Bernoulli **20**(4), 2014, 2305–2330
DOI: 10.3150/13-BEJ558

The affinely invariant distance correlation

JOHANNES DUECK[1], DOMINIC EDELMANN[1], TILMANN GNEITING[2] and DONALD RICHARDS[3]

[1]*Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany*
[2]*Heidelberg Institute for Theoretical Studies and Karlsruhe Institute of Technology, HITS gGmbH, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany*
[3]*Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA.*
E-mail: richards@stat.psu.edu

Székely, Rizzo and Bakirov (*Ann. Statist.* **35** (2007) 2769–2794) and Székely and Rizzo (*Ann. Appl. Statist.* **3** (2009) 1236–1265), in two seminal papers, introduced the powerful concept of distance correlation as a measure of dependence between sets of random variables. We study in this paper an affinely invariant version of the distance correlation and an empirical version of that distance correlation, and we establish the consistency of the empirical quantity. In the case of subvectors of a multivariate normally distributed random vector, we provide exact expressions for the affinely invariant distance correlation in both finite-dimensional and asymptotic settings, and in the finite-dimensional case we find that the affinely invariant distance correlation is a function of the canonical correlation coefficients. To illustrate our results, we consider time series of wind vectors at the Stateline wind energy center in Oregon and Washington, and we derive the empirical auto and cross distance correlation functions between wind vectors at distinct meteorological stations.

Keywords: affine invariance; distance correlation; distance covariance; hypergeometric function of matrix argument; multivariate independence; multivariate normal distribution; vector time series; wind forecasting; zonal polynomial

1. Introduction

Székely, Rizzo and Bakirov [23] and Székely and Rizzo [20], in two seminal papers, introduced the distance covariance and distance correlation as powerful measures of dependence. Contrary to the classical Pearson correlation coefficient, the population distance covariance vanishes only in the case of independence, and it applies to random vectors of arbitrary dimensions, rather than to univariate quantities only.

As noted by Newton [14], the "distance covariance not only provides a *bona fide* dependence measure, but it does so with a simplicity to satisfy Don Geman's *elevator test* (i.e., a method must be sufficiently simple that it can be explained to a colleague in the time it takes to go between floors on an elevator)." In the case of the sample distance covariance, find the pairwise distances between the sample values for the first variable, and center the resulting distance matrix; then do the same for the second variable. The square of the sample distance covariance equals the average entry in the componentwise or Schur product of the two centered distance matrices. Given the theoretical appeal of the population quantity, and the striking simplicity of the sample version, it is not surprising that the distance covariance is experiencing a wealth of applications, despite having been introduced merely half a decade ago.

1350-7265

Specifically, let p and q be positive integers. For column vectors $s \in \mathbb{R}^p$ and $t \in \mathbb{R}^q$, denote by $|s|_p$ and $|t|_q$ the standard Euclidean norms on the corresponding spaces; thus, if $s = (s_1, \dots, s_p)'$ then

$$|s|_p = \left(s_1^2 + \cdots + s_p^2\right)^{1/2},$$

and similarly for $|t|_q$. For vectors u and v of the same dimension, p, we let $\langle u, v\rangle_p$ be the standard Euclidean scalar product of u and v. For jointly distributed random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, let

$$f_{X,Y}(s,t) = \mathbb{E}\exp\left[\mathrm{i}\langle s, X\rangle_p + \mathrm{i}\langle t, Y\rangle_q\right]$$

be the joint characteristic function of (X, Y), and let $f_X(s) = f_{X,Y}(s, 0)$ and $f_Y(t) = f_{X,Y}(0, t)$ be the marginal characteristic functions of X and Y, where $s \in \mathbb{R}^p$ and $t \in \mathbb{R}^q$. Székely *et al.* [23] introduced the *distance covariance* between X and Y as the nonnegative number $\mathcal{V}(X, Y)$ defined by

$$\mathcal{V}^2(X,Y) = \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \frac{|f_{X,Y}(s,t) - f_X(s) f_Y(t)|^2}{|s|_p^{p+1} |t|_q^{q+1}} \,\mathrm{d}s\,\mathrm{d}t, \tag{1.1}$$

where $|z|$ denotes the modulus of $z \in \mathbb{C}$ and

$$c_p = \frac{\pi^{(1/2)(p+1)}}{\Gamma((1/2)(p+1))}. \tag{1.2}$$

The *distance correlation* between X and Y is the nonnegative number defined by

$$\mathcal{R}(X,Y) = \frac{\mathcal{V}(X,Y)}{\sqrt{\mathcal{V}(X,X)\mathcal{V}(Y,Y)}} \tag{1.3}$$

if both $\mathcal{V}(X, X)$ and $\mathcal{V}(Y, Y)$ are strictly positive, and defined to be zero otherwise. For distributions with finite first moments, the distance correlation characterizes independence in that $0 \le \mathcal{R}(X, Y) \le 1$ with $\mathcal{R}(X, Y) = 0$ if and only if X and Y are independent.

A crucial property of the distance correlation is that it is invariant under transformations of the form

$$(X, Y) \longmapsto (a_1 + b_1 C_1 X, a_2 + b_2 C_2 Y), \tag{1.4}$$

where $a_1 \in \mathbb{R}^p$ and $a_2 \in \mathbb{R}^q$, b_1 and b_2 are nonzero real numbers, and the matrices $C_1 \in \mathbb{R}^{p\times p}$ and $C_2 \in \mathbb{R}^{q\times q}$ are orthogonal. However, the distance correlation fails to be invariant under the group of all invertible affine transformations of (X, Y), which led Székely *et al.* [23], pages 2784–2785, and Székely and Rizzo [20], pages 1252–1253, to propose an affinely invariant sample version of the distance correlation.

Adapting this proposal to the population setting, the *affinely invariant distance covariance* between distributions X and Y with finite second moments and nonsingular population covariance matrices Σ_X and Σ_Y, respectively, can be introduced as the nonnegative number $\widetilde{\mathcal{V}}(X, Y)$ defined by

$$\widetilde{\mathcal{V}}^2(X,Y) = \mathcal{V}^2\left(\Sigma_X^{-1/2} X, \Sigma_Y^{-1/2} Y\right). \tag{1.5}$$

The *affinely invariant distance correlation* between X and Y is the nonnegative number defined by

$$\widetilde{\mathcal{R}}(X,Y) = \frac{\widetilde{\mathcal{V}}(X,Y)}{\sqrt{\widetilde{\mathcal{V}}(X,X)\widetilde{\mathcal{V}}(Y,Y)}} \tag{1.6}$$

if both $\widetilde{\mathcal{V}}(X,X)$ and $\widetilde{\mathcal{V}}(Y,Y)$ are strictly positive, and defined to be zero otherwise. In the sample versions proposed by Székely *et al.* [23], the population quantities are replaced by their natural estimators. Clearly, the population affinely invariant distance correlation and its sample version are invariant under the group of invertible affine transformations, and in addition to satisfying this often-desirable group invariance property (Eaton [2]), they inherit the desirable properties of the standard distance dependence measures. In particular, $0 \le \widetilde{\mathcal{R}}(X,Y) \le 1$ and, for populations with finite second moments and positive definite covariance matrices, $\widetilde{\mathcal{R}}(X,Y) = 0$ if and only if X and Y are independent.

The remainder of the paper is organized as follows. In Section 2, we review the sample version of the affinely invariant distance correlation introduced by Székely *et al.* [23], and we prove that the sample version is strongly consistent. In Section 3, we provide exact expressions for the affinely invariant distance correlation in the case of subvectors from a multivariate normal population of arbitrary dimension, thereby generalizing a result of Székely *et al.* [23] in the bivariate case; our result is non-trivial, being derived using the theory of zonal polynomials and the hypergeometric functions of matrix argument, and it enables the explicit and efficient calculation of the affinely invariant distance correlation in the multivariate normal case.

In Section 4, we study the behavior of the affinely invariant distance measures for subvectors of multivariate normal populations in limiting cases as the Frobenius norm of the cross-covariance matrix converges to zero, or as the dimensions of the subvectors converge to infinity. We expect that these results will motivate and provide the theoretical basis for many applications of distance correlation measures for high-dimensional data.

As an illustration of our results, Section 5 considers time series of wind vectors at the Stateline wind energy center in Oregon and Washington; we shall derive the empirical auto and cross distance correlation functions between wind vectors at distinct meteorological stations. Finally, we provide in Section 6 a discussion in which we make a case for the use of the distance correlation and the affinely invariant distance correlation, which we believe to be appealing and powerful multivariate measures of dependence.

2. The sample version of the affinely invariant distance correlation

In this section, which is written primarily to introduce readers to distance correlation measures, we describe sample versions of the affinely invariant distance covariance and distance correlation as introduced by Székely *et al.* [23], pages 2784–2785, and Székely and Rizzo [20], pages 1252–1253.

First, we review the sample versions of the standard distance covariance and distance correlation. Given a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from jointly distributed random vectors

$X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, we set

$$\mathbf{X} = [X_1, \ldots, X_n] \in \mathbb{R}^{p \times n} \quad \text{and} \quad \mathbf{Y} = [Y_1, \ldots, Y_n] \in \mathbb{R}^{q \times n}.$$

A natural way of introducing a sample version of the distance covariance is to let

$$f^n_{\mathbf{X},\mathbf{Y}}(s,t) = \frac{1}{n} \sum_{j=1}^{n} \exp\bigl[\mathrm{i}\langle s, X_j\rangle_p + \mathrm{i}\langle t, Y_j\rangle_q\bigr]$$

be the corresponding empirical characteristic function, and to write $f^n_{\mathbf{X}}(s) = f^n_{\mathbf{X},\mathbf{Y}}(s,0)$ and $f^n_{\mathbf{Y}}(t) = f^n_{\mathbf{X},\mathbf{Y}}(0,t)$ for the respective marginal empirical characteristic functions. The *sample distance covariance* then is the nonnegative number $\mathcal{V}_n(\mathbf{X},\mathbf{Y})$ defined by

$$\mathcal{V}^2_n(\mathbf{X},\mathbf{Y}) = \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \frac{|f^n_{\mathbf{X},\mathbf{Y}}(s,t) - f^n_{\mathbf{X}}(s) f^n_{\mathbf{Y}}(t)|^2}{|s|_p^{p+1} |t|_q^{q+1}} \,\mathrm{d}s\,\mathrm{d}t,$$

where c_p is the constant given in (1.2).

Székely *et al.* [23], in a *tour de force*, showed that

$$\mathcal{V}^2_n(\mathbf{X},\mathbf{Y}) = \frac{1}{n^2} \sum_{k,l=1}^{n} A_{kl} B_{kl}, \tag{2.1}$$

where

$$a_{kl} = |X_k - X_l|_p, \qquad \bar{a}_{k\cdot} = \frac{1}{n}\sum_{l=1}^{n} a_{kl}, \qquad \bar{a}_{\cdot l} = \frac{1}{n}\sum_{k=1}^{n} a_{kl}, \qquad \bar{a}_{\cdot\cdot} = \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl}$$

and

$$A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot},$$

and similarly for $b_{kl} = |Y_k - Y_l|_q$, $\bar{b}_{k\cdot}$, $\bar{b}_{\cdot l}$, $\bar{b}_{\cdot\cdot}$, and B_{kl}, where $k, l = 1, \ldots, n$. Thus, the squared sample distance covariance equals the average entry in the componentwise or Schur product of the centered distance matrices for the two variables. The *sample distance correlation* then is defined by

$$\mathcal{R}_n(\mathbf{X},\mathbf{Y}) = \frac{\mathcal{V}_n(\mathbf{X},\mathbf{Y})}{\sqrt{\mathcal{V}_n(\mathbf{X},\mathbf{X})\mathcal{V}_n(\mathbf{Y},\mathbf{Y})}} \tag{2.2}$$

if both $\mathcal{V}_n(\mathbf{X},\mathbf{X})$ and $\mathcal{V}_n(\mathbf{Y},\mathbf{Y})$ are strictly positive, and defined to be zero otherwise. Computer code for calculating these sample versions is available in an R package by Rizzo and Székely [17].

Now let $S_{\mathbf{X}}$ and $S_{\mathbf{Y}}$ denote the usual sample covariance matrices of the data $\mathbf{X}$ and $\mathbf{Y}$, respectively. Following Székely *et al.* [23], page 2785, and Székely and Rizzo [20], page 1253, the *sample affinely invariant distance covariance* is the nonnegative number $\widetilde{\mathcal{V}}_n(\mathbf{X},\mathbf{Y})$ defined by

$$\widetilde{\mathcal{V}}^2_n(\mathbf{X},\mathbf{Y}) = \mathcal{V}^2_n\bigl(S_{\mathbf{X}}^{-1/2}\mathbf{X}, S_{\mathbf{Y}}^{-1/2}\mathbf{Y}\bigr) \tag{2.3}$$

if $S_{\mathbf{X}}$ and $S_{\mathbf{Y}}$ are positive definite, and defined to be zero otherwise. The *sample affinely invariant distance correlation* is defined by

$$\widetilde{\mathcal{R}}_n(\mathbf{X}, \mathbf{Y}) = \frac{\widetilde{\mathcal{V}}_n(\mathbf{X}, \mathbf{Y})}{\sqrt{\widetilde{\mathcal{V}}_n(\mathbf{X}, \mathbf{X})\widetilde{\mathcal{V}}_n(\mathbf{Y}, \mathbf{Y})}} \tag{2.4}$$

if the quantities in the denominator are strictly positive, and defined to be zero otherwise. The sample affinely invariant distance correlation inherits the properties of the sample distance correlation; in particular

$$0 \le \widetilde{\mathcal{R}}_n(\mathbf{X}, \mathbf{Y}) \le 1,$$

and $\widetilde{\mathcal{R}}_n(\mathbf{X}, \mathbf{Y}) = 1$ implies that $p = q$, that the linear spaces spanned by $\mathbf{X}$ and $\mathbf{Y}$ have full rank, and that there exist a vector $a \in \mathbb{R}^p$, a nonzero number $b \in \mathbb{R}$, and an orthogonal matrix $C \in \mathbb{R}^{p \times p}$ such that $S_{\mathbf{Y}}^{-1/2}\mathbf{Y} = a + bCS_{\mathbf{X}}^{-1/2}\mathbf{X}$.

Our next result shows that the sample affinely invariant distance correlation is a consistent estimator of the respective population quantity.

Theorem 2.1. *Let $(X, Y) \in \mathbb{R}^{p+q}$ be jointly distributed random vectors with positive definite marginal covariance matrices $\Sigma_X \in \mathbb{R}^{p\times p}$ and $\Sigma_Y \in \mathbb{R}^{q\times q}$, respectively. Suppose that $(X_1, Y_1), \dots, (X_n, Y_n)$ is a random sample from (X, Y), and let $\mathbf{X} = [X_1, \dots, X_n] \in \mathbb{R}^{p\times n}$ and $\mathbf{Y} = [Y_1, \dots, Y_n] \in \mathbb{R}^{q\times n}$. Also, let $\widehat{\Sigma}_{\mathbf{X}}$ and $\widehat{\Sigma}_{\mathbf{Y}}$ be strongly consistent estimators for Σ_X and Σ_Y, respectively. Then*

$$\mathcal{V}_n^2\big(\widehat{\Sigma}_{\mathbf{X}}^{-1/2}\mathbf{X}, \widehat{\Sigma}_{\mathbf{Y}}^{-1/2}\mathbf{Y}\big) \to \widetilde{\mathcal{V}}^2(X, Y),$$

almost surely, as $n \to \infty$. In particular, the sample affinely invariant distance correlation satisfies

$$\widetilde{\mathcal{R}}_n(\mathbf{X}, \mathbf{Y}) \to \widetilde{\mathcal{R}}(X, Y), \tag{2.5}$$

almost surely.

Proof. As the covariance matrices Σ_X and Σ_Y are positive definite, we may assume that the strongly consistent estimators $\widehat{\Sigma}_{\mathbf{X}}$ and $\widehat{\Sigma}_{\mathbf{Y}}$ also are positive definite. Therefore, in order to prove the first statement it suffices to show that

$$\mathcal{V}_n^2\big(\widehat{\Sigma}_{\mathbf{X}}^{-1/2}\mathbf{X}, \widehat{\Sigma}_{\mathbf{Y}}^{-1/2}\mathbf{Y}\big) - \mathcal{V}_n^2\big(\Sigma_X^{-1/2}\mathbf{X}, \Sigma_Y^{-1/2}\mathbf{Y}\big) \to 0, \tag{2.6}$$

almost surely. By the decomposition of Székely *et al.* [23], page 2776, equation (2.18), the left-hand side of (2.6) can be written as an average of terms of the form

$$\big|\widehat{\Sigma}_{\mathbf{X}}^{-1/2}(X_k - X_l)\big|_p\big|\widehat{\Sigma}_{\mathbf{Y}}^{-1/2}(Y_k - Y_m)\big|_q - \big|\Sigma_X^{-1/2}(X_k - X_l)\big|_p\big|\Sigma_Y^{-1/2}(Y_k - Y_m)\big|_q.$$

Using the identity

$$\begin{aligned}
&\big|\widehat{\Sigma}_{\mathbf{X}}^{-1/2}(X_k - X_l)\big|_p\big|\widehat{\Sigma}_{\mathbf{Y}}^{-1/2}(Y_k - Y_m)\big|_q \\
&\quad = \big|\big(\widehat{\Sigma}_{\mathbf{X}}^{-1/2} - \Sigma_X^{-1/2} + \Sigma_X^{-1/2}\big)(X_k - X_l)\big|_p\big|\big(\widehat{\Sigma}_{\mathbf{Y}}^{-1/2} - \Sigma_Y^{-1/2} + \Sigma_Y^{-1/2}\big)(Y_k - Y_m)\big|_q,
\end{aligned}$$

we obtain

$$\begin{aligned}
&\big|\widehat{\Sigma}_{\mathbf{X}}^{-1/2}(X_k - X_l)\big|_p\big|\widehat{\Sigma}_{\mathbf{Y}}^{-1/2}(Y_k - Y_m)\big|_q - \big|\Sigma_X^{-1/2}(X_k - X_l)\big|_p\big|\Sigma_Y^{-1/2}(Y_k - Y_m)\big|_q \\
&\quad\le \big\|\widehat{\Sigma}_{\mathbf{X}}^{-1/2} - \Sigma_X^{-1/2}\big\|\big\|\widehat{\Sigma}_{\mathbf{Y}}^{-1/2} - \Sigma_Y^{-1/2}\big\||X_k - X_l|_p|Y_k - Y_m|_q \\
&\qquad + \big\|\widehat{\Sigma}_{\mathbf{X}}^{-1/2} - \Sigma_X^{-1/2}\big\||X_k - X_l|_p\big|\Sigma_Y^{-1/2}(Y_k - Y_m)\big|_q \\
&\qquad + \big\|\widehat{\Sigma}_{\mathbf{Y}}^{-1/2} - \Sigma_Y^{-1/2}\big\|\big|\Sigma_X^{-1/2}(X_k - X_l)\big|_p|Y_k - Y_m|_q,
\end{aligned}$$

where the matrix norm $\|\Lambda\|$ is the largest eigenvalue of Λ in absolute value. Now we can separate the three sums in the decomposition of Székely *et al.* [23], page 2776, equation (2.18) and place the factors like $\|\widehat{\Sigma}_{\mathbf{X}}^{-1/2} - \Sigma_X^{-1/2}\|$ in front of the sums, since they appear in every summand. Then, $\|\widehat{\Sigma}_{\mathbf{X}}^{-1/2} - \Sigma_X^{-1/2}\|$ and $\|\widehat{\Sigma}_{\mathbf{Y}}^{-1/2} - \Sigma_Y^{-1/2}\|$ tend to zero and the remaining averages converge to constants (representing some distance correlation components) almost surely as $n \to \infty$, and this completes the proof of the first statement. Finally, the property (2.5) of strong consistency of $\widetilde{\mathcal{R}}_n(\mathbf{X}, \mathbf{Y})$ is obtained immediately upon setting $\widehat{\Sigma}_{\mathbf{X}} = S_{\mathbf{X}}$ and $\widehat{\Sigma}_{\mathbf{Y}} = S_{\mathbf{Y}}$. □

Székely *et al.* [23], page 2783, proposed a test for independence that is based on the sample distance correlation. From their results, we see that the asymptotic properties of the test statistic are not affected by the transition from the standard distance correlation to the affinely invariant distance correlation. Hence, a completely analogous but different test can be stated in terms of the affinely invariant distance correlation. Noting the results of Kosorok [11], Section 4; [12], we raise the possibility that the specific details can be devised in a judicious, data-dependent way so that the power of the test for independence increases when the transition is made to the affinely invariant distance correlation. Alternative multivariate tests for independence based on distances have recently been proposed by Heller *et al.* [7] and Székely and Rizzo [22].

3. The affinely invariant distance correlation for multivariate normal populations

We now consider the problem of calculating the affinely invariant distance correlation between the random vectors X and Y where $(X, Y) \sim \mathcal{N}_{p+q}(\mu, \Sigma)$, a multivariate normal distribution with mean vector $\mu \in \mathbb{R}^{p+q}$, covariance matrix $\Sigma \in \mathbb{R}^{(p+q)\times(p+q)}$, where X and Y have nonsingular marginal covariance matrices $\Sigma_X \in \mathbb{R}^{p\times p}$ and $\Sigma_Y \in \mathbb{R}^{q\times q}$, respectively.

For the case in which $p = q = 1$, that is, the bivariate normal distribution, the problem was solved by Székely *et al.* [23]. In that case, the formula for the affinely invariant distance correlation depends only on ρ, the correlation coefficient, and appears in terms of the functions $\sin^{-1}\rho$ and $(1-\rho^2)^{1/2}$, both of which are well-known to be special cases of Gauss' hypergeometric series. Therefore, it is natural to expect that the general case will involve generalizations of Gauss' hypergeometric series, and Theorem 3.1 below demonstrates that such is indeed the case. To formulate this result, we need to recall the rudiments of the theory of zonal polynomials (Muirhead [13], Chapter 7).

A *partition* κ is a vector of nonnegative integers $(k_1, \ldots, k_q)$ such that $k_1 \geq \cdots \geq k_q$. The integer $|\kappa| = k_1 + \cdots + k_q$ is called the *weight* of κ; and $\ell(\kappa)$, the *length* of κ, is the largest integer j such that $k_j > 0$. The *zonal polynomial* $C_\kappa(\Lambda)$ is a polynomial mapping from the class of symmetric matrices $\Lambda \in \mathbb{R}^{q \times q}$ to the real line which satisfies several properties, the following of which are crucial for our results:

(a) Let $O(q)$ denote the group of orthogonal matrices in $\mathbb{R}^{q \times q}$. Then

$$C_\kappa\big(K' \Lambda K\big) = C_\kappa(\Lambda) \tag{3.1}$$

for all $K \in O(q)$; thus, $C_\kappa(\Lambda)$ is a symmetric function of the eigenvalues of Λ.

(b) The polynomial $C_\kappa(\Lambda)$ is homogeneous of degree $|\kappa|$ in Λ: For any $\delta \in \mathbb{R}$,

$$C_\kappa(\delta\Lambda) = \delta^{|\kappa|} C_\kappa(\Lambda). \tag{3.2}$$

(c) If Λ is of rank r, then $C_\kappa(\Lambda) = 0$ whenever $\ell(\kappa) > r$.

(d) For any nonnegative integer k,

$$\sum_{|\kappa|=k} C_\kappa(\Lambda) = (\operatorname{tr} \Lambda)^k. \tag{3.3}$$

(e) For any symmetric matrices $\Lambda_1, \Lambda_2 \in \mathbb{R}^{q \times q}$,

$$\int_{O(q)} C_\kappa\big(K' \Lambda_1 K \Lambda_2\big)\, \mathrm{d}K = \frac{C_\kappa(\Lambda_1) C_\kappa(\Lambda_2)}{C_\kappa(I_q)}, \tag{3.4}$$

where $I_q = \operatorname{diag}(1, \ldots, 1) \in \mathbb{R}^{q \times q}$ denotes the identity matrix and the integral is with respect to the Haar measure on $O(q)$, normalized to have total volume 1.

(f) Let $\lambda_1, \ldots, \lambda_q$ be the eigenvalues of Λ. Then, for a partition (k) with one part,

$$C_{(k)}(\Lambda) = \frac{k!}{(1/2)_k} \sum_{i_1 + \cdots + i_q = k} \prod_{j=1}^{q} \frac{(1/2)_{i_j} \lambda_j^{i_j}}{i_j!}, \tag{3.5}$$

where the sum is over all nonnegative integers $i_1, \ldots, i_q$ such that $i_1 + \cdots + i_q = k$, and

$$(\alpha)_k = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)} = \alpha(\alpha + 1)(\alpha + 2) \cdots (\alpha + k - 1),$$

$\alpha \in \mathbb{C}$, is standard notation for the rising factorial. In particular, on setting $\lambda_j = 1$, $j = 1, \ldots, q$, we obtain from (3.5)

$$C_{(k)}(I_q) = \frac{((1/2)q)_k}{(1/2)_k} \tag{3.6}$$

(Muirhead [13], page 237, equation (18), Gross and Richards [6], page 807, Lemma 6.8).

With these properties of the zonal polynomials, we are ready to state our key result which obtains an explicit formula for the affinely invariant distance covariance in the case of a Gaussian population of arbitrary dimension and arbitrary covariance matrix with positive definite marginal covariance matrices. This formula turns out to be a function depending only on the dimensions p and q and the eigenvalues of the matrix $\Lambda = \Sigma_Y^{-1/2}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1/2}$, that is, the squared canonical correlation coefficients of the subvectors X and Y. For fixed dimensions this implies $\widetilde{\mathcal{R}}(X, Y) = g(\lambda_1, \ldots, \lambda_r)$, where $r = \min(p, q)$ and $\lambda_1, \ldots, \lambda_r$ are the canonical correlation coefficients of X and Y. Due to the functional invariance, the maximum likelihood estimator (MLE) for the affinely invariant distance correlation in the Gaussian setting is hence defined by $g(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r)$, where $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r$ are the MLEs of the canonical correlation coefficients.

Theorem 3.1. *Suppose that* $(X, Y) \sim \mathcal{N}_{p+q}(\mu, \Sigma)$, *where*

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

with $\Sigma_X \in \mathbb{R}^{p\times p}$, $\Sigma_Y \in \mathbb{R}^{q\times q}$, *and* $\Sigma_{XY} \in \mathbb{R}^{p\times q}$. *Then*

$$\widetilde{\mathcal{V}}^2(X, Y) = 4\pi \frac{c_{p-1}}{c_p}\frac{c_{q-1}}{c_q}\sum_{k=1}^{\infty}\frac{2^{2k}-2}{k!2^{2k}}\frac{(1/2)_k(-1/2)_k(-1/2)_k}{((1/2)p)_k((1/2)q)_k}C_{(k)}(\Lambda), \tag{3.7}$$

where

$$\Lambda = \Sigma_Y^{-1/2}\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY}\Sigma_Y^{-1/2} \in \mathbb{R}^{q\times q}. \tag{3.8}$$

Proof. We may assume, with no loss of generality, that μ is the zero vector. Since Σ_X and Σ_Y both are positive definite the inverse square-roots, $\Sigma_X^{-1/2}$ and $\Sigma_Y^{-1/2}$, exist.

By considering the standardized variables $\widetilde{X} = \Sigma_X^{-1/2}X$ and $\widetilde{Y} = \Sigma_Y^{-1/2}Y$, we may replace the covariance matrix Σ by

$$\widetilde{\Sigma} = \begin{pmatrix} I_p & \Lambda_{XY} \\ \Lambda_{XY}' & I_q \end{pmatrix},$$

where

$$\Lambda_{XY} = \Sigma_X^{-1/2}\Sigma_{XY}\Sigma_Y^{-1/2}. \tag{3.9}$$

Once we have made these reductions, it follows that the matrix Λ in (3.8) can be written as $\Lambda = \Lambda_{XY}'\Lambda_{XY}$ and that it has norm less than or equal to 1. Indeed, by the *partial Iwasawa decomposition* of $\widetilde{\Sigma}$, viz., the identity,

$$\widetilde{\Sigma} = \begin{pmatrix} I_p & 0 \\ \Lambda_{XY}' & I_q \end{pmatrix}\begin{pmatrix} I_p & 0 \\ 0 & I_q - \Lambda_{XY}'\Lambda_{XY} \end{pmatrix}\begin{pmatrix} I_p & \Lambda_{XY} \\ 0 & I_q \end{pmatrix},$$

where the zero matrix of any dimension is denoted by 0, we see that the matrix $\widetilde{\Sigma}$ is positive semidefinite if and only if $I_q - \Lambda$ is positive semidefinite. Hence, $\Lambda \le I_q$ in the Loewner ordering and therefore $\|\Lambda\| \le 1$.

We proceed to calculate the distance covariance $\widetilde{\mathcal{V}}(X,Y)=\mathcal{V}(\widetilde{X},\widetilde{Y})$. It is well-known that the characteristic function of $(\widetilde{X},\widetilde{Y})$ is

$$f_{\widetilde{X},\widetilde{Y}}(s,t)=\exp\left[-\frac{1}{2}\begin{pmatrix}s\\t\end{pmatrix}'\widetilde{\Sigma}\begin{pmatrix}s\\t\end{pmatrix}\right]=\exp\left[-\frac{1}{2}\big(|s|_p^2+|t|_q^2+2s'\Lambda_{XY}t\big)\right],$$

where $s\in\mathbb{R}^p$ and $t\in\mathbb{R}^q$. Therefore,

$$\big|f_{\widetilde{X},\widetilde{Y}}(s,t)-f_{\widetilde{X}}(s)f_{\widetilde{Y}}(t)\big|^2=\big(1-\exp\big(-s'\Lambda_{XY}t\big)\big)^2\exp\big(-|s|_p^2-|t|_q^2\big),$$

and hence

$$\begin{aligned}c_pc_q\mathcal{V}^2(\widetilde{X},\widetilde{Y})&=\int_{\mathbb{R}^{p+q}}\big(1-\exp\big(-s'\Lambda_{XY}t\big)\big)^2\exp\big(-|s|_p^2-|t|_q^2\big)\frac{\mathrm{d}s}{|s|_p^{p+1}}\frac{\mathrm{d}t}{|t|_q^{q+1}}\\&=\int_{\mathbb{R}^{p+q}}\big(1-\exp\big(s'\Lambda_{XY}t\big)\big)^2\exp\big(-|s|_p^2-|t|_q^2\big)\frac{\mathrm{d}s}{|s|_p^{p+1}}\frac{\mathrm{d}t}{|t|_q^{q+1}},\end{aligned}\tag{3.10}$$

where the latter integral is obtained by making the change of variables $s\mapsto -s$ within the former integral.

By a Taylor series expansion, we obtain

$$\begin{aligned}\big(1-\exp\big(s'\Lambda_{XY}t\big)\big)^2&=1-2\exp\big(s'\Lambda_{XY}t\big)+\exp\big(2s'\Lambda_{XY}t\big)\\&=\sum_{k=2}^{\infty}\frac{2^k-2}{k!}\big(s'\Lambda_{XY}t\big)^k.\end{aligned}$$

Substituting this series into (3.10) and interchanging summation and integration, a procedure which is straightforward to verify by means of Fubini's theorem, and noting that the odd-order terms integrate to zero, we obtain

$$c_pc_q\mathcal{V}^2(\widetilde{X},\widetilde{Y})=\sum_{k=1}^{\infty}\frac{2^{2k}-2}{(2k)!}\int_{\mathbb{R}^{p+q}}\big(s'\Lambda_{XY}t\big)^{2k}\exp\big(-|s|_p^2-|t|_q^2\big)\frac{\mathrm{d}s}{|s|_p^{p+1}}\frac{\mathrm{d}t}{|t|_q^{q+1}}.\tag{3.11}$$

To calculate, for $k\ge 1$, the integral

$$\int_{\mathbb{R}^{p+q}}\big(s'\Lambda_{XY}t\big)^{2k}\exp\big(-|s|_p^2-|t|_q^2\big)\frac{\mathrm{d}s}{|s|_p^{p+1}}\frac{\mathrm{d}t}{|t|_q^{q+1}},\tag{3.12}$$

we change variables to polar coordinates, putting $s=r_x\theta$ and $t=r_y\phi$ where $r_x,r_y>0$, $\theta=(\theta_1,\dots,\theta_p)'\in S^{p-1}$, and $\phi=(\phi_1,\dots,\phi_q)'\in S^{q-1}$. Then the integral (3.12) separates into a product of multiple integrals over (r_x,r_y), and over (θ,ϕ), respectively. The integrals over r_x and r_y are standard gamma integrals,

$$\int_0^{\infty}\int_0^{\infty}r_x^{2k-2}r_y^{2k-2}\exp\big(-r_x^2-r_y^2\big)\,\mathrm{d}r_x\,\mathrm{d}r_y=\tfrac{1}{4}\big[\Gamma\big(k-\tfrac{1}{2}\big)\big]^2=\big[\big(-\tfrac{1}{2}\big)_k\big]^2\pi,\tag{3.13}$$

and the remaining factor is the integral

$$\int_{S^{q-1}} \int_{S^{p-1}} \big(\theta' \Lambda_{XY} \phi\big)^{2k} \, \mathrm{d}\theta \, \mathrm{d}\phi, \tag{3.14}$$

where $\mathrm{d}\theta$ and $\mathrm{d}\phi$ are unnormalized surface measures on S^{p-1} and S^{q-1}, respectively. By a standard invariance argument,

$$\int_{S^{p-1}} \big(\theta' v\big)^{2k} \, \mathrm{d}\theta = |v|_p^{2k} \int_{S^{p-1}} \theta_1^{2k} \, \mathrm{d}\theta,$$

$v \in \mathbb{R}^p$. Setting $v = \Lambda_{XY}\phi$ and applying some well-known properties of the surface measure $\mathrm{d}\theta$, we obtain

$$\begin{aligned} \int_{S^{p-1}} \big(\theta' \Lambda_{XY} \phi\big)^{2k} \, \mathrm{d}\theta &= |\Lambda_{XY}\phi|_p^{2k} \int_{S^{p-1}} \theta_1^{2k} \, \mathrm{d}\theta \\ &= 2c_{p-1} \frac{\Gamma(k+1/2)\Gamma(1/2p)}{\Gamma(k+(1/2)p)\Gamma(1/2)} \big(\phi' \Lambda \phi\big)^k. \end{aligned}$$

Therefore, in order to evaluate (3.14), it remains to evaluate

$$J_k(\Lambda) = \int_{S^{q-1}} \big(\phi' \Lambda \phi\big)^k \, \mathrm{d}\phi.$$

Since the surface measure is invariant under transformation $\phi \mapsto K\phi$, $K \in O(q)$, it follows that $J_k(\Lambda) = J_k(K'\Lambda K)$ for all $K \in O(q)$. Integrating with respect to the normalized Haar measure on the orthogonal group, we conclude that

$$J_k(\Lambda) = \int_{O(q)} J_k\big(K'\Lambda K\big) \, \mathrm{d}K = \int_{S^{q-1}} \int_{O(q)} \big(\phi' K' \Lambda K \phi\big)^k \, \mathrm{d}K \, \mathrm{d}\phi. \tag{3.15}$$

We now use the properties of the zonal polynomials. By (3.3),

$$\big(\phi' K' \Lambda K \phi\big)^k = \big(\operatorname{tr} K' \Lambda K \phi\phi'\big)^k = \sum_{|\kappa|=k} C_\kappa\big(K' \Lambda K \phi\phi'\big);$$

therefore, by (3.4),

$$\int_{O(q)} \big(\phi' K' \Lambda K \phi\big)^k \, \mathrm{d}K = \sum_{|\kappa|=k} \int_{O(q)} C_\kappa\big(K' \Lambda K \phi\phi'\big) \, \mathrm{d}K = \sum_{|\kappa|=k} \frac{C_\kappa(\Lambda) C_\kappa(\phi\phi')}{C_\kappa(I_q)}.$$

Since $\phi\phi'$ is of rank 1 then, by property (c), $C_\kappa(\phi\phi') = 0$ if $\ell(\kappa) > 1$; it now follows, by (3.3) and the fact that $\phi \in S^{q-1}$, that

$$C_{(k)}\big(\phi\phi'\big) = \sum_{|\kappa|=k} C_\kappa\big(\phi\phi'\big) = \big(\operatorname{tr} \phi\phi'\big)^k = \big(\phi'\phi\big)^k = |\phi|_q^{2k} = 1.$$

Therefore,

$$\int_{O(q)} \big(\phi' K' \Lambda K \phi\big)^k \, \mathrm{d}K = \frac{C_{(k)}(\Lambda)}{C_{(k)}(I_q)} = \frac{(1/2)_k}{((1/2)q)_k} C_{(k)}(\Lambda),$$

where the last equality follows by (3.6). Substituting this result at (3.15), we obtain

$$J_k(\Lambda) = 2c_{q-1} \frac{(1/2)_k}{((1/2)q)_k} C_{(k)}(\Lambda).$$

Collecting together these results, and using the well-known identity $(2k)! = k! 2^{2k} (1/2)_k$, we obtain the representation (3.7), as desired. □

We remark that by interchanging the roles of X and Y in Theorem 3.1, we would obtain (3.7) with Λ in (3.8) replaced by

$$\Lambda_0 = \Sigma_X^{-1/2} \Sigma_{XY} \Sigma_Y^{-1} \Sigma_{YX} \Sigma_X^{-1/2} \in \mathbb{R}^{p \times p}.$$

Since Λ and Λ_0 have the same characteristic polynomial and hence the same set of nonzero eigenvalues, and noting that $C_\kappa(\Lambda)$ depends only on the eigenvalues of Λ, it follows that $C_{(k)}(\Lambda) = C_{(k)}(\Lambda_0)$. Therefore, the series representation (3.7) for $\widetilde{\mathcal{V}}^2(X, Y)$ remains unchanged if the roles of X and Y are interchanged.

The series appearing in Theorem 3.1 can be expressed in terms of the generalized hypergeometric functions of matrix argument (Gross and Richards [6], James [9], Muirhead [13]). For this purpose, we introduce the *partitional rising factorial* for any $\alpha \in \mathbb{C}$ and any partition $\kappa = (k_1, \dots, k_q)$ as

$$(\alpha)_\kappa = \prod_{j=1}^{q} \big(\alpha - (1/2)(j-1)\big)_{k_j}.$$

Let $\alpha_1, \dots, \alpha_l, \beta_1, \dots, \beta_m \in \mathbb{C}$ where $-\beta_i + \frac{1}{2}(j-1)$ is not a nonnegative integer, for all $i = 1, \dots, m$ and $j = 1, \dots, q$. Then the ${}_lF_m$ generalized hypergeometric function of matrix argument is defined as

$${}_lF_m(\alpha_1, \dots, \alpha_l; \beta_1, \dots, \beta_m; S) = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{|\kappa|=k} \frac{(\alpha_1)_\kappa \cdots (\alpha_l)_\kappa}{(\beta_1)_\kappa \cdots (\beta_m)_\kappa} C_\kappa(S),$$

where S is a symmetric matrix. A complete analysis of the convergence properties of this series was derived by Gross and Richards [6], page 804, Theorem 6.3, and we refer the reader to that paper for the details.

Corollary 3.2. *In the setting of Theorem 3.1, we have*

$$\begin{aligned}\widetilde{\mathcal{V}}^2(X, Y) = 4\pi \frac{c_{p-1}}{c_p} \frac{c_{q-1}}{c_q} \bigg(& {}_3F_2\bigg(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p, \frac{1}{2}q; \Lambda\bigg) \\ & - 2\,{}_3F_2\bigg(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p, \frac{1}{2}q; \frac{1}{4}\Lambda\bigg) + 1\bigg).\end{aligned} \tag{3.16}$$

Proof. It is evident that

$$(1/2)_\kappa = \begin{cases} (1/2)_{k_1}, & \text{if } \ell(\kappa) \le 1, \\ 0, & \text{if } \ell(\kappa) > 1. \end{cases}$$

Therefore, we now can write the series in (3.7), up to a multiplicative constant, in terms of a generalized hypergeometric function of matrix argument, in that

$$\begin{aligned}
&\sum_{k=1}^{\infty} \frac{2^{2k}-2}{k!2^{2k}} \frac{(1/2)_k(-1/2)_k(-1/2)_k}{((1/2)p)_k((1/2)q)_k} C_{(k)}(\Lambda) \\
&\quad = \sum_{k=1}^{\infty} \frac{2^{2k}-2}{k!2^{2k}} \sum_{|\kappa|=k} \frac{(1/2)_\kappa(-1/2)_\kappa(-1/2)_\kappa}{((1/2)p)_\kappa((1/2)q)_\kappa} C_\kappa(\Lambda) \\
&\quad = \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{|\kappa|=k} \frac{(1/2)_\kappa(-1/2)_\kappa(-1/2)_\kappa}{((1/2)p)_\kappa((1/2)q)_\kappa} C_\kappa(\Lambda) \\
&\qquad - 2\sum_{k=1}^{\infty} \frac{1}{k!2^{2k}} \sum_{|\kappa|=k} \frac{(1/2)_\kappa(-1/2)_\kappa(-1/2)_\kappa}{((1/2)p)_\kappa((1/2)q)_\kappa} C_\kappa(\Lambda) \\
&\quad = \left[{}_3F_2\left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p, \frac{1}{2}q; \Lambda\right) - 1\right] - 2\left[{}_3F_2\left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p, \frac{1}{2}q; \frac{1}{4}\Lambda\right) - 1\right].
\end{aligned}$$

Due to property (3.2) it remains to show that the zonal polynomial series expansion for the ${}_3F_2(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p, \frac{1}{2}q; \Lambda)$ generalized hypergeometric function of matrix argument converges absolutely for all Λ with $\Lambda \le I_q$ in the Loewner ordering. By (3.6)

$$\begin{aligned}
{}_3F_2\left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p, \frac{1}{2}q; \Lambda\right) &\le \sum_{k=0}^{\infty} \frac{2^{2k}}{k!2^{2k}} \frac{(-1/2)_k(-1/2)_k}{((1/2)p)_k} \\
&= {}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p; 1\right).
\end{aligned}$$

The latter series converges due to Gauss' theorem for hypergeometric functions and so we have absolute convergence at (3.16) for all Σ with positive definite marginal covariance matrices. □

Consider the case in which $q = 1$ and p is arbitrary. Then Λ is a scalar; say, $\Lambda = \rho^2$ for some $\rho \in [-1, 1]$. Then the ${}_3F_2$ generalized hypergeometric functions in (3.16) each reduce to a Gaussian hypergeometric function, denoted by ${}_2F_1$, and (3.16) becomes

$$\widetilde{\mathcal{V}}^2(X, Y) = 4\frac{c_{p-1}}{c_p}\left({}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p; \rho^2\right) - 2{}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p; \frac{1}{4}\rho^2\right) + 1\right).$$

For the case in which $p = q = 1$, we may identify ρ with the Pearson correlation coefficient and the hypergeometric series can be expressed in terms of elementary functions. By well-known results (Andrews, Askey and Roy [1], pages 64 and 94),

$$ {}_2F_1\bigl(-\tfrac{1}{2},-\tfrac{1}{2};\tfrac{1}{2};\rho^2\bigr)=\rho\sin^{-1}\rho+\bigl(1-\rho^2\bigr)^{1/2}, \tag{3.17} $$

and thus we derive the same result for $p=q=1$ as in Székely *et al.* [23], page 2785.

For cases in which $q=1$ and p is odd, we can again obtain explicit expressions for $\widetilde{\mathcal{V}}^2(X,Y)$. In such cases, the ${}_3F_2$ generalized hypergeometric functions in (3.16) reduce to Gaussian hypergeometric functions of the form ${}_2F_1(-\frac{1}{2},-\frac{1}{2};k+\frac{1}{2};\rho^2)$, $k\in\mathbb{N}$, and it can be shown that these latter functions are expressible in closed form in terms of elementary functions and the $\sin^{-1}(\cdot)$ function. For instance, for $p=3$, the contiguous relations for the ${}_2F_1$ functions can be used to show that

$$ {}_2F_1\left(-\frac{1}{2},-\frac{1}{2};\frac{3}{2};\rho^2\right)=\frac{3(1-\rho^2)^{1/2}}{4}+\frac{(1+2\rho^2)\sin^{-1}\rho}{4\rho}. \tag{3.18} $$

Further, by repeated application of the same contiguous relations, it can be shown that for $k=2,3,4,\ldots,$

$$ {}_2F_1\bigl(-\tfrac{1}{2},-\tfrac{1}{2};k+\tfrac{1}{2};\rho^2\bigr)=\rho^{-2(k-1)}\bigl(1-\rho^2\bigr)^{1/2}P_{k-1}\bigl(\rho^2\bigr)+\rho^{-(2k-1)}Q_k\bigl(\rho^2\bigr)\sin^{-1}\rho, $$

where P_k and Q_k are polynomials of degree k. Therefore, for $q=1$ and p odd, the distance covariance $\widetilde{\mathcal{V}}^2(X,Y)$ can be expressed in closed form in terms of elementary functions and the $\sin^{-1}(\cdot)$ function.

The appearance of the generalized hypergeometric functions of matrix argument also yields a useful expression for the affinely invariant distance variance. In order to state this result, we shall define for each positive integer p the quantity

$$ A(p)=\frac{\Gamma((1/2)p)\Gamma((1/2)p+1)}{[\Gamma((1/2)(p+1))]^2}-2\,{}_2F_1\left(-\frac{1}{2},-\frac{1}{2};\frac{1}{2}p;\frac{1}{4}\right)+1. \tag{3.19} $$

Corollary 3.3. *In the setting of Theorem* 3.1, *we have*

$$ \widetilde{\mathcal{V}}^2(X,X)=4\pi\frac{c_{p-1}^2}{c_p^2}A(p). \tag{3.20} $$

Proof. We are in the special case of Theorem 3.1 for which $X=Y$, so that $p=q$ and $\Lambda=I_p$. By applying (3.6), we can write the series in (3.7) as

$$ \begin{aligned} &4\pi\frac{c_{p-1}^2}{c_p^2}\sum_{k=1}^{\infty}\frac{2^{2k}-2}{k!2^{2k}}\frac{(1/2)_k(-1/2)_k(-1/2)_k}{((1/2)p)_k((1/2)p)_k}C_{(k)}(I_p)\\ &\quad=4\pi\frac{c_{p-1}^2}{c_p^2}\sum_{k=1}^{\infty}\frac{2^{2k}-2}{k!2^{2k}}\frac{(-1/2)_k(-1/2)_k}{((1/2)p)_k} \end{aligned} $$

$$= 4\pi \frac{c_{p-1}^2}{c_p^2} \left(\left[{}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p; 1\right) - 1 \right] - 2\left[{}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p; \frac{1}{4}\right) - 1 \right] \right).$$

By Gauss' theorem for hypergeometric functions the series ${}_2F_1(-\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p; z)$ also converges for the special value $z = 1$, and then

$${}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}p; 1\right) = \frac{\Gamma((1/2)p)\Gamma((1/2)p+1)}{[\Gamma((1/2)(p+1))]^2},$$

thereby completing the proof. □

For cases in which p is odd, we can proceed as explained at (3.18) to obtain explicit values for the Gaussian hypergeometric function remaining in (3.20). This leads in such cases to explicit expressions for the exact value of $\widetilde{\mathcal{V}}^2(X, X)$. In particular, if $p = 1$ then it follows from (1.2) and (3.17) that

$$\widetilde{\mathcal{V}}^2(X, X) = \frac{4}{3} - \frac{4(\sqrt{3}-1)}{\pi};$$

and for $p = 3$, we deduce from (1.2) and (3.18) that

$$\widetilde{\mathcal{V}}^2(X, X) = 2 - \frac{4(3\sqrt{3}-4)}{\pi}.$$

Corollaries 3.2 and 3.3 enable the explicit and efficient calculation of the affinely invariant distance correlation (1.6) in the case of subvectors of a multivariate normal population. In doing so, we use the algorithm of Koev and Edelman [10] to evaluate the generalized hypergeometric function of matrix argument, with C and Matlab code being available at these authors' websites.

Figure 1 concerns the case $p = q = 2$ in various settings, in which the matrix Λ_{22} depends on a single parameter r only. The dotted line shows the affinely invariant distance correlation when

$$\Lambda_{XY} = \begin{pmatrix} 0 & 0 \\ 0 & r \end{pmatrix};$$

this is the case with the weakest dependence considered here. The dash-dotted line applies when

$$\Lambda_{XY} = \begin{pmatrix} r & 0 \\ 0 & r \end{pmatrix}.$$

The strongest dependence corresponds to the dashed line, which shows the affinely invariant distance correlation when

$$\Lambda_{XY} = \begin{pmatrix} r & r \\ r & r \end{pmatrix};$$

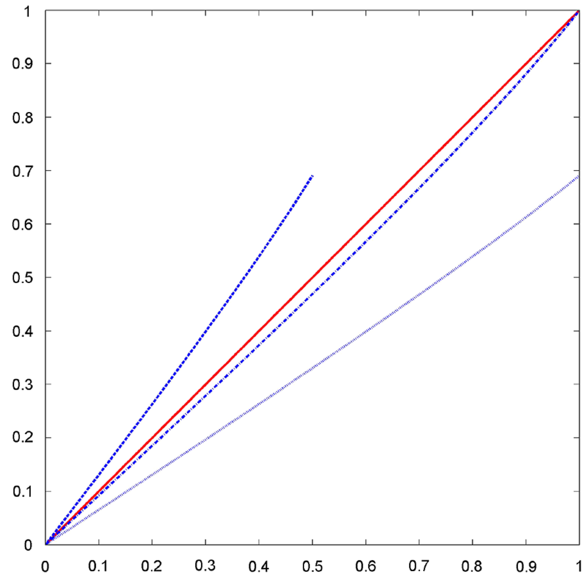

Figure 1. The affinely invariant distance correlation for subvectors of a multivariate normal population, where $p = q = 2$, as a function of the parameter r in three distinct settings. The solid diagonal line is the identity function and is provided to serve as a reference for the three distance correlation functions. See the text for details.

in this case we need to assume that $0 \le r \le \frac{1}{2}$ in order to retain positive definiteness.

In Figure 2, panel (a) shows the affinely invariant distance correlation when $p = q = 2$ and

$$\Lambda_{XY} = \begin{pmatrix} r & 0 \\ 0 & s \end{pmatrix},$$

where $0 \le r, s \le 1$. With reference to Figure 1, the margins correspond to the dotted line and the diagonal corresponds to the dash-dotted line.

Panel (b) of Figure 2 concerns the case in which $p = 2$, $q = 1$ and $\Lambda_{XY} = (r, s)'$, where $r^2 + s^2 \le 1$. Here, the affinely invariant distance correlation attains an upper limit as $r^2 + s^2 \uparrow 1$, and we have evaluated that limit numerically as 0.8252.

4. Limit theorems

We now study the limiting behavior of the affinely invariant distance correlation measures for subvectors of multivariate normal populations.

Our first result quantifies the asymptotic decay of the affinely invariant distance correlation in the case in which the cross-covariance matrix converges to the zero matrix, in that

$$\operatorname{tr}(\Lambda) = \|\Lambda_{XY}\|_F^2 \longrightarrow 0,$$

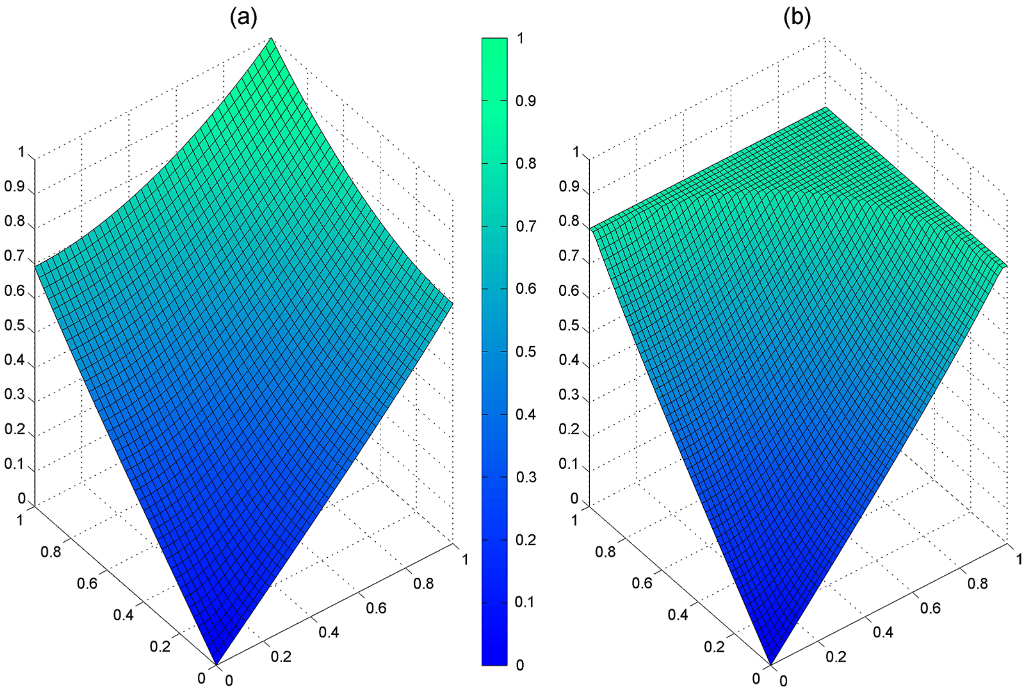

Figure 2. The affinely invariant distance correlation between the p- and q-dimensional subvectors of a $(p+q)$-dimensional multivariate normal population, where (a) $p = q = 2$ and $\Lambda_{XY} = \mathrm{diag}(r, s)$, and (b) $p = 2$, $q = 1$ and $\Lambda_{XY} = (r, s)'$.

where $\|\cdot\|_F$ denotes the Frobenius norm, and the matrices $\Lambda = \Lambda_{XY}{}'\Lambda_{XY}$ and Λ_{XY} are defined in (3.8) and (3.9), respectively.

Theorem 4.1. *Suppose that* $(X, Y) \sim \mathcal{N}_{p+q}(\mu, \Sigma)$, *where*

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

with $\Sigma_X \in \mathbb{R}^{p\times p}$ *and* $\Sigma_Y \in \mathbb{R}^{q\times q}$ *being positive definite, and suppose that the matrix* Λ *in* (3.8) *has positive trace. Then,*

$$\lim_{\mathrm{tr}(\Lambda)\to 0} \frac{\widetilde{\mathcal{R}}^2(X, Y)}{\mathrm{tr}(\Lambda)} = \frac{1}{4pq\sqrt{A(p)A(q)}}, \tag{4.1}$$

where $A(p)$ *is defined in* (3.19).

Proof. We first note that $\widetilde{\mathcal{V}}^2(X, X)$ and $\widetilde{\mathcal{V}}^2(Y, Y)$ do not depend on Σ_{XY}, as can be seen from their explicit representations in terms of $A(p)$ and $A(q)$ given in (3.20).

In studying the asymptotic behavior of $\widetilde{\mathcal{V}}^2(X,Y)$, we may interchange the limit and the summation in the series representation (3.7). Hence, it suffices to find the limit term-by-term. Since $C_{(1)}(\Lambda) = \operatorname{tr}(\Lambda)$ then the ratio of the term for $k = 1$ and $\operatorname{tr}(\Lambda)$ equals

$$\frac{c_{p-1}}{c_p}\frac{c_{q-1}}{c_q}\frac{\pi}{pq}.$$

For $k \geq 2$, it follows from (3.5) that $C_{(k)}(\Lambda)$ is a sum of monomials in the eigenvalues of Λ, with each monomial being of degree k, which is greater than the degree, viz. 1, of $\operatorname{tr}(\Lambda)$; therefore,

$$\lim_{\operatorname{tr}(\Lambda)\to 0}\frac{C_{(k)}(\Lambda)}{\operatorname{tr}(\Lambda)} = \lim_{\Lambda\to 0}\frac{C_{(k)}(\Lambda)}{\operatorname{tr}(\Lambda)} = 0.$$

Collecting these facts together, we obtain (4.1). $\square$

If $p = q = 1$, we are in the situation of Theorem 7(iii) in Székely *et al.* [23]. Applying the identity (3.17), we obtain

$${}_2F_1\left(-\frac{1}{2}, -\frac{1}{2}; \frac{1}{2}; \frac{1}{4}\right) = \frac{\pi}{12} + \frac{\sqrt{3}}{2},$$

and $(\operatorname{tr}(\Lambda))^{1/2} = |\rho|$. Thus, we obtain

$$\lim_{\rho\to 0}\frac{\widetilde{\mathcal{R}}(X,Y)}{|\rho|} = \frac{1}{2(1+(1/3)\pi - \sqrt{3})^{1/2}},$$

as shown by Székely *et al.* [23], page 2785.

In the remainder of this section, we consider situations in which one or both of the dimensions p and q grow without bound. We will repeatedly make use of the fact that, with c_p defined as in (1.2),

$$\frac{c_{p-1}}{\sqrt{p}c_p} \longrightarrow \frac{1}{\sqrt{2\pi}} \tag{4.2}$$

as $p \to \infty$, which follows easily from the functional equation for the gamma function along with Stirling's formula.

Theorem 4.2. *For each positive integer p, suppose that $(X_p, Y_p) \sim \mathcal{N}_{2p}(\mu_p, \Sigma_p)$, where*

$$\Sigma_p = \begin{pmatrix} \Sigma_{X,p} & \Sigma_{XY,p} \\ \Sigma_{YX,p} & \Sigma_{Y,p} \end{pmatrix}$$

with $\Sigma_{X,p} \in \mathbb{R}^{p\times p}$ and $\Sigma_{Y,p} \in \mathbb{R}^{p\times p}$ being positive definite and such that

$$\Lambda_p = \Sigma_{Y,p}^{-1/2}\Sigma_{YX,p}\Sigma_{X,p}^{-1}\Sigma_{XY,p}\Sigma_{Y,p}^{-1/2} \neq 0.$$

Then

$$\lim_{p\to\infty} \frac{p}{\operatorname{tr}(\Lambda_p)} \widetilde{\mathcal{V}}^2(X_p, Y_p) = \frac{1}{2} \tag{4.3}$$

and

$$\lim_{p\to\infty} \frac{p}{\operatorname{tr}(\Lambda_p)} \widetilde{\mathcal{R}}^2(X_p, Y_p) = 1. \tag{4.4}$$

In particular, if $\Lambda_p = r^2 I_p$ for some $r \in [0, 1]$ then $\operatorname{tr}(\Lambda_p) = r^2 p$, and so (4.3) and (4.4) reduce to

$$\lim_{p\to\infty} \widetilde{\mathcal{V}}^2(X_p, Y_p) = \tfrac{1}{2} r^2 \quad \text{and} \quad \lim_{p\to\infty} \widetilde{\mathcal{R}}(X_p, Y_p) = r,$$

respectively. The following corollary concerns the special case in which $r = 1$; we state it separately for emphasis.

Corollary 4.3. *For each positive integer p, suppose that $X_p \sim \mathcal{N}_p(\mu_p, \Sigma_p)$, with Σ_p being positive definite. Then*

$$\lim_{p\to\infty} \widetilde{\mathcal{V}}^2(X_p, X_p) = \tfrac{1}{2}. \tag{4.5}$$

Proof of Theorem 4.2 and Corollary 4.3. In order to prove (4.3), we study the limit for the terms corresponding separately to $k = 1$, $k = 2$, and $k \ge 3$ in (3.7).

For $k = 1$, on recalling that $C_{(1)}(\Lambda_p) = \operatorname{tr}(\Lambda_p)$, it follows from (4.2) that the ratio of that term to $\operatorname{tr}(\Lambda_p)/p$ tends to $1/2$.

For $k = 2$, we first deduce from (3.3) that $C_{(2)}(\Lambda_p) \le (\operatorname{tr}\Lambda_p)^2$. Moreover, $\operatorname{tr}(\Lambda_p) \le p$ because $\Lambda_p \le I_p$ in the Loewner ordering. Thus, the ratio of the second term in (3.7) to $\operatorname{tr}(\Lambda_p)/p$ is a constant multiple of

$$\begin{aligned} \frac{p}{\operatorname{tr}(\Lambda_p)} \frac{c_{p-1}^2}{c_p^2} \frac{C_{(2)}(\Lambda_p)}{((1/2)p)_2((1/2)p)_2} &\le \frac{c_{p-1}^2}{c_p^2} \frac{p^2}{((1/2)p)_2((1/2)p)_2} \\ &= 4\frac{p}{(p+1)^2} \frac{c_{p-1}^2}{p c_p^2} \end{aligned}$$

which, by (4.2), converges to zero as $p \to \infty$.

Finally, suppose that $k \ge 3$. Obviously, $\Lambda_p \le \|\Lambda_p\| I_p$ in the Loewner ordering inequality, and so it follows from (3.5) that $C_{(k)}(\Lambda_p) \le \|\Lambda_p\|^k C_{(k)}(I_p)$. Also, since $\operatorname{tr}(\Lambda_p) \ge \|\Lambda_p\|$ then by again applying the Loewner ordering inequality and (3.6) we obtain

$$\frac{C_{(k)}(\Lambda_p)}{\operatorname{tr}(\Lambda_p)} \le \frac{\|\Lambda_p\|^k C_{(k)}(I_p)}{\|\Lambda_p\|} = \|\Lambda_p\|^{k-1} C_{(k)}(I_p) \le C_{(k)}(I_p) = \frac{((1/2)p)_k}{(1/2)_k}. \tag{4.6}$$

Therefore,

$$\begin{aligned}
&4\pi \frac{p}{\operatorname{tr}(\Lambda_p)} \frac{c_{p-1}^2}{c_p^2} \sum_{k=3}^{\infty} \frac{2^{2k}-2}{k!2^{2k}} \frac{(1/2)_k(-1/2)_k(-1/2)_k}{((1/2)p)_k((1/2)p)_k} C_{(k)}(\Lambda_p) \\
&\quad \le 4\pi p \frac{c_{p-1}^2}{c_p^2} \sum_{k=3}^{\infty} \frac{2^{2k}-2}{k!2^{2k}} \frac{(-1/2)_k(-1/2)_k}{((1/2)p)_k}.
\end{aligned}$$

By (4.2), each term $pc_{p-1}^2/(\frac{1}{2}p)_k c_p^2$ converges to zero as $p \to \infty$, and this proves both (4.3) and its special case, (4.5). Then, (4.4) follows immediately. □

Finally, we consider the situation in which q, the dimension of Y, is fixed while p, the dimension of X, grows without bound.

Theorem 4.4. *For each positive integer p, suppose that $(X_p, Y) \sim \mathcal{N}_{p+q}(\mu_p, \Sigma_p)$, where*

$$\Sigma_p = \begin{pmatrix} \Sigma_{X,p} & \Sigma_{XY,p} \\ \Sigma_{YX,p} & \Sigma_Y \end{pmatrix}$$

with $\Sigma_{X,p} \in \mathbb{R}^{p\times p}$ and $\Sigma_Y \in \mathbb{R}^{q\times q}$ being positive definite and such that

$$\Lambda_p = \Sigma_Y^{-1/2} \Sigma_{YX,p} \Sigma_{X,p}^{-1} \Sigma_{XY,p} \Sigma_Y^{-1/2} \neq 0.$$

Then

$$\lim_{p\to\infty} \frac{\sqrt{p}}{\operatorname{tr}(\Lambda_p)} \widetilde{\mathcal{V}}^2(X_p, Y) = \sqrt{\frac{\pi}{2}} \frac{c_{q-1}}{qc_q} \tag{4.7}$$

and

$$\lim_{p\to\infty} \frac{\sqrt{p}}{\operatorname{tr}(\Lambda_p)} \widetilde{\mathcal{R}}^2(X_p, Y) = \frac{1}{2q\sqrt{A(q)}}. \tag{4.8}$$

Proof. By (3.7),

$$\widetilde{\mathcal{V}}^2(X_p, Y) = 4\pi \frac{c_{p-1}}{c_p} \frac{c_{q-1}}{c_q} \sum_{k=1}^{\infty} \frac{2^{2k}-2}{k!2^{2k}} \frac{(1/2)_k(-1/2)_k(-1/2)_k}{((1/2)p)_k((1/2)q)_k} C_{(k)}(\Lambda_p).$$

We now examine the limiting behavior, as $p \to \infty$, of the terms in this sum for $k = 1$ and, separately, for $k \ge 2$.

For $k = 1$, the limiting value of the ratio of the corresponding term to $\operatorname{tr}(\Lambda_p)/\sqrt{p}$ equals

$$\pi \frac{c_{q-1}}{qc_q} \lim_{p\to\infty} \frac{\sqrt{p}}{\operatorname{tr}(\Lambda_p)} \frac{c_{p-1}}{pc_p} C_{(1)}(\Lambda_p) = \sqrt{\frac{\pi}{2}} \frac{c_{q-1}}{qc_q}$$

by (4.2) and the fact that $C_{(1)}(\Lambda_p) = \operatorname{tr}(\Lambda_p)$.

For $k \geq 2$, the ratio of the sum to $\operatorname{tr}(\Lambda_p)/\sqrt{p}$ equals

$$\begin{aligned} &4\pi \frac{\sqrt{p}}{\operatorname{tr}(\Lambda_p)} \frac{c_{p-1}}{c_p} \frac{c_{q-1}}{c_q} \sum_{k=2}^{\infty} \frac{2^{2k}-2}{k!2^{2k}} \frac{(1/2)_k(-1/2)_k(-1/2)_k}{((1/2)p)_k((1/2)q)_k} C_{(k)}(\Lambda_p) \\ &\leq 4\pi \frac{\sqrt{p}}{\|\Lambda_p\|} \frac{c_{p-1}}{c_p} \frac{c_{q-1}}{c_q} \sum_{k=2}^{\infty} \frac{2^{2k}-2}{k!2^{2k}} \frac{(-1/2)_k(-1/2)_k}{((1/2)p)_k} \|\Lambda_p\|^k \\ &\leq 4\pi \sqrt{p} \frac{c_{p-1}}{c_p} \frac{c_{q-1}}{c_q} \sum_{k=2}^{\infty} \frac{2^{2k}-2}{k!2^{2k}} \frac{(-1/2)_k(-1/2)_k}{((1/2)p)_k}, \end{aligned}$$

where we have used (4.6) to obtain the last two inequalities. By applying (4.2), we see that the latter upper bound converges to 0 as $p \to \infty$, which proves (4.7), and then (4.8) follows immediately. □

The results in this section have practical implications for affine distance correlation analysis of large-sample, high-dimensional Gaussian data. In the setting of Theorem 4.4, $\operatorname{tr}(\Lambda_p) \leq q$ is bounded, and so

$$\lim_{p \to \infty} \widetilde{\mathcal{R}}(X_p, Y) = 0.$$

As a consequence of Theorem 2.1 on the consistency of sample measures, it follows that the direct calculation of affine distance correlation measures for such data will return values which are virtually zero. In practice, in order to obtain values of the sample affine distance correlation measures which permit statistical inference, it will be necessary to calculate $\widehat{\Lambda}_p$, the maximum likelihood estimator of Λ_p, and then to rescale the distance correlation measures with the factor $\sqrt{p}/\operatorname{tr}(\widehat{\Lambda}_p)$. In the scenario of Theorem 4.2, the asymptotic behavior of the affine distance correlation measures depends on the ratio $p/\operatorname{tr}(\Lambda_p)$; and as $\operatorname{tr}(\Lambda_p)$ can attain any value in the interval $[0, p]$, a wide range of asymptotic rates of convergence is conceivable.

In all these settings, the series representation (3.7) can be used to obtain complete asymptotic expansions in powers of p^{-1} or q^{-1}, of the affine distance covariance or correlation measures, as p or q tend to infinity.

5. Time series of wind vectors at the Stateline wind energy center

Rémillard [15] proposed the use of the distance correlation to explore nonlinear dependencies in time series data. Zhou [24] pursued this approach recently and defined the auto distance covariance function and the auto distance correlation function, along with natural sample versions, for a strongly stationary vector-valued time series, say $(X_j)_{j=-\infty}^{\infty}$.

It is straightforward to extend these notions to the affinely invariant distance correlation. Thus, for an integer k, we refer to

$$\widetilde{\mathcal{R}}_X(k) = \frac{\widetilde{\mathcal{V}}(X_j, X_{j+k})}{\widetilde{\mathcal{V}}(X_j, X_j)} \tag{5.1}$$

as the *affinely invariant auto distance correlation* at the lag k. Similarly, given jointly strongly stationary, vector-valued time series $(X_j)_{j=-\infty}^{\infty}$ and $(Y_j)_{j=-\infty}^{\infty}$, we refer to

$$\widetilde{\mathcal{R}}_{X,Y}(k) = \frac{\widetilde{\mathcal{V}}(X_j, Y_{j+k})}{\sqrt{\widetilde{\mathcal{V}}(X_j, X_j)\widetilde{\mathcal{V}}(Y_j, Y_j)}} \tag{5.2}$$

as the *affinely invariant cross distance correlation* at the lag k. The corresponding sample versions can be defined in the natural way, as in the case of the non-affine distance correlation (Zhou [24]).

We illustrate these concepts on time series data of wind observations at and near the Stateline wind energy center in the Pacific Northwest of the United States. Specifically, we consider time series of bivariate wind vectors at the meteorological towers at Vansycle, right at the Stateline wind farm at the border of the states of Washington and Oregon, and at Goodnoe Hills, 146 km west of Vansycle along the Columbia River Gorge. Further information can be found in the paper by Gneiting *et al.* [3], who developed a regime-switching space-time (RST) technique for 2-hour-ahead forecasts of hourly average wind speed at the Stateline wind energy center, which was then the largest wind farm globally. For our purposes, we follow Hering and Genton [8] in studying the time series at the original 10-minute resolution, and we restrict our analysis to the longest continuous record, the 75-day interval from August 14, 2002 to October 28, 2002.

Thus, we consider time series of bivariate wind vectors over 10 800 consecutive 10-minute intervals. We write V_j^{NS} and V_j^{EW} to denote the north–south and the east–west component, respectively, of the wind vector at Vansycle at time j, with positive values corresponding to northerly and easterly winds. Similarly, we write G_j^{NS} and G_j^{EW} for the north–south and the east–west component, respectively, of the wind vector at Goodnoe Hills at time j.

Figure 3 shows the classical (Pearson) sample auto and cross correlation functions for the four univariate time series. The auto correlation functions generally decay with the temporal, but do so non-monotonously, due to the presence of a diurnal component. The cross correlation functions between the wind vector components at Vansycle and Goodnoe Hills show remarkable asymmetries and peak at positive lags, due to the prevailing westerly and southwesterly wind (Gneiting *et al.* [3]). In another interesting feature, the cross correlations between the north–south and east–west components at lag zero are strongly positive, documenting the dominance of southwesterly winds.

Figure 4 shows the sample auto and cross distance correlation functions for the four time series; as these variables are univariate, there is no distinction between the standard and the affinely invariant version of the distance correlation. The patterns seen resemble those in the case of the Pearson correlation. For comparison, we also display values of the distance correlation based on the sample Pearson correlations shown in Figure 3, and converted to distance correlation under the assumption of bivariate Gaussianity, using the results of Székely *et al.* [23], page 2785, and Section 3; in every single case, these values are smaller than the original ones.

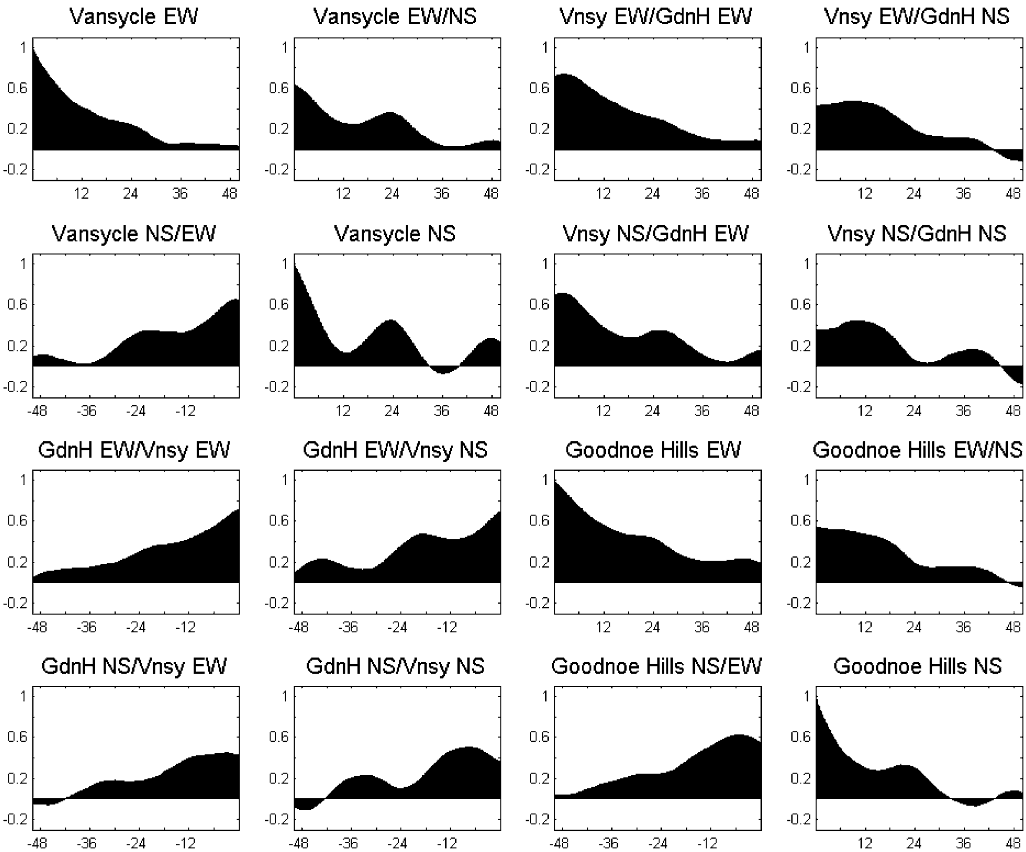

Figure 3. Sample auto and cross Pearson correlation functions for the univariate time series V_j^{EW}, V_j^{NS}, G_j^{EW}, and G_j^{NS}, respectively. Positive lags indicate observations at the westerly site (Goodnoe Hills) leading those at the easterly site (Vansycle), or observations of the north–south component leading those of the east–west component, in units of hours.

Having considered the univariate time series setting, it is natural and complementary to look at the wind vector time series $(V_j^{\mathrm{EW}}, V_j^{\mathrm{NS}})$ at Vansycle and $(G_j^{\mathrm{EW}}, G_j^{\mathrm{NS}})$ at Goodnoe Hills from a genuinely multivariate perspective. To this end, Figure 5 shows the sample affinely invariant auto and cross distance correlation functions for the bivariate wind vector series at the two sites. Again, a diurnal component is visible, and there is a remarkable asymmetry in the cross-correlation functions, which peak at lags of about two to three hours.

In light of our analytical results in Section 3, we can compute the affinely invariant distance correlation between subvectors of a multivariate normally distributed random vector. In particular, we can compute the affinely invariant auto and cross distance correlation between bivariate subvectors of a 4-variate Gaussian process with Pearson auto and cross correlations as shown in Figure 3. In Figure 5, values of the affinely invariant distance correlation that have been derived

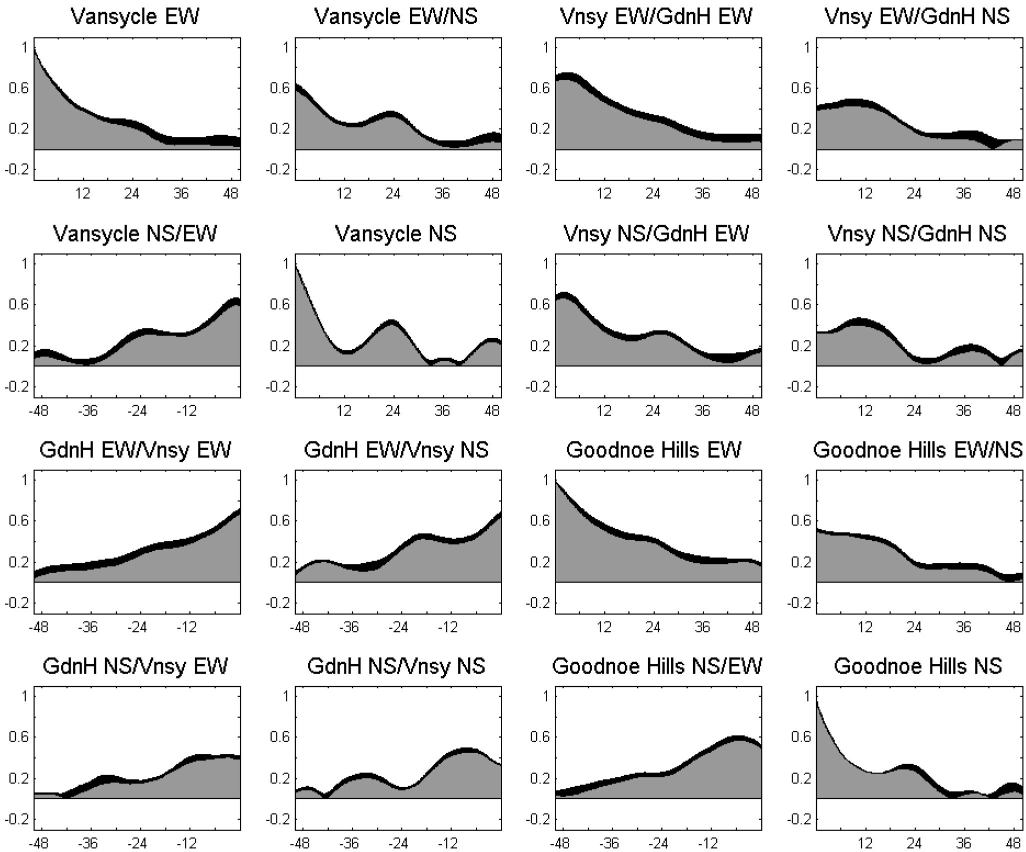

Figure 4. Sample auto and cross distance correlation functions for the univariate time series V_j^{EW}, V_j^{NS}, G_j^{EW}, and G_j^{NS}, respectively. For comparison, we also display, in grey, the values that arise when the sample Pearson correlations in Figure 3 are converted to distance correlation under the assumption of Gaussianity; these values generally are smaller than the original ones. Positive lags indicate observations at Goodnoe Hills leading those at Vansycle, or observations of the north–south component leading those of the east–west component, in units of hours.

from Pearson correlations in these ways are shown in grey; the differences from those values that are computed directly from the data are substantial, with the converted values being smaller, possibly suggesting that assumptions of Gaussianity may not be appropriate for this particular data set.

We wish to emphasize that our study is purely exploratory: it is provided for illustrative purposes and to serve as a basic example. In future work, the approach hinted at here may have the potential to be developed into parametric or nonparametric bootstrap tests for Gaussianity. For this purpose recall that, in the Gaussian setting, the affinely invariant distance correlation is a function of the canonical correlation coefficients, that is, $\widetilde{\mathcal{R}} = g(\lambda_1, \ldots, \lambda_r)$. For a parametric

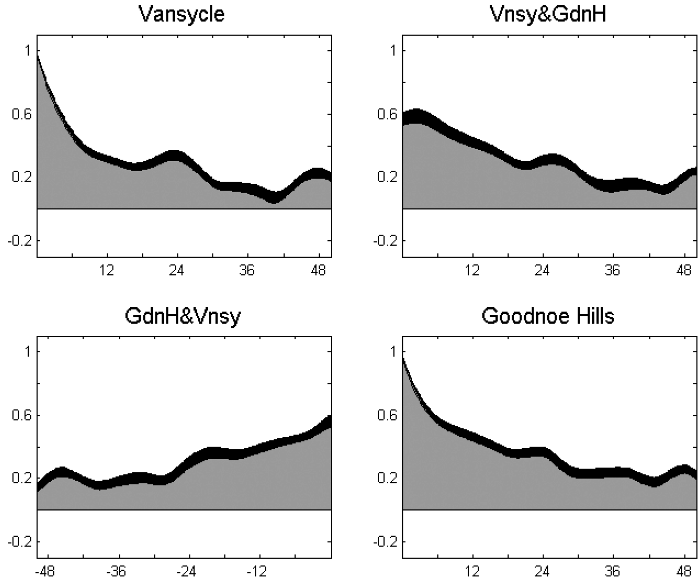

Figure 5. Sample auto and cross affinely invariant distance correlation functions for the bivariate time series $(V_j^{\mathrm{EW}}, V_j^{\mathrm{NS}})'$ and $(G_j^{\mathrm{EW}}, G_j^{\mathrm{NS}})'$ at Vansycle and Goodnoe Hills. For comparison, we also display, in grey, the values that are generated when the Pearson correlation in Figure 3 is converted to the affinely invariant distance correlation under the assumption of Gaussianity; these converted values generally are smaller than the original ones. Positive lags indicate observations at Goodnoe Hills leading those at Vansycle, in units of hours.

bootstrap test, one could generate B replicates of $g(\lambda_1^\star, \ldots, \lambda_r^\star)$, leading to a pointwise $(1-\alpha)$-confidence band. The test would now reject Gaussianity if the sample affinely invariant distance correlation function does not lie within this band. For the nonparametric bootstrap test, one could obtain ensembles $\widetilde{\mathcal{R}}_n^\star$ by resampling methods, again defining a pointwise $(1-\alpha)$-confidence band and checking if $g(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_r)$ is located within this band.

Following the pioneering work of Zhou [24], the distance correlation may indeed find a wealth of applications in exploratory and inferential problems for time series data.

6. Discussion

In this paper, we have studied an affinely invariant version of the distance correlation measure introduced by Székely *et al.* [23] and Székely and Rizzo [20] in both population and sample settings (see Székely and Rizzo [21] for further aspects of the role of invariance in properties of distance correlation measures). The affinely invariant distance correlation shares the desirable properties of the standard version of the distance correlation and equals the latter in the univariate case. In the multivariate case, the affinely invariant distance correlation remains unchanged under invertible affine transformations, unlike the standard version, which is preserved under orthogonal transformations only. Furthermore, the affinely invariant distance correlation admits an exact and readily computable expression in the case of subvectors from a multivariate normal population. We have shown elsewhere that the standard version allows for a series expansion too, but this does not appear to be a series that generally can be made simple, and further research will be necessary to make it accessible to efficient numerical computation. Related asymptotic results can be found in Gretton *et al.* [5] and Székely and Rizzo [22].

Competing measures of dependence also have featured prominently recently (Reshef *et al.* [16], Speed [19]). However, those measures are restricted to univariate settings, and claims of superior performance in exploratory data analysis have been disputed (Gorfine, Heller and Heller [4], Simon and Tibshirani [18]). We therefore share much of Newton's [14] enthusiasm about the use of the distance correlation as a measure of dependence and association. A potential drawback for large data sets is the computational cost required to compute the sample distance covariance, and the development of computationally efficient algorithms or subsampling techniques for doing this is highly desirable.

Acknowledgements

The research of Johannes Dueck, Dominic Edelmann and Tilmann Gneiting has been supported by the *Deutsche Forschungsgemeinschaft* (German Research Foundation) within the programme "Spatio/Temporal Graphical Models and Applications in Image Analysis," grant GRK 1653.

References

[1] Andrews, G.E., Askey, R. and Roy, R. (1999). *Special Functions*. *Encyclopedia of Mathematics and Its Applications* **71**. Cambridge: Cambridge Univ. Press. MR1688958

[2] Eaton, M.L. (1989). *Group Invariance Applications in Statistics*. *NSF-CBMS Regional Conference Series in Probability and Statistics* **1**. Hayward, CA: IMS. MR1089423

[3] Gneiting, T., Larson, K., Westrick, K., Genton, M.G. and Aldrich, E. (2006). Calibrated probabilistic forecasting at the Stateline wind energy center: The regime-switching space-time method. *J. Amer. Statist. Assoc.* **101** 968–979. MR2324108

[4] Gorfine, M., Heller, R. and Heller, Y. (2012). Comment on "Detecting novel associations in large data sets." Unpublished manuscript. Available at http://iew3.technion.ac.il/~gorfinm/files/science6.pdf.

[5] Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B. and Smola, A. (2012). A kernel two-sample test. *J. Mach. Learn. Res.* **13** 723–773. MR2913716

[6] Gross, K.I. and Richards, D.S.P. (1987). Special functions of matrix argument. I. Algebraic induction, zonal polynomials, and hypergeometric functions. *Trans. Amer. Math. Soc.* **301** 781–811. MR0882715

[7] Heller, R., Heller, Y. and Gorfine, M. (2013). A consistent multivariate test of association based on ranks of distances. *Biometrika* **100** 503–510. MR3068450

[8] Hering, A.S. and Genton, M.G. (2010). Powering up with space-time wind forecasting. *J. Amer. Statist. Assoc.* **105** 92–104. MR2757195

[9] James, A.T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 475–501. MR0181057

[10] Koev, P. and Edelman, A. (2006). The efficient evaluation of the hypergeometric function of a matrix argument. *Math. Comp.* **75** 833–846. MR2196994

[11] Kosorok, M.R. (2009). Discussion of: Brownian distance covariance. *Ann. Appl. Stat.* **3** 1270–1278. MR2752129

[12] Kosorok, M.R. (2013). Correction: Discussion of Brownian distance covariance. *Ann. Appl. Stat.* **7** 1247. MR3113509

[13] Muirhead, R.J. (1982). *Aspects of Multivariate Statistical Theory*. New York: Wiley. MR0652932

[14] Newton, M.A. (2009). Introducing the discussion paper by Székely and Rizzo. *Ann. Appl. Stat.* **3** 1233–1235. MR2752126

[15] Rémillard, B. (2009). Discussion of: Brownian distance covariance. *Ann. Appl. Stat.* **3** 1295–1298. MR2752133

[16] Reshef, D.N., Reshef, J.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., Lander, E.S., Mitzenmacher, M. and Sabeti, P.C. (2011). Detecting novel associations in large data sets. *Science* **334** 1518–1524.

[17] Rizzo, M.L. and Székely, G.J. (2011). Energy: E-statistics (energy statistics). R package, Version 1.4-0. Available at http://cran.us.r-project.org/web/packages/energy/index.html.

[18] Simon, N. and Tibshirani, R. (2012). Comment on "Detecting novel associations in large data sets," by Reshef *et al. Science* **334** (2011) 1518–1524. Unpublished manuscript. Available at http://www-stat.stanford.edu/~tibs/reshef/comment.pdf.

[19] Speed, T. (2011). Mathematics. A correlation for the 21st century. *Science* **334** 1502–1503.

[20] Székely, G.J. and Rizzo, M.L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. MR2752127

[21] Székely, G.J. and Rizzo, M.L. (2012). On the uniqueness of distance covariance. *Statist. Probab. Lett.* **82** 2278–2282. MR2979766

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

[23] Székely, G.J., Rizzo, M.L. and Bakirov, N.K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. MR2382665

[24] Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.* **33** 438–457. MR2915095

Received April 2013 and revised August 2013