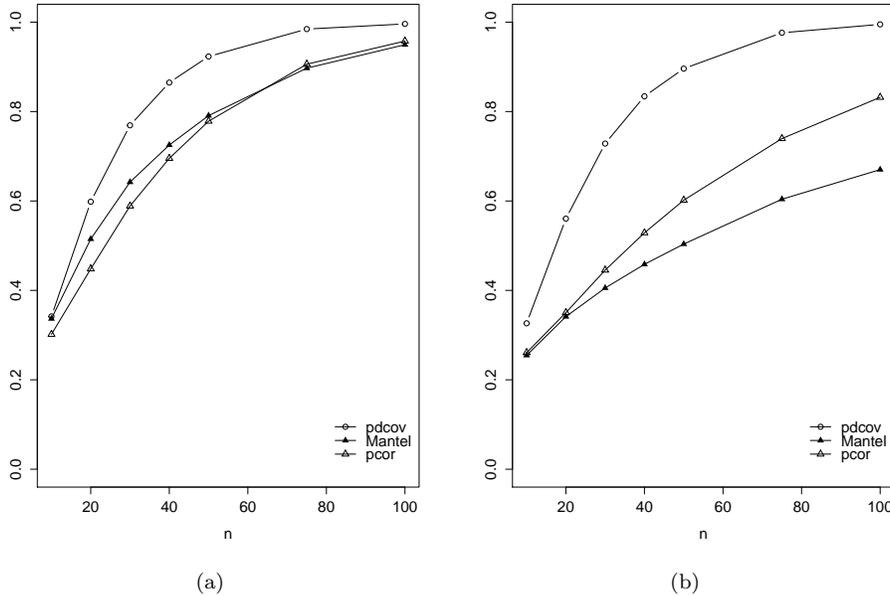


FIGURE 2. Power comparisons for partial distance covariance, partial Mantel test, and partial correlation test at significance level $\alpha = 0.10$. Figure (a) summarizes Example 4 (correlated standard normal data). Figure (b) summarizes Example 5 (correlated non-normal data).

TABLE 3. Dissimilarities for Example 6: Two genetic distances for seven maize populations from Reif et al. [23].

	Pool24	Pop21	Pop22	Pop25	Pop29	Pop32	Pop43
Pool24		0.50	-0.40	0.70	-0.30	-0.70	-1.30
Pop21	0.22		-0.40	-0.40	0.40	-0.70	-1.20
Pop22	0.20	0.22		-0.60	-1.50	-1.20	-1.80
Pop25	0.22	0.27	0.25		-0.90	-0.90	-0.50
Pop29	0.22	0.24	0.23	0.26		-0.70	-0.20
Pop32	0.27	0.30	0.28	0.26	0.28		-0.90
Pop43	0.25	0.29	0.27	0.28	0.27	0.32	

test with 100,000 permutations. Piepho found that the tests lead to different decisions. The Pearson correlation $r = -0.44$ had p -value 0.0466 for a two-sided t -test, which is significant at the 5% level. The permutation Mantel test gave a non-significant p -value of 0.1446. The different conclusions may be due to fact that the t test applied for the Mantel statistic is on the liberal side.

We repeated the analysis and also use the data to illustrate the application of `dCor` and the inner product test to dissimilarities. Using the `cor.test` function in R we compute sample correlation $r = -0.4388$ and a two-sided p -value of .04662.

For the permutation Mantel test, we used the *ecodist* implementation of Mantel test. The three reported p -values are for the alternative hypotheses $\rho \leq 0$, $\rho \geq 0$, and $\rho \neq 0$, respectively, so for a two tailed Mantel test we have p -value 0.14370.

```
ecodist::mantel(dy ~ dx, nperm = 1e+05, nboot = 0)
## mantelr   pval1   pval2   pval3   llim   ulim
## -0.43875  0.92562  0.07486  0.14370  0.00000  0.00000
```

The distance covariance inner product test can be applied to dissimilarities following the method that we developed for partial distance covariance. Starting with the two dissimilarities which we stored in matrices `d1` and `d2`, we \mathcal{U} -center them, obtain the two representations in Euclidean space, and apply the inner product test. If we need only the distance correlation statistic R_{xy}^* , this can be computed by the `Rxy` function in *pdcor* package, which implements formula (3.8).

```
AU <- Ucenter.mat(d1)
BU <- Ucenter.mat(d2)
u <- cmdscale(as.dist(AU), add = TRUE, k = 5)$points
v <- cmdscale(as.dist(BU), add = TRUE, k = 5)$points

Rxy(u, v)
[1] -0.2734559
```

The function `dcovIP.test` in the *pdcor* package [26] for R implements the inner product `dCov` test. Alternately one could apply the original `dCov` test to the points in the Euclidean space.

```
dcovIP.test(u, v, R = 9999)
##
## dCov inner product test of independence
##
## data:  replicates 9999
## n * V^* = -0.0124, p-value = 0.832
## sample estimates:
##      R^*
## -0.2735
```

The `dCov` inner product test is not significant, with p -value 0.832.

This example illustrates that the distance correlation or distance covariance in the Hilbert space \mathcal{H}_n is readily applied to measuring or testing dependence between samples represented by dissimilarities. It is more general than a test of association such as the Mantel test, as we have shown in Remark 4.

Example 7 (Variable selection). This example considers the prostate cancer data from a study by Stamey et al. [30], and is discussed in Hastie, Tibshirani, and Friedman [8, Ch. 3] in the context of variable selection. The data is from men who were about to have a radical prostatectomy. The response variable *lpsa* measures log PSA (log of the level of prostate-specific antigen). The predictor variables under consideration are eight clinical measures:

lcavol	log cancer volume
lweight	log prostate weight
age	age
lbph	log of the amount of benign prostatic hyperplasia
svi	seminal vesicle invasion
lcp	log of capsular penetration
gleason	Gleason score
pgg45	percent of Gleason scores 4 or 5

Here the goal was to fit a linear model to predict the response *lpsa* given one or more of the predictor variables above. The train/test set indicator is in the last column of the data set. For comparison with the discussion and analysis in Hastie et al. [8], we standardized each variable, and used the training set of 67 observations for variable selection.

Feature screening by distance correlation has been investigated by Li et al. [15]. In this example we introduce a partial distance correlation criterion for variable selection. For simplicity, we implement a simple variant of forward selection. This criterion, when applied for a linear model, can help to identify possible important variables that have strong non-linear association with the response, and thus help researchers to improve a linear model by transforming variables or improve prediction by extending to a nonlinear model.

In the initial step of pdCor forward selection, the first variable to enter the model is the variable x_j for which distance correlation $R_{x_j,y}$ with response y is largest. After the initial step, we have a model with one predictor x_j , and we compute $\text{pdCor}(y, x_k; x_j)$, for the variables $x_k \neq x_j$ not in the model, then select the variable x_k for which $\text{pdCor}(y, x_k; x_j)$ is largest. Then continue, at each step computing $\text{pdCor}(y, x_j; w)$ for every variable x_j not yet in the model, where w is the vector of predictors currently in the model. The variable to enter next is the one that maximizes $\text{pdCor}(y, x_j; w)$.

According to the pdCor criterion, the variables selected enter the model in the order:

lcavol, lweight, svi, gleason, lbph, pgg45.

If we set a stopping rule at 5% significance for the pdCor coefficient, then we stop after adding *lbph* (or possibly after adding *pgg45*). At the step where *gleason* and *lbph* enter, the p -values for significance of pdCor are 0.016 and 0.001, respectively. The p -value for *pgg45* is approximately 0.05.

The models selected by pdCor, best subset method, and lasso (Hastie et al. [8, Table 3.3, p. 63]) are:

pdCor:	$\text{lpsa} \sim \text{lcavol} + \text{lweight} + \text{svi} + \text{gleason};$
best subsets:	$\text{lpsa} \sim \text{lcavol} + \text{lweight};$
lasso:	$\text{lpsa} \sim \text{lcavol} + \text{lweight} + \text{svi} + \text{lbph}.$

The order of selection for ordinary forward stepwise selection (Cp) is

lcavol, lweight, svi, lbph, pgg45, age, lcp, gleason.

Comparing pdCor forward selection with lasso and forward stepwise selection, we see that the results are similar, but *gleason* is not in the lasso model and enters last in the forward stepwise selection, while it is the fourth variable to enter the pdCor selected model. The raw Gleason Score is an integer from 2 to 10 which is used to measure how aggressive is the tumor, based on a prostate biopsy. Plotting

the data (see Figure 3) we can observe that there is a strong *non-linear* relation between *gleason* and *lpsa*.

This example illustrates that partial distance correlation has practical use in variable selection and in model checking. If we were using pdCor only to check the model selected by another procedure, it would show in this case that there is some nonlinear dependence remaining between the response and the predictors excluded from the lasso model or the best subsets model. Using the pdCor selection approach, we also learn which of the remaining predictors may be important.

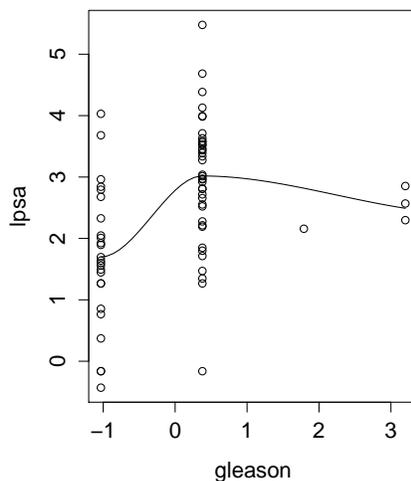


FIGURE 3. Scatter plot of response *lpsa* vs Gleason score with a loess smoother.

Finally, it is useful to recall that pdCor has more flexibility to handle predictors that are multi-dimensional. One may want groups of variables to enter or leave the model as a set. It is often the case that when dimension of the feature space is high, many of the predictor variables are highly correlated. In this case, methods such as Partial Least Squares are sometimes applied, where a small set of derived predictors (linear combinations of features) become the predictor variables. Using methods of partial distance correlation, one could evaluate the subsets (as predictor sets) without reducing the multivariate observation to a real number via linear combination. The pdCor coefficient can be computed for multivariate predictors (and for multivariate response).

6. SUMMARY

Partial distance covariance and partial distance correlation coefficients are measures of dependence of two random vectors X , Y , controlling for a third random vector Z , where X , Y , and Z are in arbitrary, not necessarily equal dimensions. These definitions are based on an unbiased distance covariance statistic, replacing the original double centered distance matrices of the original formula with \mathcal{U} -centered distance matrices. The \mathcal{U} -centered distance covariance is the inner product

in the Hilbert space \mathcal{H}_n of \mathcal{U} -centered distance matrices for samples of size n , and it is unbiased for the squared population distance covariance.

Each element in the Hilbert \mathcal{H}_n is shown to be the \mathcal{U} -centered distance matrix of some configuration of n points in a Euclidean space \mathbb{R}^p , $p \leq n - 2$. The proof and the solution are obtained through application of theory of classical MDS. This allows one to define dCor and dCov for dissimilarity matrices, and provides a statistically consistent test of independence based on the inner product statistic. This methodology also provides a similar test of the hypothesis of zero partial distance correlation based on the inner product.

All pdCor and pdCov methods have been implemented and simulation studies illustrate that the tests control the type 1 error rate at its nominal level. Power performance was superior compared with power of partial correlation and partial Mantel tests for association. Methods have been implemented and illustrated for non-Euclidean dissimilarity matrices. A ‘pdCor forward selection’ method was applied to select variables for a linear model, with a significance test as the stopping rule. More sophisticated selection methods will be investigated in future work. With the flexibility to handle multivariate response and/or multivariate predictor, pdCor offers a new method to extend the variable selection toolbox. Software is available in the *energy* [25] package for R, and the *pdcor* [26] package for R.

APPENDIX A. PROOFS OF STATEMENTS

A.1. Proof of Proposition 1. Proposition 1 asserts that $(\tilde{A} \cdot \tilde{B})$ is an unbiased estimator of the population coefficient $\mathcal{V}^2(X, Y)$. When the terms of $(\tilde{A} \cdot \tilde{B})$ are expanded, we have a linear combination of terms $a_{ij}b_{kl}$. The expected values of these terms differ according to the number of subscripts that agree. Define

$$\begin{aligned} \alpha &:= E[a_{kl}] = E[|X - X'|], & \beta &:= E[b_{kl}] = E[|Y - Y'|], & k \neq l, \\ \delta &:= E[a_{kl}b_{kj}] = E[|X - X'| |Y - Y''|], & & j, k, l \text{ distinct}, \\ \gamma &:= E[a_{kl}b_{kl}] = E[|X - X'| |Y - Y'|], & & k \neq l, \end{aligned}$$

where $(X, Y), (X', Y'), (X'', Y'')$ are iid. Due to symmetry, the expected value of each term in the expanded expression $(\tilde{A} \cdot \tilde{B})$ is proportional to one of $\alpha\beta$, δ , or γ , so the expected value of $(\tilde{A} \cdot \tilde{B})$ can be written as a linear combination of $\alpha\beta$, δ , and γ .

The population coefficient can be written as (see [32, Theorem 7])

$$\begin{aligned} \mathcal{V}^2(X, Y) &= E[|X - X'| |Y - Y'|] + E[|X - X'|] E[|Y - Y'|] - 2E[|X - X'| |Y - Y''|] \\ &= \gamma + \alpha\beta - 2\delta. \end{aligned}$$

Let us adopt the notation $\tilde{a}_{k.} := \frac{a_{k.}}{n-2}$, $\tilde{a}_{.l} := \frac{a_{.l}}{n-2}$, and $\tilde{a}_{..} := \frac{a_{..}}{(n-1)(n-2)}$, where $a_{k.} = \sum_{l=1}^n a_{kl}$, $a_{.l} = \sum_{k=1}^n a_{kl}$, and $a_{..} = \sum_{k,l=1}^n a_{kl}$. Similarly define $\tilde{b}_{k.}$, $\tilde{b}_{.l}$, and $\tilde{b}_{..}$. Then

$$n(n-3)(\tilde{A} \cdot \tilde{B}) = \sum_{k \neq l} \begin{Bmatrix} a_{kl}b_{kl} & -a_{kl}\tilde{b}_{k.} & -a_{kl}\tilde{b}_{.l} & +a_{kl}\tilde{b}_{..} \\ -\tilde{a}_{k.}b_{kl} & +\tilde{a}_{k.}\tilde{b}_{k.} & +\tilde{a}_{k.}\tilde{b}_{.l} & -\tilde{a}_{k.}\tilde{b}_{..} \\ -\tilde{a}_{.l}b_{kl} & +\tilde{a}_{.l}\tilde{b}_{k.} & +\tilde{a}_{.l}\tilde{b}_{.l} & -\tilde{a}_{.l}\tilde{b}_{..} \\ +\tilde{a}_{..}b_{kl} & -\tilde{a}_{..}\tilde{b}_{k.} & -\tilde{a}_{..}\tilde{b}_{.l} & +\tilde{a}_{..}\tilde{b}_{..} \end{Bmatrix}$$

$$\begin{aligned}
&= \sum_{k \neq l} a_{kl} b_{kl} - \sum_k a_k \tilde{b}_k. & - \sum_l a_l \tilde{b}_l & + a_{..} \tilde{b}_{..} \\
&- \sum_k \tilde{a}_k b_k. & + (n-1) \sum_k \tilde{a}_k \tilde{b}_k. & + \sum_{k \neq l} \tilde{a}_k \tilde{b}_l & - (n-1) \sum_k \tilde{a}_k \tilde{b}_{..} \\
&- \sum_l \tilde{a}_l b_l & + \sum_{k \neq l} \tilde{a}_l \tilde{b}_k. & + (n-1) \sum_l \tilde{a}_l \tilde{b}_l & - (n-1) \sum_l \tilde{a}_l \tilde{b}_{..} \\
&+ \tilde{a}_{..} b_{..} & - (n-1) \sum_k \tilde{a}_{..} \tilde{b}_k. & - (n-1) \sum_l \tilde{a}_{..} \tilde{b}_l & + n(n-1) \tilde{a}_{..} \tilde{b}_{..}
\end{aligned}$$

Let

$$T_1 = \sum_{k \neq l} a_{kl} b_{kl}, \quad T_2 = a_{..} b_{..}, \quad T_3 = \sum_k a_k b_k.$$

Then

$$\begin{aligned}
n(n-3)(\tilde{A} \cdot \tilde{B}) &= \left\{ \begin{array}{cccc} T_1 & -\frac{T_3}{n-2} & -\frac{T_3}{n-2} & +\frac{T_2}{(n-1)(n-2)} \\ -\frac{T_3}{n-2} & +\frac{(n-1)T_3}{(n-2)^2} & +\frac{T_2-T_3}{(n-2)^2} & -\frac{T_2}{(n-2)^2} \\ -\frac{T_3}{n-2} & +\frac{T_2-T_3}{(n-2)^2} & +\frac{(n-1)T_3}{(n-2)^2} & -\frac{T_2}{(n-2)^2} \\ +\frac{T_2}{(n-1)(n-2)} & -\frac{T_2}{(n-2)^2} & -\frac{T_2}{(n-2)^2} & +\frac{nT_2}{(n-1)(n-2)^2} \end{array} \right\} \\
&= T_1 - \frac{T_2}{(n-1)(n-2)^2} - \frac{2T_3}{n-2}.
\end{aligned}$$

It is easy to see that $E[T_1] = n(n-1)\gamma$. By expanding the terms of T_2 and T_3 , and combining terms that have equal expected values, one can obtain

$$E[T_2] = n(n-2)\{(n-2)(n-3)\alpha\beta + 2\gamma + 4(n-2)\delta\},$$

and

$$E[T_3] = n(n-1)\{(n-2)\delta + \gamma\}.$$

Then

$$\begin{aligned}
E[(\tilde{A} \cdot \tilde{B})] &= \frac{1}{n(n-3)} E \left[T_1 - \frac{T_2}{(n-1)(n-2)^2} - \frac{2T_3}{n-2} \right] \\
&= \frac{1}{n(n-3)} \left\{ \frac{n^3 - 5n^2 + 6n}{n-2} \gamma + n(n-3)\alpha\beta + (6n - 2n^2)\delta \right\} \\
&= \gamma + \alpha\beta - 2\delta = \mathcal{V}^2(X, Y).
\end{aligned}$$

□

A.2. Proof of Proposition 2. Here the inner product is (3.3) and the ‘vectors’ are \mathcal{U} -centered elements of the Hilbert space \mathcal{H}_n . Equation (3.9) can be derived from (3.7) using similar algebra with inner products as used to obtain the representation

$$\frac{\langle x_{z^\perp}, y_{z^\perp} \rangle}{\sqrt{\langle x_{z^\perp}, x_{z^\perp} \rangle \langle y_{z^\perp}, y_{z^\perp} \rangle}} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{1-r_{xz}^2} \sqrt{1-r_{yz}^2}}.$$

for the linear correlation r (see e.g. Huber [9]). The details for pdCor are as follows. If either $|\tilde{A}|$ or $|\tilde{B}| = 0$ then $(\tilde{A}, \tilde{B}) = 0$ so by definition $R^*(x, y; z) = 0$, $R_{xz}^* R_{yz}^* = 0$, and (3.9) is also zero.

If $|\tilde{A}||\tilde{B}| \neq 0$ but $|\tilde{C}| = 0$, then $R_{x,z}^* = R_{y,z}^* = 0$. In this case $P_{z^\perp}(x) = \tilde{A}$ and $P_{z^\perp}(y) = \tilde{B}$, so that

$$R^*(x, y; z) = \frac{(P_{z^\perp}(x) \cdot P_{z^\perp}(y))}{|P_{z^\perp}(x)||P_{z^\perp}(y)|} = \frac{(\tilde{A} \cdot \tilde{B})}{|\tilde{A}||\tilde{B}|} = R_{xy}^*,$$

which equals expression (3.9) since $R_{x,z}^* = R_{y,z}^* = 0$.

Suppose that none of $|\tilde{A}|, |\tilde{B}|, |\tilde{C}|$ are zero. Then

$$\begin{aligned}
R^*(x, y; z) &= (P_{z^\perp}(x) \cdot P_{z^\perp}(y)) \\
&= \left(\left\{ \tilde{A} - \frac{(\tilde{A} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})} \cdot \tilde{C} \right\} \cdot \left\{ \tilde{B} - \frac{(\tilde{B} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})} \cdot \tilde{C} \right\} \right) \\
&= (\tilde{A} \cdot \tilde{B}) - \frac{2(\tilde{A} \cdot \tilde{C})(\tilde{B} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})} + \frac{(\tilde{A} \cdot \tilde{C})(\tilde{B} \cdot \tilde{C})(\tilde{C} \cdot \tilde{C})}{(\tilde{C} \cdot \tilde{C})^2} \\
&= |\tilde{A}||\tilde{B}| \left\{ \frac{(\tilde{A} \cdot \tilde{B})}{|\tilde{A}||\tilde{B}|} - \frac{(\tilde{A} \cdot \tilde{C})(\tilde{B} \cdot \tilde{C})}{|\tilde{A}||\tilde{C}||\tilde{B}||\tilde{C}|} \right\} \\
&= |\tilde{A}||\tilde{B}| \{R_{xy}^* - R_{xz}^*R_{yz}^*\}.
\end{aligned}$$

Similarly, in the denominator of (3.7) we have

$$\begin{aligned}
&\sqrt{|\tilde{A}|^2 (1 - (R_{xz}^*)^2) |\tilde{B}|^2 (1 - (R_{yz}^*)^2)} \\
&= |\tilde{A}||\tilde{B}| \sqrt{(1 - (R_{xz}^*)^2)(1 - (R_{yz}^*)^2)}.
\end{aligned}$$

Thus, if the denominator of (3.7) is not zero, the factor $|\tilde{A}||\tilde{B}|$ cancels from the numerator and denominator and we obtain (3.9). \square

A.3. Proof of Lemma 1. The sum of the first row of \tilde{A} is

$$\begin{aligned}
\tilde{A}_{1.} &= \sum_{j=2}^n \left(a_{1j} - \frac{a_{1.}}{n-2} - \frac{a_{.j}}{n-2} + \frac{a_{..}}{(n-1)(n-2)} \right) \\
&= a_{1.} - \frac{(n-1)a_{1.}}{n-2} - \frac{a_{..} - a_{.1}}{n-2} + \frac{(n-1)a_{..}}{(n-1)(n-2)} = 0.
\end{aligned}$$

Similarly each of the rows of \tilde{A} sum to zero, and by symmetry the columns also sum to zero, which proves statement (i).

Statements (ii) and (iii) follow immediately from (i). In (iv) let b_{ij} be the ij -th element of B . Then $b_{ii} = 0$, and for $i \neq j$, $b_{ij} = \tilde{A}_{ij} + c$. Hence $b_{i.} = b_{.j} = (n-1)c$, $i \neq j$ and $b_{..} = n(n-1)c$. Therefore

$$\tilde{B}_{ij} = b_{ij} - \frac{2(n-1)c}{n-2} + \frac{nc}{n-2} = \tilde{A}_{ij} + c + \frac{(2-n)c}{n-2} = \tilde{A}_{ij}, \quad i \neq j,$$

which proves (iv). \square

A.4. Proof of Lemma 2. For $x = x'$ we have $\tilde{A}_X(x, x) := 0$ and $\tilde{A}_X^{(c)}(x, x) = 0$ by definition. Write $a^{(c)}(x, x') = a(x, x') + cI(x \neq x')$, where $I(\cdot)$ is the indicator function. Suppose that $Pr(X = x) < 1$, $Pr(X = x') < 1$, and $Pr(X = X') < 1$.

Then applying definition (4.3), we have

$$\begin{aligned}
\tilde{A}_X^{(c)}(x, x') &= a(x, y) + c - \frac{E[a(x, X') + cI(X' \neq x)]}{Pr(X' \neq x)} \\
&\quad - \frac{E[a(X, x') + cI(X \neq x')]}{Pr(X \neq x')} + \frac{E[a(X, X') + cI(X \neq X')]}{Pr(X \neq X')} \\
&= a(x, y) + c - \frac{E[a(x, X')] + cPr(X' \neq x)}{Pr(X' \neq x)} \\
&\quad - \frac{E[a(X, x')] + cPr(X \neq x')]}{Pr(X \neq x')} + \frac{E[a(X, X')] + cPr(X \neq X')]}{Pr(X \neq X')} \\
&= a(x, x') - \frac{Ea(x, X')}{Pr(X' \neq x)} - \frac{Ea(X, x')}{Pr(X \neq x')} + \frac{Ea(X, X')}{Pr(X \neq X')} \\
&= \tilde{A}_X(x, x').
\end{aligned}$$

Since $a(x, x') + cI(x \neq x')$ is a symmetric dissimilarity such that $a(x, x) = 0$, for the cases $P(X \neq x') = 0$, $P(X' \neq x) = 0$, and $P(X \neq X') = 0$ we have $\tilde{A}_X^{(c)}(x, x') = \tilde{A}_X(x, x') = 0$, by definition of \mathcal{U} -centering. \square

A.5. Proof of Lemma 3. It is clear that $c_1\tilde{A}_X(x, x')$ is identical to the \mathcal{U} -centered version of the dissimilarity $c_1a(x, x')$. It remains to show that the sum of two arbitrary elements $\tilde{A}_X(x, x') + \tilde{B}_Y(y, y')$ is a \mathcal{U} -centered dissimilarity function.

Let $T = [X, Y] \in \mathbb{R}^p \times \mathbb{R}^q$. Consider the dissimilarity function $d(t, t') := a(x, x') + b(y, y')$, where $t = [x, y]$ and $t' = [x', y']$. Then for $t \neq t'$, $Pr(T \neq t') > 0$, $Pr(T \neq t) > 0$, and $Pr(T \neq T') > 0$,

$$\begin{aligned}
\tilde{D}_X(t, t') &= a(x, x') + b(y, y') - \frac{\int [a(x, x') + b(y, y')] I(t \neq t') dF_T(t')}{Pr(T' \neq t)} \\
&\quad - \frac{\int [a(x, x') + b(y, y')] I(t \neq t') dF_T(t)}{Pr(T \neq t')} \\
&\quad + \frac{\iint [a(x, x') + b(y, y')] I(t \neq t') dF_T(t') dF_T(t)}{Pr(T \neq T')} \\
&= a(x, x') + b(y, y') - E[a(x, X') + b(y, Y') | T' \neq t] \\
&\quad - E[a(X, x') + b(Y, y') | T \neq t'] + E[a(X, X') + b(Y, Y') | T \neq T'].
\end{aligned}$$

Thus

$$\begin{aligned}
\tilde{D}_X(t, t') &= a(x, x') + b(y, y') - E[a(x, X') | X' \neq x] - E[b(y, Y') | Y' \neq y] \\
&\quad - E[a(X, x') | X \neq x'] - E[b(Y, y') | Y \neq y'] \\
&\quad + E[a(X, X') | X \neq X'] + E[b(Y, Y') | Y \neq Y'] \\
&= \tilde{A}_X(x, x') + \tilde{B}_Y(y, y').
\end{aligned}$$

If $t = t'$, or $Pr(T = t') = 1$, or $Pr(T = t) = 1$, or $Pr(T = T') = 1$, then by definition $\tilde{D}_X(t, t') = 0$. In any of these cases we have by definition that $\tilde{A}_X(x, x') = 0$ and $\tilde{B}_Y(y, y') = 0$, so

$$\tilde{D}_X(t, t') = \tilde{A}_X(x, x') + \tilde{B}_Y(y, y')$$

holds for all (x, x') and (y, y') . \square

ACKNOWLEDGEMENTS

The authors thank Russell Lyons for several helpful suggestions and comments on a preliminary draft of this paper.

REFERENCES

- [1] Baba, K., Shibata, R. and Sibuya, M. (2004). Partial correlation and conditional correlation as measures of conditional independence. *Australian and New Zealand Journal of Statistics* **46** (4): 657–664.
- [2] Cailliez, F. (1983). The analytical solution of the additive constant problem. *Psychometrika*, **48**, 343–349.
- [3] Cox, T. F. and Cox, M. A. A. (2001). *Multidimensional Scaling*, Second edition. Chapman and Hall.
- [4] Dueck, J., Edelmann, D. Gneiting, T. and Richards, D. (2012). The affinely invariant distance correlation, submitted for publication. <http://arxiv.org/pdf/1210.2482.pdf>
- [5] Feuerverger, A. (1993). A consistent test for bivariate dependence. *International Statistical Review*, **61**, 419–433.
- [6] Goslee, S. C. and Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**(7), 1–19.
- [7] Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis, *Biometrika* **53**, 325–328.
- [8] Hastie, T., Tibshirani, R. and Friedman, J. (2009). *Elements of Statistical Learning, second edition*, Springer, New York.
- [9] Huber, John (1981). Partial and Semipartial Correlation—A Vector Approach, *The Two-Year College Mathematics Journal*, **12**/2, 151–153, JSTOR: 3027381.
- [10] Josse, J. and Holmes, S. (2013). Measures of dependence between random vectors and tests of independence. Literature review. <http://arxiv.org/abs/1307.7383>.
- [11] Kim, S. (2012). ppcor: Partial and Semi-partial (Part) correlation. R package version 1.0. <http://CRAN.R-project.org/package=ppcor>
- [12] Kong, J., Klein, B. E. K., Klein, R., Lee, K., and Wahba, G. (2012). Using distance correlation and SS-ANOVA to assess associations of familial relationships, lifestyle factors, diseases, and mortality, *Proc. of the National Acad. of Sciences*, **109** (50), 20352–20357. DOI:10.1073/pnas.1217269109
- [13] Legendre, P. (2000). Comparison of permutation methods for the partial correlation and partial Mantel tests, *J. Statist. Comput. Simul.*, **67**, 37–73.
- [14] Legendre, P. and Legendre, L. (2012). *Numerical Ecology*, 3rd English Edition, Elsevier.
- [15] Li, R., Zhong, W. and Zhu, L. (2012). Feature Screening via Distance Correlation Learning, *Journal of the American Statistical Association*, **107**/499, 1129–1139. DOI: 10.1080/01621459.2012.695654
- [16] Lyons, R. (2013). Distance covariance in metric spaces, *Ann. Probab.* **41**/5 , 3284–3305.
- [17] Mantel, N. (1967). The detection of disease clustering and a generalized regression approach, *Cancer Res.*, **27**, 209–220.
- [18] Mardia, K. V. (1978) Some properties of classical multidimensional scaling. *Communications in Statistics: Theory and Methods*, **7**(13), 1233–1241.
- [19] Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*, London: Academic Press.
- [20] Jari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O’Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens and Helene Wagner (2013). vegan: Community Ecology Package. R package version 2.0-7. <http://CRAN.R-project.org/package=vegan>
- [21] Piepho, H. P. (2005). Permutation tests for the correlation among genetic distances and measures of heterosis, *Theor. Appl. Genet.*, **111**, 95–99, DOI 10.1007/s00122-005-1995-7.
- [22] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- [23] Reif, J. C., A. E. Melchinger, X. C. Xia, M. L. Warburton, D. A. Hoisington, S. K. Vasal, G. Srinivasan, M. Bohn, and M. Frisch (2003). Genetic distance based on simple sequence repeats and heterosis in tropical maize populations, *Crop Science* **43**, no. 4, 1275–1282.

- [24] Riesz, F. and Sz.-Nagy, B. (1990). *Functional Analysis*, Dover Publications.
- [25] Rizzo, M. L. and Székely, G. J. (2014). energy: E-statistics (energy statistics). R package version 1.6.1. <http://CRAN.R-project.org/package=energy>
- [26] Rizzo, M. L. and Székely, G. J. (2013). pdcor: Partial distance correlation. R package version 1.0.0.
- [27] Schoenberg, I. J. (1935). Remarks to Maurice Fréchet’s article “Sur la definition axiomatique d’une classe d’espace distanciés vectoriellement applicable sur l’espace de Hilbert.” *Ann. Math.*, **36**, 724–732.
- [28] 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.
- [29] Smouse, P. E., Long, J. C. and Sokal, R. R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, **35**:62, 7–632.
- [30] Stamey, T. A., Kabalin, J. N., McNeal, J. E., Johnstone, I. M., Freiha, F., Redwine, E. A. and Yang, N. (1989). Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate: II. radical prostatectomy treated patients, *Journal of Urology* **141**(5), 1076–1083.
- [31] Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007) Measuring and testing independence by correlation of distances, *Ann. Statist.* **35**:6, 2769–2794. DOI: 10.1214/009053607000000505
- [32] Székely, G. J. and Rizzo, M. L. (2009). Brownian Distance Covariance, *Annals of Applied Statistics*, **3**(4), 1236–1265. DOI: 10.1214/09-AOAS312
- [33] Székely, G. J. and Rizzo, M. L. (2012). On the uniqueness of distance covariance. *Statistics & Probability Letters*, **82**:12, 2278–2282. DOI: 10.1016/j.spl.2012.08.007
- [34] Székely, G. J. and Rizzo, M. L. (2013). The distance correlation t-test of independence in high dimension, *J. Multivar. Anal.*, **117**, 193–213. DOI: 10.1016/j.jmva.2013.02.012
- [35] Székely, G. J. and Rizzo, M. L. (2013). Energy statistics: statistics based on distances. *Journal of Statistical Planning and Inference*, **143**:8, 1249–1272. DOI: 10.1016/j.jspi.2013.03.018
- [36] Torgerson, W. S. (1958). *Theory and Methods of Scaling*, New York: Wiley.
- [37] Wermuth, N. and Cox, D. R. (2013). Concepts and a case study for a flexible class of graphical Markov models. In Becker, C., Fried, R. and Kuhnt, S. (eds.) *Robustness and complex data structures. Festschrift in honour of Ursula Gather*. Springer, Heidelberg, 331–350.
- [38] Young, G. and Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, **3**, 19–22.

GÁBOR J. SZÉKELY, NATIONAL SCIENCE FOUNDATION, 4201 WILSON BLVD. #1025, ARLINGTON, VA 22230

E-mail address: gszekely@nsf.gov

MARIA L. RIZZO, DEPARTMENT OF MATHEMATICS AND STATISTICS, BOWLING GREEN STATE UNIVERSITY, BOWLING GREEN, OH 43403

E-mail address: mrizzo@bgsu.edu

URL: personal.bgsu.edu/~mrizzo