Journal of Multivariate Analysis 182 (2021) 104711

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# Testing independence of functional variables by angle covariance

Tingyu Lai [a], Zhongzhan Zhang [a,*], Yafei Wang [b], Linglong Kong [b]

[a] Beijing University of Technology, Beijing, 100124, China
[b] University of Alberta, Edmonton, AB, T6G 2R3, Canada

A R T I C L E I N F O

*Article history:*
Received 23 August 2020
Received in revised form 21 November 2020
Accepted 22 November 2020
Available online 8 December 2020

*AMS 2000 subject classifications:*
62G10
62H20

*Keywords:*
Angle covariance
Distance covariance
Hilbert space
Projection correlation
Test of independence

A B S T R A C T

We propose a new nonparametric independence test for two functional random variables. The test is based on a new dependence metric, the so-called angle covariance, which fully characterizes the independence of the random variables and generalizes the projection covariance proposed for random vectors. The angle covariance has a number of desirable properties, including the equivalence of its zero value and the independence of the two functional variables, and it can be applied to any functional data without finite moment conditions. We construct a $V$-statistic estimator of the angle covariance, and show that it has a Gaussian chaos limiting distribution under the independence null hypothesis and a normal limiting distribution under the alternative hypothesis. The test based on the estimated angle covariance is consistent against all alternatives and easy to be implemented by the given random permutation method. Simulations show that the test based on the angle covariance outperforms other competing tests for functional data.



## 1. Introduction

Over recent decades, functional data analysis (FDA) has been developed rapidly and become an important area of statistics. FDA offers effective tools for the analysis of high or infinite dimensional data, and meets the growing needs for data collection and analysis with the progress in technology. Many aspects of FDA, such as functional regression [11], clustering and classification of functional data [27], have been extensively investigated, and a number of excellent monographs about FDA have been published, see, for example, [3,7,8,14]. However, relatively little works focus on measuring and testing the dependence of two functional random variables.

Testing the independence of random elements is a fundamental problem in statistics and has important applications. It has been studied by many authors, and several excellent methods have been developed, including the distance correlation [21,22], the kernel based criterion [4,5,20], the maximal information coefficient [15], the copula based measures [16,19], the projection correlation [28], the ball covariance [12], and others (see, for example, [23]). Many of these methods were developed for random variables in Euclidean spaces and may not be directly applied to functional data, which have infinite dimensions. Among them, an indispensable method is the distance correlation [21,22], which can be used to measure and test the dependence between random vectors $X$ and $Y$, provided that $E(\|X\| + \|Y\|) < \infty$. It has been shown to work well in a number of situations, and extended to strong negative type metric space [10], and thus can be applicable for functional data. In spite of its advantages, the distance covariance requires the finite first moment.

* Corresponding author.
*E-mail address:* zzhang@bjut.edu.cn (Z. Zhang).

https://doi.org/10.1016/j.jmva.2020.104711

When the condition is violated, it may not behave well, as illustrated in Zhu et al. [28]. To remove the moment condition, Zhu et al. [28] proposed a projection correlation for two random vectors based on the projections of the random vectors. The projection correlation test is powerful in some cases, but it only deals with the random vectors in the Euclidean setting, and cannot be used straightforwardly to functional data. Actually, our simulations show that a naive use of the projection correlation test could suffer a loss of power for functional data. Another powerful tool suitable for functional data is the ball covariance [12]. It is designed to measure and test the dependence of random elements in Banach spaces, and can be applied to many kinds of data. Both the distance covariance and the ball covariance test statistics are constructed with the metrics of the underlying spaces. Besides the metrics endowed in the underlying spaces, one may utilize more information such as the features of the data or the geometric structures of the underlying spaces. For example, a functional variable is usually considered as an element of a Hilbert space and has some properties of functions such as periodicity or monotonicity. It could improve the performance of an independence test if we combine these kinds of information in an appropriate way.

In view of the preceding discussion, we would like to establish a general independence test for two random elements valued in two separable Hilbert spaces, which requires less strict conditions such as finite moments, and uses more information of the data in hand. More precisely, let $(\Omega, \mathcal{B}, P)$ be a probability space, and $(X, Y) : \Omega \mapsto \mathcal{H}_1 \times \mathcal{H}_2$ be a vector of random elements, where $\mathcal{H}_1, \mathcal{H}_2$ are two separable Hilbert spaces. We aim to test

$$H_0 : X \text{ and } Y \text{ are independent} \qquad \text{vs.} \qquad H_1 : \text{otherwise}. \tag{1}$$

To attain this goal, we first construct a quantity with projection and integration skill in separable Hilbert space, called angle covariance, to measure and test the dependence of the random elements. The angle covariance involves the inner products of the Hilbert spaces and some nondegenerate Gaussian measures on the spaces, thus it can combine the geometric properties of Hilbert spaces and the features of the data by choosing appropriate Gaussian measures. The angle covariance of $X$ and $Y$ is always nonnegative and is equal to zero if and only if $X$ and $Y$ are independent, and it is equal to the projection covariance in Zhu et al. [28] in finite dimensional settings with the choice of the standard multivariate normal distributions (see Remark 1 in Section 2). Then, we provide an empirical estimator of the angle covariance, and give its asymptotic properties. It is shown that the estimator is $n$-consistent if $X$ and $Y$ are independent, and $\sqrt{n}$-consistent otherwise. Correspondingly, the test of independence based on the angle covariance is consistent against all alternatives, and easy to be implemented. The proposed test does not require any restriction to the underlining distributions. Through numerical simulations, we show that it is more powerful than those based on other dependence metrics, such as distance covariance [22], projection covariance [28], and ball covariance [12] for functional data.

The rest of the paper is organized as follows: In Section 2, we explore the procedure of constructing angle covariance in Hilbert spaces. In Section 3, we give an estimator of the angle covariance and show its asymptotic properties. In Section 4, we present the test procedure based on angle covariance in practice and provide an empirical criterion for the choice of Gaussian measures. A group of finite sample simulation studies is carried out in Section 5. In Section 6, some discussions are included. All technical proofs are presented in the Appendix.

## 2. Angle covariance

In this section, we define the angle covariance of two random elements in separable Hilbert spaces (hereinafter all Hilbert spaces mentioned are separable). Suppose that $X$ and $Y$ are the two random elements defined in the previous section. For simplification of notations, we denote the associated norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$ for both $\mathcal{H}_1$ and $\mathcal{H}_2$.

To test the independence of two random elements in Hilbert spaces, a natural approach is to convert the independence of the functional random elements into that of real-valued random variables. The following lemma ensures that this is feasible.

**Lemma 1.** *$X$ and $Y$ are independent if and only if $\langle X, f\rangle$ and $\langle Y, g\rangle$ are independent for all $f \in \mathcal{H}_1$, $g \in \mathcal{H}_2$.*

Instead of considering the original random elements, Lemma 1 shows that we only need to consider the projections of the random elements, which are real variables. Given the projection directions $f$ and $g$, let $V = \langle X, f\rangle$, $W = \langle Y, g\rangle$, $F(v, w)$ be the joint distribution of $(V, W)$, and $F_1(v)$ and $F_2(w)$ be the marginal distributions of $V$ and $W$, respectively. Then $V$ and $W$ are independent if and only if $F(v, w) - F_1(v)F_2(w) = \operatorname{cov}\{I(\langle X, f\rangle \le v), I(\langle Y, g\rangle \le w)\} = 0$, for all $v, w \in \mathbb{R}$, where $I(\cdot)$ is the indicator function. Therefore, by integrating the squared covariance, the dependence of $X$ and $Y$ can be measured by the quantity

$$\int_{\mathcal{H}_1}\int_{\mathcal{H}_2}\int_{\mathbb{R}^2} \operatorname{cov}^2\{I(\langle X, f\rangle \le v), I(\langle Y, g\rangle \le w)\} dF(v, w)\mu_1(df)\mu_2(dg), \tag{2}$$

where $\mu_1$ and $\mu_2$ are two nondegenerate Gaussian measures in $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. The integral (2) is complex and difficult to deal with. Motivated by Zhu et al. [28], we try to give an explicit form of the integral. Let $(X^i, Y^i)$, $i \in \{1, \ldots, 5\}$

be independent copies of $(X, Y)$ and denote $\langle X^i, f\rangle = X^i_f$, $\langle Y^i, g\rangle = Y^i_g$. Then by Fubini's theorem the integral can be rewritten as

$$\mathrm{E}\Bigg[\int_{\mathcal{H}_1}\int_{\mathcal{H}_2}\left\{I(X^1_f \le X^3_f)I(Y^1_g \le Y^3_g) - I(X^1_f \le X^3_f)I(Y^2_g \le Y^3_g)\right\} \times \left\{I(X^4_f \le X^3_f)I(Y^4_g \le Y^3_g) - I(X^4_f \le X^3_f)I(Y^5_g \le Y^3_g)\right\}\mu_1(df)\mu_2(dg)\Bigg]. \quad (3)$$

To compute the integrals over $\mathcal{H}_1$ and $\mathcal{H}_2$, recall the fact that a Gaussian measure on a separable Hilbert space can be expressed as countable products of measures (see Theorem 1.11 of [2]). With this merit we obtain the following theorem, which is a generalization of Lemma 1 in Zhu et al. [28].

Suppose that $\mu$ is a nondegenerate Gaussian measure on a separable Hilbert space $\mathcal{H}$ with mean zero and covariance operator $Q$, and the eigenvalues and orthonormal eigenfunctions of $Q$ are $(\lambda_i, e_i)_{i=1}^{\infty}$. For $x \in \mathcal{H}$, define $Q^{1/2}x = \sum_{i=1}^{\infty}\sqrt{\lambda_i}\langle x, e_i\rangle e_i$. Then we have the following theorem.

**Theorem 1.** *For any two nonzero elements $U, V$ in a separable Hilbert space $\mathcal{H}$, we have*

$$\int_{\mathcal{H}} I(\langle U, t\rangle \le 0)I(\langle V, t\rangle \le 0)\mu(dt) = \frac{1}{2} - \frac{1}{2\pi}\arccos\left(\frac{\langle Q^{1/2}U, Q^{1/2}V\rangle}{\|Q^{1/2}U\|\cdot\|Q^{1/2}V\|}\right),$$

*where* $\arccos(\cdot)$ *is the inverse cosine function.*

With the help of Theorem 1, we can simplify the integral (3) into an explicit formula. Denote

$$\theta_Q(X, X', X'') = \arccos\left(\frac{\langle Q^{1/2}(X - X''), Q^{1/2}(X' - X'')\rangle}{\|Q^{1/2}(X - X'')\|\cdot\|Q^{1/2}(X' - X'')\|}\right),$$

then $\theta_Q(X, X', X'')$ is the angle between $Q^{1/2}(X - X'')$ and $Q^{1/2}(X' - X'')$. When the denominator is zero, we define $\theta_Q(X, X', X'')$ as zero. Choosing nondegenerate Gaussian measures $\mu_1$, $\mu_2$ on $\mathcal{H}_1$, $\mathcal{H}_2$ with mean zero and covariance operators $Q_1$, $Q_2$ respectively, and ignoring the multiplier constant, we define the resulting quantity from integral (3) as the squared angle covariance.

**Definition 1.** The squared angle covariance between $X$ and $Y$ is defined as

$$\begin{aligned}\mathrm{Acov}^2(X, Y) =& \mathrm{E}\left\{\theta_{Q_1}(X^1, X^4, X^3)\theta_{Q_2}(Y^1, Y^4, Y^3)\right\} + \mathrm{E}\left\{\theta_{Q_1}(X^1, X^4, X^3)\theta_{Q_2}(Y^2, Y^5, Y^3)\right\} \\ &- 2\mathrm{E}\left\{\theta_{Q_1}(X^1, X^4, X^3)\theta_{Q_2}(Y^2, Y^4, Y^3)\right\},\end{aligned}$$

and the angle covariance $\mathrm{Acov}(X, Y)$ is the square root of $\mathrm{Acov}^2(X, Y)$.

**Remark 1.** The angle covariance is a natural extension of projection covariance [28] to functional data. Specifically, if the dimensions of the underlying Hilbert spaces are finite and the chosen Gaussian measures are the standard multivariate Gaussian measures (with mean zero and identity covariance matrix), then the angle covariance coincides with the projection covariance. However, when the dimension of an underlying Hilbert space, say $\mathcal{H}_1$, is infinite, it is different. In this case, the covariance operator $Q_1$ is a compact operator, and $Q_1^{1/2}(\mathcal{H}_1)$ is the Cameron–Martin space of the measure $\mu_1$, which is a dense subspace of $\mathcal{H}_1$ (see Chapter 2 of [2]).

The rest of this section reveals some properties of $\mathrm{Acov}^2(X, Y)$. Let

$$\phi(x, x', x'') = \theta_{Q_1}(x, x', x'') - \mathrm{E}\left\{\theta_{Q_1}(X^1, x', x'')\right\} - \mathrm{E}\left\{\theta_{Q_1}(x, X^2, x'')\right\} + \mathrm{E}\left\{\theta_{Q_1}(X^1, X^2, x'')\right\},$$

and

$$\varphi(y, y', y'') = \theta_{Q_2}(y, y', y'') - \mathrm{E}\left\{\theta_{Q_2}(Y^1, y', y'')\right\} - \mathrm{E}\left\{\theta_{Q_2}(y, Y^2, y'')\right\} + \mathrm{E}\left\{\theta_{Q_2}(Y^1, Y^2, y'')\right\}.$$

**Proposition 1.** *Let $(X^i, Y^i)$, $i = 1, 2, 3$ be independent copies of $(X, Y)$. Then we have*

(i) $\mathrm{Acov}^2(X, Y) = \mathrm{E}\left\{\phi(X^1, X^2, X^3)\varphi(Y^1, Y^2, Y^3)\right\}$;
(ii) $\mathrm{Acov}^2(X, Y) \le \sqrt{\mathrm{E}\left\{\phi^2(X^1, X^2, X^3)\right\}\mathrm{E}\left\{\varphi^2(Y^1, Y^2, Y^3)\right\}}$;
(iii) $\mathrm{Acov}(X, Y) = 0$ *if and only if $X$ and $Y$ are independent;*
(iv) *Let $a_1 \in \mathcal{H}_1$, $a_2 \in \mathcal{H}_2$, $b_1, b_2$ be two nonzero scalar constants, then* $\mathrm{Acov}(b_1X + a_1, b_2Y + a_2) = \mathrm{Acov}(X, Y)$.

Proposition 1 is parallel to Proposition 1 of Zhu et al. [28]. However, for unitary operators $U_1$, $U_2$ on $\mathcal{H}_1$, $\mathcal{H}_2$ respectively, equality $\mathrm{Acov}(b_1U_1X + a_1, b_2U_2Y + a_2) = \mathrm{Acov}(X, Y)$ is usually not true, since a nondegenerate Gaussian measure on a Hilbert space of infinite dimension is not symmetric with respect to its eigenfunctions. In addition, Proposition 1(ii) shows that the ratio of $\mathrm{Acov}^2(X, Y)$ to the right hand side might be used to record the dependence level between $X$ and $Y$, but it should be noted that the ratio depends on the choices of $\mu_1$ and $\mu_2$.

**Remark 2.** Theorem 1 and Proposition 1 show the advantages of employing the Gaussian measures in the dependence measure (3), although other choices of $\mu_1$ and $\mu_2$ might be possible. Firstly, by choosing the Gaussian measures, we obtain the explicit expression of quantity (3), which allows us to express the dependence of the random elements with their geometric characteristics. Secondly, we get the equivalence between the zero angle covariance and the independence of two random elements. These advantages make it easier to estimate the angle covariance, and to obtain a consistent test for independence.

## 3. Empirical estimator and asymptotic results

We now give an estimator of $\mathrm{Acov}^2(X, Y)$ by Proposition 1(i). Let $\{(X_i, Y_i), i \in \{1, \ldots, n\}\}$ be an independent and identically distributed (i.i.d.) sample of $(X, Y)$. Define, for $i, j, k \in \{1, \ldots, n\}$,

$$a_{ijk} = \theta_{Q_1}(X_i, X_j, X_k), \qquad a_{i\cdot k} = n^{-1}\sum_{j=1}^{n} a_{ijk},$$
$$a_{\cdot jk} = n^{-1}\sum_{i=1}^{n} a_{ijk}, \qquad a_{\cdot\cdot k} = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ijk},$$
$$A_{ijk} = a_{ijk} - a_{i\cdot k} - a_{\cdot jk} + a_{\cdot\cdot k}.$$

Similarly, define $b_{ijk} = \theta_{Q_2}(Y_i, Y_j, Y_k)$ and $B_{ijk} = b_{ijk} - b_{i\cdot k} - b_{\cdot jk} + b_{\cdot\cdot k}$, for $i, j, k \in \{1, \ldots, n\}$. When $i = k$ or $j = k$, to avoid possible confusion, we define $a_{ijk} = b_{ijk} = 0$. Then an estimator of $\mathrm{Acov}^2(X, Y)$ is given by

$$\mathrm{Acov}_n^2(X, Y) = n^{-3}\sum_{i,j,k=1}^{n} A_{ijk}B_{ijk}. \tag{4}$$

This estimator is easy to compute when using the permutation method given below to obtain the critical value of the test. Furthermore, we have the following theorems.

**Theorem 2.** *For an i.i.d. random sample $\{(X_i, Y_i), i \in \{1, \ldots, n\}\}$, the estimator (4) equals*

$$2\pi n^{-1}\sum_{i=1}^{n}\left[\int\int\{\hat{F}(\langle X_i, f\rangle, \langle Y_i, g\rangle) - \hat{F}_1(\langle X_i, f\rangle)\hat{F}_2(\langle Y_i, g\rangle)\}^2\mu_1(df)\mu_2(dg)\right],$$

*where $\hat{F}$, $\hat{F}_1$ and $\hat{F}_2$ stand for the empirical distributions of $(\langle X, f\rangle, \langle Y, g\rangle)$, $\langle X, f\rangle$ and $\langle Y, g\rangle$, respectively.*

Theorem 2 recovers the original formulation (2), which is more intuitive, and indicates that $\mathrm{Acov}_n^2(X, Y) \geq 0$. In the Appendix, we show that $\mathrm{Acov}_n^2(X, Y)$ is a $V$-statistic with a kernel of degree 7, and it is degenerate with rank 2 under $H_0$. Therefore, we obtain its consistency and asymptotic distributions according to the limit theory of $V$-statistics (or $U$-statistics).

**Theorem 3.** *For an i.i.d. random sample $\{(X_i, Y_i), i \in \{1, \ldots, n\}\}$, we have*

$$\lim_{n\to\infty} \mathrm{Acov}_n(X, Y) = \mathrm{Acov}(X, Y)$$

*almost surely.*

**Theorem 4.** *For an i.i.d. random sample $\{(X_i, Y_i), i \in \{1, \ldots, n\}\}$, we have*

(i) *Under $H_0$, $n \cdot \mathrm{Acov}_n^2(X, Y) \xrightarrow{d} \sum_{i=1}^{\infty}\gamma_i Z_i^2$, where $Z_i$s are i.i.d. standard normal random variables and the nonnegative constants $\gamma_i$s depend on the distribution of $(X, Y)$ and the chosen Gaussian measures.*
(ii) *Under $H_1$, $n^{1/2}\{\mathrm{Acov}_n^2(X, Y) - \mathrm{Acov}^2(X, Y)\} \xrightarrow{d} \mathcal{N}(0, 7^2\zeta_1)$, where $\zeta_1$ is defined in the proof.*

## 4. Test procedure

According to the result of Theorem 4, we define the test statistic for the hypothesis (1) as

$$T_n = n \cdot \mathrm{Acov}_n^2(X, Y), \tag{5}$$

and reject $H_0$ when $T_n$ is large. Theorems 3 and 4 indicate that the test statistic $T_n$ converges in distribution to a weighted sum of independent $\chi^2$ variables with 1 degree of freedom (which is a special Gaussian chaos) if $X$ and $Y$ are independent, and diverges to $\infty$ otherwise. Therefore, the test based on the angle covariance is consistent against all alternatives without requiring any moment conditions.

Since the $\gamma_i$s in Theorem 4 are unknown, the asymptotic distribution cannot be used directly to compute the critical value of the test statistic $T_n$. In practice, we use a permutation method to approximate the asymptotic null distribution of $T_n$. The algorithm generating an approximated $p$-value is given as follows:

**Algorithm 1** (*p-value*)**.**

(1) For an i.i.d. random sample $\{(X_i, Y_i), i \in \{1, \ldots, n\}\}$, one computes the statistic $T_n$ using (5).
(2) For each $\ell$, $1 \le \ell \le L$, generate a random permutation of $Y = (Y_1, \ldots, Y_n)$, denoted as $Y^{(\ell)} = (Y_1^{(\ell)}, \ldots, Y_n^{(\ell)})$.
(3) Compute the test statistic with $\{(X_i, Y_i^{(\ell)}), i \in \{1, \ldots, n\}\}$, denoted as $T_n^{(\ell)}$.
(4) Repeat steps (2) and (3) $L$ times and collect data $T_n^{(1)}, \ldots, T_n^{(L)}$. The $p$-value obtained from this permutation procedure is

$$p\text{-value} = \frac{\sum_{\ell=1}^{L} I(T_n^{(\ell)} \ge T_n)}{L}.$$

One rejects the null hypothesis $H_0$ if the $p$-value is smaller than the given significance level.

The permutation method is widely used in independence testing problems, see, for example, Székely et al. [22], Székely and Rizzo [21], Gretton et al. [5], Pfister et al. [13] and Shen et al. [18]. It has desirable performance even if the sample size is small. In our simulations, $L = 300$ permutations achieve a good control of Type-I error frequencies.

**Remark 3.** The choice of the Gaussian measures with zero means influences the practical performance of the angle covariance test. We make the choice with the eigenvalues and the corresponding orthonormal eigenfunctions of their covariance operators, and give the following empirical criterion to determine a suitable nondegenerate Gaussian measure. Firstly, choose a suitable orthonormal basis $(e_i)_{i=1}^{\infty}$ for the expression of functional data. For example, if the data has periodic property then we may use the Fourier basis. Some useful principles for choosing basis system are found in Chapter 3 of Ramsay and Silverman [14]. After choosing a suitable basis, we choose a positive series $(\lambda_i)_{i=1}^{\infty}$ satisfying $\sum_{i=1}^{\infty} \lambda_i < \infty$, and construct the covariance operator by $Qx = \sum_{i=1}^{\infty} \lambda_i \langle x, e_i \rangle e_i$, such that $Q$ has eigenvalues and eigenfunctions $(\lambda_i, e_i)_{i=1}^{\infty}$ and decide a Gaussian measure uniquely. Note that these eigenfunctions are the basis system for the expression of the associated functional data and are used in the computation of $T_n$. In the simulations followed, we take $\lambda_i = 1/i^a$ with $a > 1$. See Sections 5 and 6 for more discussions.

## 5. Simulation study

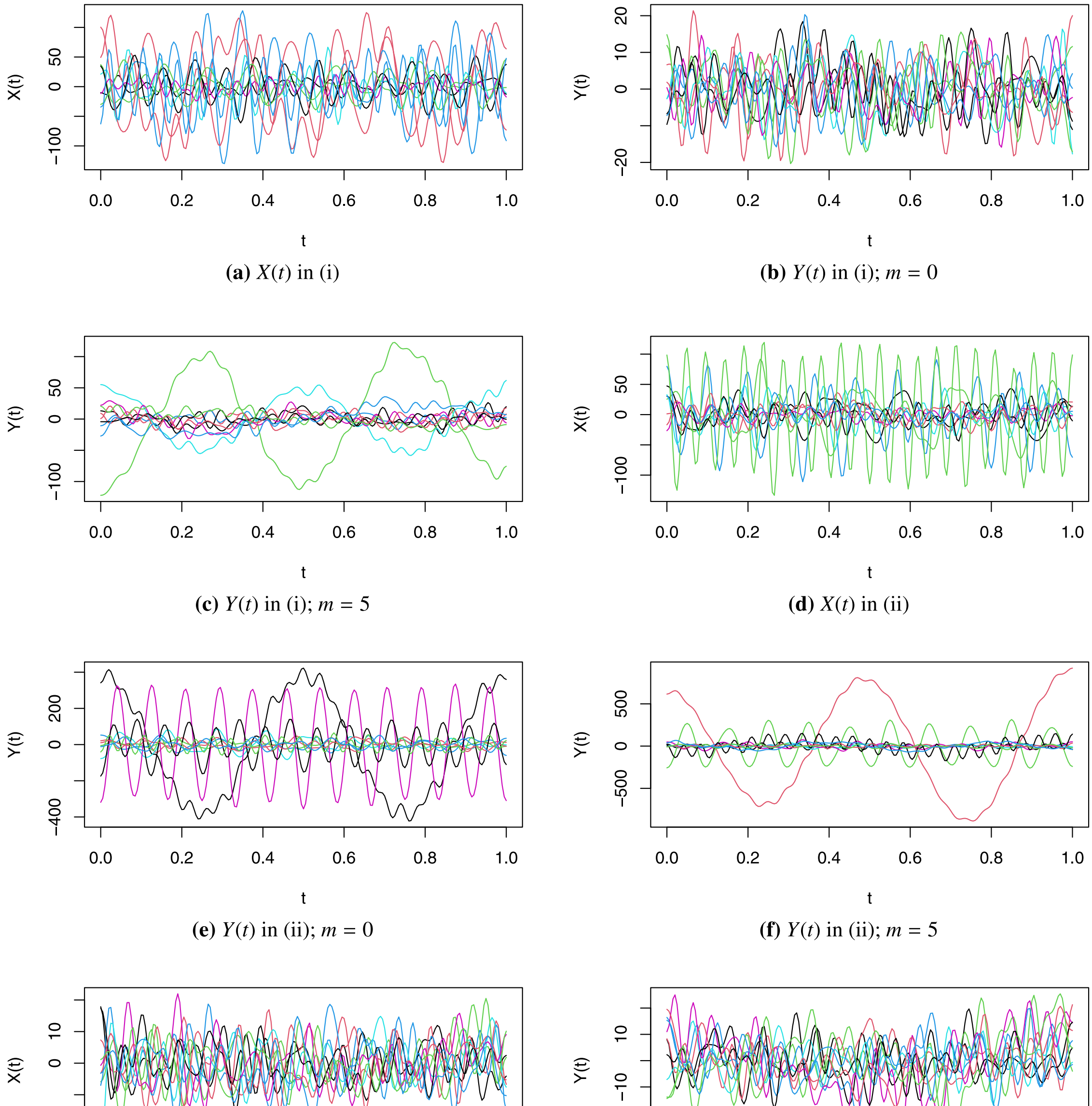
In this section, we investigate the finite sample performance of the proposed independence test by simulations. Three experiment examples are designed to evaluate the real size and the power of the test. We consider to test the independence of two functional variables in Examples 1 and 2, and the independence between a functional variable and a scalar variable in Example 3. All the functional data produced in the simulations are defined on [0, 1], and are observed at 201 equal-spaced points in [0, 1]. The eigenfunctions $(e_i)_{i=1}^{\infty}$ of the covariance operator of the Gaussian measures are chosen according to the feature of the data, and the corresponding eigenvalues are set to $\lambda_i = 1/i^a$ with $a = 3$ and $a = 2.5$. The permutation number is taken as $L = 300$.

Several independence tests are compared in the simulations. For convenience, we denote our test as acov (acov$_1$ for $a = 3$ and acov$_2$ for $a = 2.5$), and the tests based on the projection covariance [28], the distance covariance [22], and the ball covariance [12] as pcov, dcov and bcov respectively. Here, the pcov is carried out for the observed 201 dimension vectors of the functions. We implement dcov by calling `dcov.test` function in `R` package `energy` and bcov by calling `bcov.test` function in `R` package `Ball`.

Different sample sizes and dependence levels between the two random variables are considered in the simulations to evaluate the performance of tests, and the empirical size or power of the tests (the rejection proportions) are recorded through 1000 repetitions at significance level 0.05 for each setting.

**Example 1.** This example is designed to be parallel to Example 1 in Zhu et al. [28], which was for real vectors. It consists of the following three scenarios.

(i) Similar to the functional data considered in Hall and Hosseini-Nasab [6], we take $X(t) = \sum_{i=1}^{p} \xi_i \phi_i^*(t)$, where $\phi_i^*(t) = \sqrt{2}\cos(i\pi t)$ and $\xi_i$s are independent random variables distributed as the Cauchy distribution with location zero and scale 0.5. Let $Y(t) = \sum_{i=1}^{p} \gamma_i \phi_i^*(t)$, where $\gamma_i = f(\xi_i)$ for $i \in \{1, \ldots, m\}$ and $\gamma_i$ ($i \in \{m+1, \ldots, p\}$) are independently generated from the standard normal distribution. We consider four relationships: $f(x) = x^3$, $f(x) = x^2$, $f(x) = \sin(x)$ and $f(x) = \cos(x)$. Note that the last three functions are not monotone.
(ii) The conditions are the same as scenario (i), except that $\gamma_i$ ($i \in \{m+1, \ldots, p\}$) are sampled independently from the Cauchy distribution with location zero and scale 0.5.
(iii) The conditions are the same as scenario (i), except that $\xi_i$ ($i \in \{1, \ldots, p\}$) are sampled independently from the standard normal distribution.



**(a)** $X(t)$ in (i) **(b)** $Y(t)$ in (i); $m = 0$

**(c)** $Y(t)$ in (i); $m = 5$ **(d)** $X(t)$ in (ii)

**(e)** $Y(t)$ in (ii); $m = 0$ **(f)** $Y(t)$ in (ii); $m = 5$

**(g)** $X(t)$ in (iii) **(h)** $Y(t)$ in (iii); $m = 5$

**Fig. 1.** Observed curves of $X$ and $Y$ in Example 1 with $f(x) = x^3$ and 10 observed curves are shown in each graph.

In this example, the parameter $m$ is designed to produce different dependence levels between the two random variables, where $m = 0$ implies that $X$ and $Y$ are independent, and the dependence level increases with $m$. In all the three scenarios, we fix the sample size to $n = 30$ and $p = 50$, and set $m = 0, m = 1, m = 3, m = 5, m = 10$ respectively to generate $X$s and $Y$s. To implement our test, we plot the functional data. As shown in Fig. 1, the curves of $X$ and $Y$ reveal periodic feature, hence we choose the Fourier basis as the eigenfunctions of the covariance operators of the Gaussian measures.

Tables 1–3 summarize the empirical sizes and powers of the tests for all the settings. The empirical sizes (for the settings with $m = 0$) are very close to the nominal significance level, and the acov tests have higher powers (for the settings with $m > 0$) than other tests in all the scenarios. It seems that the angle covariance is more sensitive to the weak dependence relationships, as is shown, for example, in the case of $f(x) = x^3$ and $m = 1$. From these results, we can

**Table 1**
Empirical sizes and powers: Example 1(i). $H_0$ is true when $m = 0$, and $H_1$ holds when $m > 0$. The empirical sizes of all the tests are close to the significance level and the acov tests have higher powers than other tests.

| Relationship | Method | $m = 0$ | $m = 1$ | $m = 3$ | $m = 5$ | $m = 10$ |
|---|---|---|---|---|---|---|
| $f(x) = x^3$ | $acov_1$ | 0.037 | 0.933 | 1.000 | 1.000 | 1.000 |
| | $acov_2$ | 0.038 | 0.928 | 1.000 | 1.000 | 1.000 |
| | pcov | 0.051 | 0.156 | 0.372 | 0.521 | 0.804 |
| | dcov | 0.054 | 0.083 | 0.155 | 0.235 | 0.370 |
| | bcov | 0.044 | 0.102 | 0.178 | 0.265 | 0.629 |
| $f(x) = x^2$ | $acov_1$ | 0.043 | 0.381 | 0.781 | 0.901 | 0.989 |
| | $acov_2$ | 0.047 | 0.373 | 0.779 | 0.896 | 0.980 |
| | pcov | 0.052 | 0.105 | 0.192 | 0.271 | 0.515 |
| | dcov | 0.054 | 0.075 | 0.163 | 0.210 | 0.387 |
| | bcov | 0.054 | 0.079 | 0.149 | 0.244 | 0.553 |
| $f(x) = \sin(x)$ | $acov_1$ | 0.057 | 0.154 | 0.479 | 0.584 | 0.619 |
| | $acov_2$ | 0.054 | 0.185 | 0.542 | 0.644 | 0.676 |
| | pcov | 0.056 | 0.055 | 0.042 | 0.047 | 0.057 |
| | dcov | 0.055 | 0.052 | 0.048 | 0.051 | 0.053 |
| | bcov | 0.051 | 0.048 | 0.048 | 0.039 | 0.066 |
| $f(x) = \cos(x)$ | $acov_1$ | 0.048 | 0.106 | 0.196 | 0.241 | 0.255 |
| | $acov_2$ | 0.053 | 0.110 | 0.208 | 0.247 | 0.247 |
| | pcov | 0.053 | 0.056 | 0.050 | 0.048 | 0.062 |
| | dcov | 0.046 | 0.063 | 0.042 | 0.042 | 0.056 |
| | bcov | 0.046 | 0.060 | 0.053 | 0.054 | 0.061 |

**Table 2**
Empirical sizes and powers: Example 1(ii). $H_0$ is true when $m = 0$, and $H_1$ holds when $m > 0$. The empirical sizes of all the tests are close to the significance level and the acov tests have higher powers than other tests.

| Relationship | Method | $m = 0$ | $m = 1$ | $m = 3$ | $m = 5$ | $m = 10$ |
|---|---|---|---|---|---|---|
| $f(x) = x^3$ | $acov_1$ | 0.053 | 0.872 | 1.000 | 1.000 | 1.000 |
| | $acov_2$ | 0.053 | 0.880 | 1.000 | 1.000 | 1.000 |
| | pcov | 0.051 | 0.133 | 0.293 | 0.522 | 0.797 |
| | dcov | 0.050 | 0.080 | 0.141 | 0.214 | 0.366 |
| | bcov | 0.050 | 0.072 | 0.140 | 0.281 | 0.591 |
| $f(x) = x^2$ | $acov_1$ | 0.049 | 0.344 | 0.747 | 0.897 | 0.991 |
| | $acov_2$ | 0.052 | 0.370 | 0.749 | 0.882 | 0.982 |
| | pcov | 0.053 | 0.092 | 0.165 | 0.284 | 0.504 |
| | dcov | 0.046 | 0.075 | 0.136 | 0.214 | 0.346 |
| | bcov | 0.047 | 0.072 | 0.108 | 0.206 | 0.470 |
| $f(x) = \sin(x)$ | $acov_1$ | 0.042 | 0.089 | 0.158 | 0.230 | 0.409 |
| | $acov_2$ | 0.043 | 0.094 | 0.222 | 0.301 | 0.444 |
| | pcov | 0.046 | 0.050 | 0.052 | 0.056 | 0.043 |
| | dcov | 0.048 | 0.050 | 0.053 | 0.054 | 0.061 |
| | bcov | 0.053 | 0.046 | 0.051 | 0.049 | 0.039 |
| $f(x) = \cos(x)$ | $acov_1$ | 0.049 | 0.057 | 0.079 | 0.107 | 0.155 |
| | $acov_2$ | 0.048 | 0.061 | 0.074 | 0.127 | 0.152 |
| | pcov | 0.052 | 0.054 | 0.055 | 0.042 | 0.046 |
| | dcov | 0.054 | 0.046 | 0.046 | 0.047 | 0.037 |
| | bcov | 0.042 | 0.066 | 0.045 | 0.047 | 0.042 |

see that when the data have no finite first order moment, acov, pcov, bcov have higher power than dcov, for which the moment condition is needed theoretically. It is also shown that all the tests are powerful for the monotone relationship (i.e. $f(x) = x^3$), and acov tests are better to recognize the non-monotone dependence relationships. In addition, it seems that the relationship $f(x) = \cos(x)$ is hard to be recognized, for which all the tests performs poorly. The powers increase as $m$ gets larger (or the dependence becomes stronger) for the angle covariance tests, but stay at about the nominal significance level for others.

**Example 2.** Consider 4 models with the formula $Y(t) = f(X(t)) + \varepsilon(t), t \in [0, 1]$, where $\varepsilon$ is generated by Wiener process, $X$ and $Y$ are the focused stochastic processes. Three processes, including Ornstein–Uhlenbeck process (OU) (a Gaussian process with mean 0 and covariance $\mathrm{E}\{X(s)X(t)\} = 3e^{-(s+t)/3}\{e^{2(s+t)/3} - 1\}, s, t \in [0, 1]$), Gaussian process (GP) with mean 0 and covariance $\mathrm{E}\{X(s)X(t)\} = 1$ when $s = t$ and 0 otherwise, $s, t \in [0, 1]$, and Gaussian process with exponential variogram (VP) (a Gaussian process with mean 0 and covariance $\mathrm{E}\{X(s)X(t)\} = e^{-5|s-t|}, s, t \in [0, 1]$)

**Table 3**
Empirical sizes and powers: Example 1(iii). $H_0$ is true when $m = 0$, and $H_1$ holds when $m > 0$. The empirical sizes of all the tests are close to the significance level and the acov tests have higher powers than other tests.

| Relationship | Method | $m = 0$ | $m = 1$ | $m = 3$ | $m = 5$ | $m = 10$ |
|---|---|---|---|---|---|---|
| $f(x) = x^3$ | $acov_1$ | 0.050 | 0.995 | 1.000 | 1.000 | 1.000 |
| | $acov_2$ | 0.053 | 0.995 | 1.000 | 1.000 | 1.000 |
| | pcov | 0.054 | 0.390 | 0.889 | 0.991 | 1.000 |
| | dcov | 0.046 | 0.392 | 0.871 | 0.980 | 1.000 |
| | bcov | 0.047 | 0.175 | 0.510 | 0.721 | 0.958 |
| $f(x) = x^2$ | $acov_1$ | 0.049 | 0.198 | 0.424 | 0.426 | 0.424 |
| | $acov_2$ | 0.049 | 0.208 | 0.411 | 0.413 | 0.408 |
| | pcov | 0.041 | 0.060 | 0.075 | 0.105 | 0.132 |
| | dcov | 0.039 | 0.068 | 0.086 | 0.136 | 0.205 |
| | bcov | 0.058 | 0.078 | 0.119 | 0.203 | 0.420 |
| $f(x) = \sin(x)$ | $acov_1$ | 0.062 | 0.969 | 1.000 | 1.000 | 1.000 |
| | $acov_2$ | 0.057 | 0.972 | 1.000 | 1.000 | 1.000 |
| | pcov | 0.050 | 0.062 | 0.127 | 0.191 | 0.499 |
| | dcov | 0.049 | 0.054 | 0.131 | 0.191 | 0.492 |
| | bcov | 0.059 | 0.052 | 0.080 | 0.104 | 0.211 |
| $f(x) = \cos(x)$ | $acov_1$ | 0.042 | 0.101 | 0.238 | 0.327 | 0.422 |
| | $acov_2$ | 0.040 | 0.099 | 0.262 | 0.336 | 0.414 |
| | pcov | 0.034 | 0.055 | 0.049 | 0.054 | 0.055 |
| | dcov | 0.036 | 0.053 | 0.045 | 0.055 | 0.065 |
| | bcov | 0.043 | 0.062 | 0.055 | 0.071 | 0.088 |

for $X$ are employed, and they are generated by `rproc2fdata` function with default parameters in `fda.usc` package (https://cran.r-project.org/web/packages/fda.usc/index.html). The models for generating $Y(t)$ are as follows:

(i) $Y(t) = \varepsilon(t)$, $t \in [0, 1]$.
(ii) $Y(t) = r\sin(X(t)) + \varepsilon(t)$, $t \in [0, 1]$, $r$ takes values 0.5, 0.25, 0.05 for GP, OU and VP respectively.
(iii) $Y(t) = re^{X(t)} + \varepsilon(t)$, $t \in [0, 1]$, $r$ takes values 0.5, 0.25, 0.05 for GP, OU and VP respectively.
(iv) $Y(t) = r\tan(X(t)) + \varepsilon(t)$, $t \in [0, 1]$, $r$ takes values 0.5, 0.5, 0.05 for GP, OU and VP respectively.

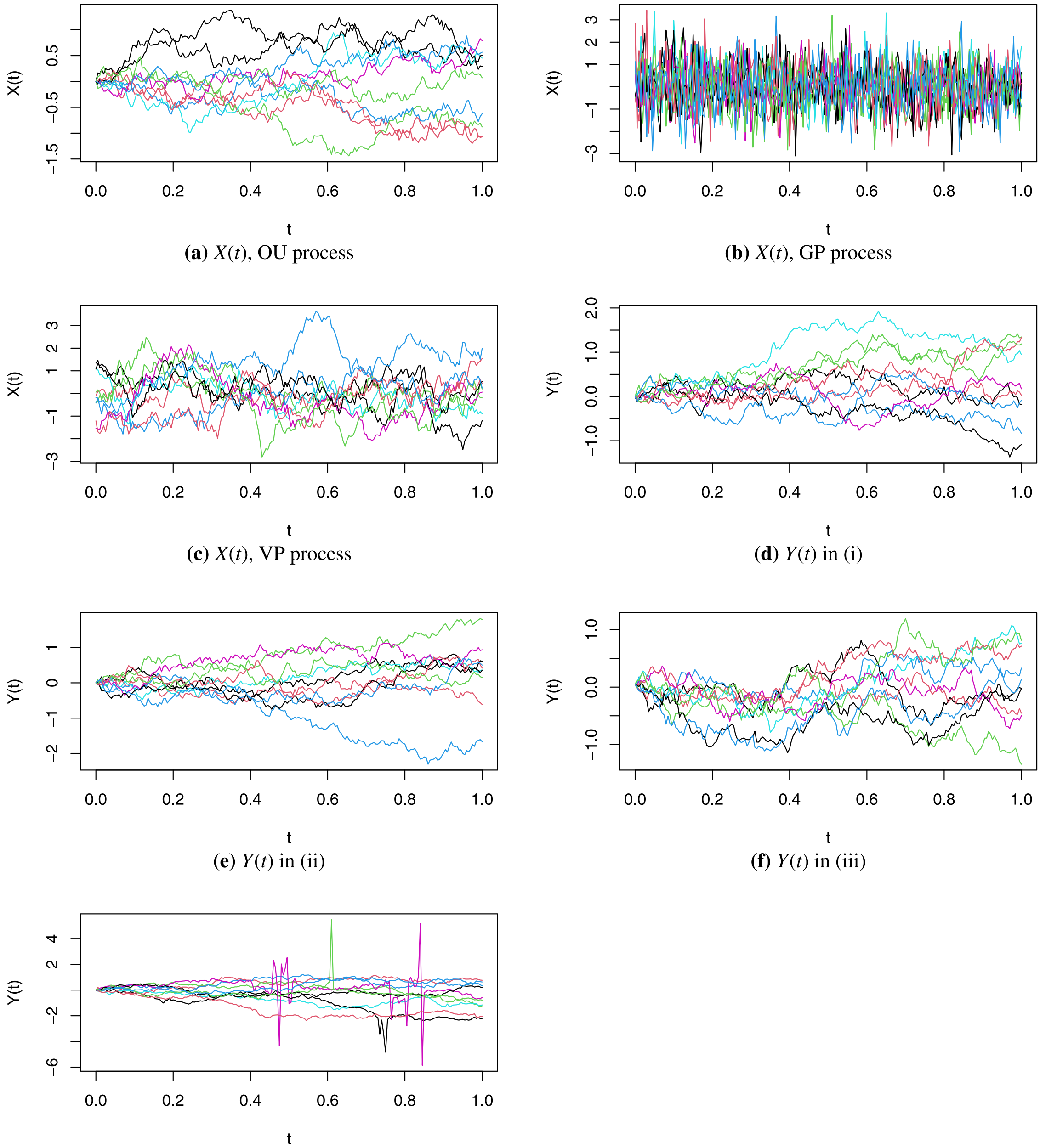
In Example 2, the sample size varies from 20 to 40, and the significance level is still 0.05. As shown in Fig. 2, not all observed trajectories reveal obvious periodic feature. In this case, we try to choose the spline basis as the eigenfunctions of the covariance operators of the Gaussian measures. Table 4 summarizes the results. The empirical sizes of all the tests approximate the significance level well. Our test performs better in power than other tests except that when the relationship functions are $\exp(\cdot)$, $\tan(\cdot)$ and the underlying process is OU. In these cases, our test is slightly worse than other tests and the differences become smaller when the sample size gets larger. However, in other cases, especially when the underlying $X$ process is GP or VP, the performance of our test is much more powerful than other tests. It should be noted that in some cases, the powers of other tests increase slowly with the sample size. For example, when the $X$ process is generated from GP or VP in Scenario (ii), the powers of pcov, docv, bcov are almost the same even though the sample size goes larger.

**Example 3.** Consider $X(t) = \sum_{i=1}^{50} \xi_i \phi_i^*(t)$, where $\phi_i^*(t) = \sqrt{2}\cos(i\pi t)$ and $\xi_i$s are independent random variables distributed as the Cauchy distribution with location zero and scale 1 or the standard normal distribution.

(i) $Y$ is generated from the normal distribution $\mathcal{N}(0, 1)$.
(ii) $Y$ is generated from the uniform distribution $\mathcal{U}(0, 1)$.
(iii) $Y = 5\langle X, \beta\rangle + \epsilon$, where $\beta(t) = \sin(t) + \cos(t)$, and the error $\epsilon$ is from the normal distribution $\mathcal{N}(0, 1)$.
(iv) $Y = \xi_1 + \xi_2^2 + \epsilon$, where $\epsilon$ is from the normal distribution $\mathcal{N}(0, 1)$.

We set the significance level as 0.05, and vary the sample size $n$ from 20 to 40 in Example 3. Similar to Example 1, we choose the Fourier basis. The results are summarized in Table 5. The empirical sizes of all the tests are close to the significance level. The tests based on the angle covariance outperform other tests. The powers of all the tests increase as the sample size gets larger. Note that pcov is slightly worse than dcov in many cases, which shows the improvement of the proposed method to the projection correlation again.

Totally, we find that the tests based on the angle covariance work well in most considered scenarios, and the powers increase rationally with the sample size. The results have shown the difference between the angle covariance test and the others. Although they are all of omnibus tests, their original working orientation are different. The pcov test is essentially for finite dimension data, and is corresponding to the choice of identity covariance operator (which is not compact in infinite dimension case) in our working frame. The distance covariance test was originally designed for finite dimension data; although it has been generalized to functional data, the moment condition should be satisfied. The ball covariance test is for data in metric space, less information on geometry is available in general.




**(a)** $X(t)$, OU process

**(b)** $X(t)$, GP process

**(c)** $X(t)$, VP process

**(d)** $Y(t)$ in (i)

**(e)** $Y(t)$ in (ii)

**(f)** $Y(t)$ in (iii)

**(g)** $Y(t)$ in (iv)

**Fig. 2.** Trajectories of $X$ and $Y$ in Example 2. The curves of $Y(t)$ in (ii), (iii) and (iv) are obtained under the condition that $X$ is OU process. Ten realizations of the given process are shown in each graph.

## 6. Discussion

In this paper, we use the angle covariance to test the independence of two random elements in Hilbert spaces, which is especially useful in functional data analysis. In the Hilbert setting, the proposed angle covariance has desirable properties, including the equivalence of zero angle covariance and the independence, the consistency of the test against all kinds of discrepancies from the independence null hypothesis. The finite sample performance shows that the proposed test is often more powerful than the tests based on other dependence measures for functional data.

The proposed test depends on the chosen Gaussian measures, or the eigenvalues and the eigenfunction basis systems of their covariance operators. We only present some experience from simulations in this paper. In general, the eigenvalues of the form $1/i^a$ with $a > 1$ works well. Note that when $a$ is large, the projection of $Q^{1/2}X$ on eigenfunction $e_i$ decrease

**Table 4**
Empirical sizes and powers: Example 2. $H_0$ holds for scenario (i), and $H_1$ holds for scenarios (ii)–(iv). OU, GP and VP stand for the Ornstein–Uhlenbeck, Gaussian, and Gaussian process with exponential variogram, respectively. The empirical sizes of all the tests are close to the significance level 0.05 and the acov tests have higher powers than other tests in most considered cases.

| Size | Method | $n = 20$ | | | $n = 30$ | | | $n = 40$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | OU | GP | VP | OU | GP | VP | OU | GP | VP |
| (i) | $acov_1$ | 0.053 | 0.053 | 0.062 | 0.051 | 0.055 | 0.047 | 0.055 | 0.043 | 0.049 |
| | $acov_2$ | 0.047 | 0.047 | 0.059 | 0.051 | 0.058 | 0.041 | 0.046 | 0.044 | 0.049 |
| | pcov | 0.050 | 0.063 | 0.044 | 0.039 | 0.052 | 0.034 | 0.061 | 0.052 | 0.035 |
| | dcov | 0.056 | 0.063 | 0.040 | 0.042 | 0.054 | 0.039 | 0.064 | 0.045 | 0.036 |
| | bcov | 0.044 | 0.049 | 0.043 | 0.041 | 0.060 | 0.042 | 0.062 | 0.042 | 0.050 |
| Power | | | | | | | | | | |
| (ii) | $acov_1$ | 0.510 | 1.000 | 0.457 | 0.711 | 1.000 | 0.761 | 0.855 | 1.000 | 0.936 |
| | $acov_2$ | 0.542 | 1.000 | 0.930 | 0.746 | 1.000 | 0.997 | 0.871 | 1.000 | 1.000 |
| | pcov | 0.278 | 0.072 | 0.047 | 0.431 | 0.067 | 0.044 | 0.581 | 0.109 | 0.055 |
| | dcov | 0.280 | 0.066 | 0.048 | 0.411 | 0.054 | 0.044 | 0.565 | 0.087 | 0.062 |
| | bcov | 0.184 | 0.059 | 0.053 | 0.248 | 0.062 | 0.051 | 0.347 | 0.069 | 0.059 |
| (iii) | $acov_1$ | 0.635 | 1.000 | 0.920 | 0.811 | 1.000 | 0.992 | 0.922 | 1.000 | 1.000 |
| | $acov_2$ | 0.615 | 1.000 | 0.992 | 0.799 | 1.000 | 1.000 | 0.908 | 1.000 | 1.000 |
| | pcov | 0.606 | 0.220 | 0.062 | 0.757 | 0.356 | 0.067 | 0.881 | 0.463 | 0.093 |
| | dcov | 0.654 | 0.183 | 0.057 | 0.805 | 0.274 | 0.057 | 0.901 | 0.335 | 0.094 |
| | bcov | 0.460 | 0.324 | 0.065 | 0.628 | 0.477 | 0.073 | 0.760 | 0.595 | 0.077 |
| (iv) | $acov_1$ | 0.692 | 0.330 | 0.541 | 0.868 | 0.388 | 0.713 | 0.957 | 0.484 | 0.799 |
| | $acov_2$ | 0.638 | 0.438 | 0.683 | 0.829 | 0.527 | 0.841 | 0.937 | 0.646 | 0.910 |
| | pcov | 0.759 | 0.086 | 0.126 | 0.908 | 0.070 | 0.164 | 0.968 | 0.082 | 0.141 |
| | dcov | 0.800 | 0.074 | 0.155 | 0.858 | 0.059 | 0.188 | 0.921 | 0.060 | 0.177 |
| | bcov | 0.888 | 0.088 | 0.360 | 0.988 | 0.101 | 0.518 | 0.998 | 0.115 | 0.635 |

**Table 5**
Empirical sizes and powers: Example 3. $H_0$ hold for scenarios (i) and (ii), and $H_1$ holds for scenarios (iii) and (iv). The 'Cauchy' and 'normal' indicate the distributions of $\xi_i$s in $X$s. The empirical sizes of all the tests are close to the significance level 0.05 and the acov tests have higher powers than other tests.

| Size | Method | $n = 20$ | | $n = 30$ | | $n = 40$ | |
|---|---|---|---|---|---|---|---|
| | | Cauchy | Normal | Cauchy | Normal | Cauchy | Normal |
| (i) | $acov_1$ | 0.044 | 0.041 | 0.045 | 0.040 | 0.047 | 0.045 |
| | $acov_2$ | 0.044 | 0.041 | 0.053 | 0.047 | 0.042 | 0.048 |
| | pcov | 0.053 | 0.044 | 0.051 | 0.043 | 0.051 | 0.050 |
| | dcov | 0.052 | 0.049 | 0.056 | 0.041 | 0.045 | 0.049 |
| | bcov | 0.052 | 0.053 | 0.057 | 0.039 | 0.043 | 0.049 |
| (ii) | $acov_1$ | 0.049 | 0.046 | 0.053 | 0.033 | 0.046 | 0.052 |
| | $acov_2$ | 0.050 | 0.051 | 0.048 | 0.035 | 0.038 | 0.056 |
| | pcov | 0.040 | 0.050 | 0.042 | 0.042 | 0.048 | 0.062 |
| | dcov | 0.047 | 0.048 | 0.048 | 0.044 | 0.046 | 0.059 |
| | bcov | 0.044 | 0.051 | 0.051 | 0.048 | 0.038 | 0.069 |
| Power | | | | | | | |
| (iii) | $acov_1$ | 0.786 | 0.351 | 0.923 | 0.561 | 0.987 | 0.696 |
| | $acov_2$ | 0.754 | 0.356 | 0.909 | 0.546 | 0.985 | 0.683 |
| | pcov | 0.233 | 0.107 | 0.377 | 0.164 | 0.516 | 0.220 |
| | dcov | 0.272 | 0.112 | 0.315 | 0.180 | 0.401 | 0.233 |
| | bcov | 0.223 | 0.088 | 0.302 | 0.113 | 0.385 | 0.102 |
| (iv) | $acov_1$ | 0.566 | 0.592 | 0.769 | 0.790 | 0.899 | 0.909 |
| | $acov_2$ | 0.593 | 0.591 | 0.786 | 0.796 | 0.910 | 0.911 |
| | pcov | 0.086 | 0.119 | 0.093 | 0.223 | 0.118 | 0.276 |
| | dcov | 0.093 | 0.150 | 0.104 | 0.228 | 0.109 | 0.282 |
| | bcov | 0.093 | 0.110 | 0.100 | 0.169 | 0.109 | 0.211 |

rapidly with $i$, and the angle obtained depends mainly on the first several coordinates of $X$. As for the eigenfunction basis, we have only tried the Fourier basis and spline basis. Both are easy to implement. Other basis, like Legendre polynomials, should be explored in the future. Furthermore, it may be desirable to provide a data-driven and powerful criterion to decide suitable nondegenerate Gaussian measures. The functional principal components and canonical correlation analysis might be useful to construct a data-driven method.

**Acknowledgments**

The research is partly supported by National Natural Science Foundation of China under Grant Nos. 11771032 and 10971045. We would like to thank the Editor, Associate Editor and the referees for constructive comments that led to a substantial improvement of the paper.

## Appendix. Proofs of the theorems

**Proof of Lemma 1.** Suppose that $X$ and $Y$ are independent. For any $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$, $\langle X, f\rangle$ and $\langle Y, g\rangle$ are measurable functions of $X$ and $Y$ respectively. Hence, $\langle X, f\rangle$ and $\langle Y, g\rangle$ are independent.

Now suppose that $\langle X, f\rangle$ and $\langle Y, g\rangle$ are independent for any $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$. Then, by Proposition 2.10 of Da Prato [2], $\mathrm{E}e^{is\langle X,f\rangle + it\langle Y,g\rangle} = \mathrm{E}e^{is\langle X,f\rangle}\mathrm{E}e^{it\langle Y,g\rangle}$ for any $f \in \mathcal{H}_1$, $g \in \mathcal{H}_2$, $s, t \in \mathbb{R}$. Let $s = t = 1$, we have $\mathrm{E}e^{i\langle X,f\rangle + i\langle Y,g\rangle} = \mathrm{E}e^{i\langle X,f\rangle}\mathrm{E}e^{i\langle Y,g\rangle}$ for any $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$. Using Proposition 2.10 of Da Prato [2] again, $X$ and $Y$ are independent. □

To prove Theorem 1, we firstly recall a representation of Gaussian measures. Suppose that $\mu$ is a nondegenerate Gaussian measure in a separable Hilbert space $\mathcal{H}$ with mean $\nu$ and covariance operator $Q$, then the characteristic functional (or the Fourier transform) has the form $\varphi_\mu(t) = \exp\{i\langle \nu, t\rangle - \frac{1}{2}\langle Qt, t\rangle\}$. $\mu$ is nondegenerate if the operator $Q$ is strictly positive. Let $\{e_k\}$ be the orthonormal eigenvectors of $Q$ and $\{\lambda_k\}$ the corresponding eigenvalues, i.e., $Qe_k = \lambda_k e_k$, for any $k \in \{1, 2, \ldots\}$. Let $\nu_k = \langle \nu, e_k\rangle$ and $\mu_k = N(\nu_k, \lambda_k)$ be the normal probability measure with mean $\nu_k$ and variance $\lambda_k$, that is,

$$\mu_k(dx) = \frac{1}{\sqrt{2\pi\lambda_k}} e^{-\frac{(x-\nu_k)^2}{2\lambda_k}} dx, x \in \mathbb{R}.$$

For any $x \in \mathcal{H}$, let $x_k = \langle x, e_k\rangle$. Then $\mathcal{H}$ is isomorphic to the Hilbert space $\ell^2$ of all sequences $(x_1, x_2, \ldots)$ of real numbers such that $\sum_{i=1}^\infty x_i^2 < \infty$. By Theorem 1.11 of Da Prato [2], $\mu = \prod_{k=1}^\infty \mu_k$, where $\prod_{k=1}^\infty \mu_k$ is countable product of measures defined through $\prod_{k=1}^\infty \mu_k(C_{m,A}) = (\mu_1 \times \cdots \times \mu_m)(A)$ for any cylindrical set $C_{m,A} = A \times \mathbb{R} \times \mathbb{R} \times \cdots$ of $\mathbb{R}^\infty$, $A$ is a Borel set of $\mathbb{R}^m$, see also e.g. Section 6.3 Ambrosio et al. [1].

Secondly, we need following Lemma 2.

**Lemma 2.** *Suppose that $\mu$ is a Gaussian measure in $\mathbb{R}^m$ with mean zero and covariance matrix $I_{m\times m}$. For any nonzero $U, V \in \mathbb{R}^m$, we have*

$$\int_{\mathbb{R}^m} I(\langle t, U\rangle \le 0) I(\langle t, V\rangle \le 0)\mu(dt) = \frac{1}{2} - \frac{1}{2\pi}\arccos\left(\frac{\langle U, V\rangle}{\|U\| \cdot \|V\|}\right).$$

**Proof.** Without loss of generality, we assume the norms of $U$ and $V$ equal to 1. We first prove the special case $m = 2$. Let $t_1 = r\cos(\theta)$, $t_2 = r\sin(\theta)$, $r \ge 0$, $0 \le \theta \le 2\pi$, $U = (\cos(\theta_1), \sin(\theta_1))$ and $V = (\cos(\theta_2), \sin(\theta_2))$. We will assume that $0 \le \theta_1 \le \theta_2 \le \pi$, otherwise we can rotate the coordinates since the standard multivariate normal distribution is invariant under rotation.

$$\begin{aligned}\int_{\mathbb{R}^2} I(\langle t, U\rangle \le 0) I(\langle t, V\rangle \le 0)\mu(dt) =& \frac{1}{2\pi}\int_{\mathbb{R}^2} I(t_1\cos\theta_1 + t_2\sin\theta_1 \le 0) I(t_1\cos\theta_2 + t_2\sin\theta_2 \le 0) e^{-\frac{t_1^2+t_2^2}{2}} dt_1 dt_2 \\ =& \frac{1}{2\pi}\int_{r\ge 0, 0\le\theta\le 2\pi} I(\cos(\theta-\theta_1) \le 0) I(\cos(\theta-\theta_2) \le 0) e^{-\frac{r^2}{2}} r dr d\theta \\ =& \frac{1}{2} - \frac{1}{2\pi}|\theta_2 - \theta_1| = \frac{1}{2} - \frac{1}{2\pi}\arccos\left(\frac{\langle U, V\rangle}{\|U\| \cdot \|V\|}\right).\end{aligned}$$

For the general case, by Proposition 3.3.2 of Vershynin [26], if $g \sim N(0, I_{m\times m})$, then $\Sigma g \sim N(0, I_{m\times m})$, where $\Sigma$ is an orthogonal matrix. Hence, we can assume that $U = (u_1, u_2, 0, \ldots, 0)$, $V = (v_1, v_2, 0, \ldots, 0)$. Therefore,

$$\begin{aligned}\int_{\mathbb{R}^m} I(\langle t, U\rangle \le 0) I(\langle t, V\rangle \le 0)\mu(dt) =& \frac{1}{2\pi}\int_{\mathbb{R}^2} I(t_1 u_1 + t_2 u_2 \le 0) I(t_1 v_1 + t_2 v_2 \le 0) e^{-\frac{t_1^2+t_2^2}{2}} dt_1 dt_2 \\ =& \frac{1}{2} - \frac{1}{2\pi}\arccos\left(\frac{\langle U, V\rangle}{\|U\| \cdot \|V\|}\right).\end{aligned}$$

The last equation follows from the previous result. We complete the proof. □

**Proof of Theorem 1.** By Theorem 1.11 of Da Prato [2], $\mu = \prod_{k=1}^\infty \mu_k$, where

$$\mu_k(dx) = \frac{1}{\sqrt{2\pi\lambda_k}} e^{-\frac{x_k^2}{2\lambda_k}} dx_k, \quad x_k = \langle x, e_k\rangle \in \mathbb{R}.$$

For any $U, V, t \in \mathcal{H}$, write $U = \sum_{k=1}^{\infty} u_k e_k$, $V = \sum_{k=1}^{\infty} v_k e_k$, $t = \sum_{k=1}^{\infty} t_k e_k$. Let $U_m = \sum_{k=1}^{m} u_k e_k$, $V_m = \sum_{k=1}^{m} v_k e_k$, $\tau_m = \sum_{k=1}^{m} t_k e_k$, $f_m(t) = I(\langle t, U_m\rangle \le 0)I(\langle t, V_m\rangle \le 0)$ and $f(t) = I(\langle t, U\rangle \le 0) I(\langle t, V\rangle \le 0)$. Then for all $t \in \mathcal{H}$, $f_m(t) \to f(t)$, as $m \to \infty$. Note that

$$\begin{aligned}\int_{\mathcal{H}} I(\langle t, U_m\rangle \le 0)I(\langle t, V_m\rangle \le 0)\mu(dt) &= \int_{\mathbb{R}^m} I(\langle \tau_m, U_m\rangle \le 0) I(\langle \tau_m, V_m\rangle \le 0) \mu_1(dt_1)\dots\mu_m(dt_m)\\ &= \int_{\mathbb{R}^m} I(t_1u_1 + \cdots + t_m u_m \le 0) I(t_1 v_1 + \cdots + t_m v_m \le 0)\mu_1(dt_1)\dots\mu_m(dt_m)\\ &= \int_{\mathbb{R}^m} I\left(\langle \tau_m', U_m \circ \lambda_m\rangle \le 0\right) I\left(\langle \tau_m', V_m \circ \lambda_m\rangle \le 0\right) \mu_1'(dt_1')\dots\mu_m'(dt_m')\\ &= \frac{1}{2} - \frac{1}{2\pi}\arccos\left(\frac{\langle U_m\circ\lambda_m, V_m\circ\lambda_m\rangle}{\|U_m\circ\lambda_m\|\|V_m\circ\lambda_m\|}\right),\end{aligned}$$

where $\tau_m' = \left(t_1/\sqrt{\lambda_1}, \dots, t_m/\sqrt{\lambda_m}\right)$, $U_m \circ \lambda_m = \left(u_1\sqrt{\lambda_1}, \dots, u_m\sqrt{\lambda_m}\right)$, $t_i' = t_i/\sqrt{\lambda}$, $\mu_i'$ are the standard normal measures. The last equality follows from Lemma 2. By the dominated convergence theorem,

$$\begin{aligned}\int_{\mathcal{H}} I(\langle t, U\rangle \le 0)I(\langle t, V\rangle \le 0)\mu(dt) &= \lim_{m\to\infty}\int_{\mathcal{H}} I(\langle t, U_m\rangle \le 0)I(\langle t, V_m\rangle \le 0)\mu(dt)\\ &= \lim_{m\to\infty}\left\{\frac{1}{2} - \frac{1}{2\pi}\arccos\left(\frac{\langle U_m\circ\lambda_m, V_m\circ\lambda_m\rangle}{\|U_m\circ\lambda_m\|\cdot\|V_m\circ\lambda_m\|}\right)\right\}\\ &= \frac{1}{2} - \frac{1}{2\pi}\arccos\left(\frac{\langle Q^{1/2}U, Q^{1/2}V\rangle}{\|Q^{1/2}U\|\cdot\|Q^{1/2}V\|}\right).\end{aligned}$$

We complete the proof. □

**Proof of Proposition 1.** (i) Note that

$$\begin{aligned}&\mathrm{E}\left\{\phi(X^1, X^2, X^3)\varphi(Y^1, Y^2, Y^3)\right\}\\ =&\mathrm{E}\Big[\left\{\theta_{Q_1}(X^1, X^2, X^3) - \theta_{Q_1}(X^4, X^2, X^3) - \theta_{Q_1}(X^1, X^5, X^3) + \theta_{Q_1}(X^4, X^5, Y^3)\right\}\\ &\quad\times\left\{\theta_{Q_2}(Y^1, Y^2, Y^3) - \theta_{Q_2}(Y^6, Y^2, Y^3) - \theta_{Q_1}(Y^1, Y^7, Y^3) + \theta_{Q_1}(Y^6, Y^7, Y^3)\right\}\Big]\\ =&\mathrm{E}\left\{\theta_{Q_1}(X^1, X^2, X^3)\theta_{Q_2}(Y^1, Y^2, Y^3)\right\} - \mathrm{E}\left\{\theta_{Q_1}(X^1, X^2, X^3)\theta_{Q_2}(Y^6, Y^2, Y^3)\right\}\\ &- \mathrm{E}\left\{\theta_{Q_1}(X^1, X^2, X^3)\theta_{Q_2}(Y^1, Y^7, Y^3)\right\} + \mathrm{E}\left\{\theta_{Q_1}(X^1, X^2, X^3)\theta_{Q_2}(Y^6, Y^7, Y^3)\right\}\\ &- \mathrm{E}\left\{\theta_{Q_1}(X^4, X^2, X^3)\theta_{Q_2}(Y^1, Y^2, Y^3)\right\} + \mathrm{E}\left\{\theta_{Q_1}(X^4, X^2, X^3)\theta_{Q_2}(Y^6, Y^2, Y^3)\right\}\\ &+ \mathrm{E}\left\{\theta_{Q_1}(X^4, X^2, X^3)\theta_{Q_2}(Y^1, Y^7, Y^3)\right\} - \mathrm{E}\left\{\theta_{Q_1}(X^4, X^2, X^3)\theta_{Q_2}(Y^6, Y^7, Y^3)\right\}\\ &- \mathrm{E}\left\{\theta_{Q_1}(X^1, X^5, X^3)\theta_{Q_2}(Y^1, Y^2, Y^3)\right\} + \mathrm{E}\left\{\theta_{Q_1}(X^1, X^5, X^3)\theta_{Q_2}(Y^6, Y^2, Y^3)\right\}\\ &+ \mathrm{E}\left\{\theta_{Q_1}(X^1, X^5, X^3)\theta_{Q_2}(Y^1, Y^7, Y^3)\right\} - \mathrm{E}\left\{\theta_{Q_1}(X^1, X^5, X^3)\theta_{Q_2}(Y^6, Y^7, Y^3)\right\}\\ &+ \mathrm{E}\left\{\theta_{Q_1}(X^4, X^5, X^3)\theta_{Q_2}(Y^1, Y^2, Y^3)\right\} - \mathrm{E}\left\{\theta_{Q_1}(X^4, X^5, X^3)\theta_{Q_2}(Y^6, Y^2, Y^3)\right\}\\ &- \mathrm{E}\left\{\theta_{Q_1}(X^4, X^5, X^3)\theta_{Q_2}(Y^1, Y^7, Y^3)\right\} + \mathrm{E}\left\{\theta_{Q_1}(X^4, X^5, X^3)\theta_{Q_2}(Y^6, Y^7, Y^3)\right\}\\ :=&J_1 - J_2 - J_3 + \cdots + J_{16}.\end{aligned}$$

Since

$$J_2 = J_3 = J_5 = J_6 = J_9 = J_{11}$$

and

$$J_4 = J_7 = J_8 = J_{10} = J_{12} = J_{13} = J_{14} = J_{15} = J_{16},$$

we have

$$\begin{aligned}J_1 - J_2 - J_3 + \cdots + J_{16} =& J_1 - 2J_2 + J_4\\ =&\mathrm{E}\left\{\theta_{Q_1}(X^1, X^2, X^3)\theta_{Q_2}(Y^1, Y^2, Y^3)\right\} - 2\mathrm{E}\left\{\theta_{Q_1}(X^1, X^2, X^3)\theta_{Q_2}(Y^6, Y^2, Y^3)\right\}\\ &+ \mathrm{E}\left\{\theta_{Q_1}(X^1, X^2, X^3)\theta_{Q_2}(Y^6, Y^7, Y^3)\right\}\\ =&\mathrm{Acov}^2(X, Y).\end{aligned}$$

(ii) Using Cauchy–Schwarz inequality, the conclusion follows.

(iii) By the definition of $\text{Acov}(X, Y)$, if $X$ and $Y$ are independent, then $\text{Acov}(X, Y) = 0$ obviously.

Now suppose that $\text{Acov}(X, Y) = 0$. By the construction of $\text{Acov}^2(X, Y)$, it follows that there exist $A \subset \mathcal{H}_1$ and $B \subset \mathcal{H}_2$ such that $\mu_1(A)\times\mu_2(B) = 1$ and for any $g \in A$ and $f \in B$, $\langle X, g\rangle$ and $\langle Y, f\rangle$ are independent. Then $\mu_1(\bar{A}) = 1$ and $\mu_2(\bar{B}) = 1$, where $\bar{A}$ and $\bar{B}$ denote the closures of $A$ and $B$ respectively. Since the Gaussian measures $\mu_1$ and $\mu_2$ are nondegenerate, by Theorem 1 and Corollary of Vakhania [24], $\bar{A} = \mathcal{H}_1$ and $\bar{B} = \mathcal{H}_2$. This means that for any $g \in \mathcal{H}_1, f \in \mathcal{H}_2$, there exists $g_n \in A$ and $f_n \in B$, such that $g_n \to g$ and $f_n \to f$ as $n \to \infty$. Since $\langle X, g_n\rangle$ and $\langle Y, f_n\rangle$ are independent, by the result on page 251 of Vakhania et al. [25] , $\langle X, g\rangle$ and $\langle Y, f\rangle$ are independent. Then by Lemma 1, $X$ and $Y$ are independent.

(iv) We only need to verify that $\theta_{Q_1}(b_1X^1 + a_1, b_1X^2 + a_1, b_1X^3 + a_1) = \theta_{Q_1}(X^1, X^2, X^3)$ and $\theta_{Q_2}(b_2Y^1 + a_2, b_2Y^2 + a_2, b_2Y^3 + a_2) = \theta_{Q_2}(Y^1, Y^2, Y^3)$. We show the first equation; the other can be verified similarly. Obviously we have $Q^{1/2}bX = bQ^{1/2}X$ for any constant $b$. Therefore,

$$\begin{aligned}\theta_{Q_1}(b_1X^1 + a_1, b_1X^2 + a_1, b_1X^3 + a_1) &= \arccos\left(\frac{\langle Q^{1/2}b_1(X^1 - X^3), Q^{1/2}b_1(X^2 - X^3)\rangle}{\|Q^{1/2}b_1(X^1 - X^3)\| \cdot \|Q^{1/2}b_1(X^2 - X^3)\|}\right)\\ &= \arccos\left(\frac{\langle Q^{1/2}(X^1 - X^3), Q^{1/2}(X^2 - X^3)\rangle}{\|Q^{1/2}(X^1 - X^3)\| \cdot \|Q^{1/2}(X^2 - X^3)\|}\right) = \theta_{Q_1}(X^1, X^2, X^3).\end{aligned}$$

The proof is complete. □

**Proof of Theorem 2.** We prove that $\text{Acov}_n^2(X, Y)$ is equal to

$$2\pi n^{-1}\sum_{k=1}^{n}\left(\int\int\{\hat{F}(\langle X_k, f\rangle, \langle Y_k, g\rangle) - \hat{F}_1(\langle X_k, f\rangle)\hat{F}_2(\langle Y_k, g\rangle)\}^2\mu_1(df)\mu_2(dg)\right).$$

Using Theorem 1, we have

$$\begin{aligned}&\int\left\{I(\langle X_i, f\rangle \le \langle X_k, f\rangle) - n^{-1}\sum_{i=1}^{n}I(\langle X_i, f\rangle \le \langle X_k, f\rangle)\right\}\left\{I(\langle X_j, f\rangle \le \langle X_k, f\rangle) - n^{-1}\sum_{i=1}^{n}I(\langle X_i, f\rangle \le \langle X_k, f\rangle)\right\}\mu_1(df)\\ &= -\frac{1}{2\pi}(a_{ijk} - a_{i\cdot k} - a_{\cdot jk} + a_{\cdot\cdot k}) = -\frac{1}{2\pi}A_{ijk}.\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}&\int\left\{I(\langle Y_i, g\rangle \le \langle Y_k, g\rangle) - n^{-1}\sum_{i=1}^{n}I(\langle Y_i, g\rangle \le \langle Y_k, g\rangle)\right\}\left\{I(\langle Y_j, g\rangle \le \langle Y_k, g\rangle) - n^{-1}\sum_{i=1}^{n}I(\langle Y_i, g\rangle \le \langle Y_k, g\rangle)\right\}\mu_2(dg)\\ &= -\frac{1}{2\pi}(b_{ijk} - b_{i\cdot k} - b_{\cdot jk} + b_{\cdot\cdot k}) = -\frac{1}{2\pi}B_{ijk}.\end{aligned}$$

The above two results yield

$$2\pi n^{-1}\sum_{k=1}^{n}\left[\int\int\left\{\hat{F}(\langle X_k, f\rangle, \langle Y_k, g\rangle) - \hat{F}_1(\langle X_k, f\rangle)\hat{F}_2(\langle Y_k, g\rangle)\right\}^2\mu_1(df)\mu_2(dg)\right] = n^{-3}\sum_{i,j,k=1}^{n}A_{ijk}B_{ijk}.$$

This completes the proof. □

**Proof of Theorem 3.** Let $(X^i, Y^i)$, $1 \le i \le 7$, be independent copies of $(X, Y)$. Denote $W = (X, Y)$, $W_i = (X^i, Y^i)$, $1 \le i \le 7$ and

$$p_Q(Z_1, Z_2, Z_3, Z_4, Z_5) = \theta_Q(Z_1, Z_2, Z_3) - \theta_Q(Z_1, Z_4, Z_3) - \theta_Q(Z_2, Z_5, Z_3) + \theta_Q(Z_4, Z_5, Z_3),$$

where, $Z_i \in \mathcal{H}_1$ or $Z_i \in \mathcal{H}_2$, $i = 1, \ldots, 5$, $Q = Q_1$ or $Q_2$ accordingly. Since $0 \le \theta \le \pi$, then

$$h\left((X^1, Y^1), \ldots, (X^7, Y^7)\right) := p_{Q_1}(X^1, X^2, X^3, X^4, X^5)p_{Q_2}(Y^1, Y^2, Y^3, Y^6, Y^7)$$

is integrable, that is, $Eh\left((X^1, Y^1), \ldots, (X^7, Y^7)\right) < \infty$. Thus, Fubini's theorem shows that

$$Eh\left((X^1, Y^1), \ldots, (X^7, Y^7)\right) = \text{Acov}^2(X, Y).$$

By definition,

$$\begin{aligned}\text{Acov}_n^2(X, Y) =& n^{-3}\sum_{i,j,k=1}^{n}A_{ijk}B_{ijk}\\ =& n^{-3}\sum_{i,j,k=1}^{n}\left[\left\{\theta_{Q_1}(X_i, X_j, X_k) - n^{-1}\sum_{l=1}^{n}\theta_{Q_1}(X_i, X_l, X_k) - n^{-1}\sum_{r=1}^{n}\theta_{Q_1}(X_r, X_j, X_k) + n^{-2}\sum_{r,l=1}^{n}\theta_{Q_1}(X_r, X_l, X_k)\right\}\right.\end{aligned}$$

$$\times \left\{\theta_{Q_2}(Y_i, Y_j, Y_k) - n^{-1}\sum_{s=1}^{n}\theta_{Q_2}(Y_i, Y_s, Y_k) - n^{-1}\sum_{t=1}^{n}\theta_{Q_2}(Y_t, Y_j, Y_k) + n^{-2}\sum_{t,s=1}^{n}\theta_{Q_2}(Y_t, Y_s, Y_k)\right\}\Bigg]$$

$$= n^{-7}\sum_{i,j,k,l,r,s,t=1}^{n}\Bigg[\left\{\theta_{Q_1}(X_i, X_j, X_k) - \theta_{Q_1}(X_i, X_l, X_k) - \theta_{Q_1}(X_r, X_j, X_k) + \theta_{Q_1}(X_r, X_l, X_k)\right\}$$

$$\times \left\{\theta_{Q_2}(Y_i, Y_j, Y_k) - \theta_{Q_2}(Y_i, Y_s, Y_k) - \theta_{Q_2}(Y_t, Y_j, Y_k) + \theta_{Q_2}(Y_t, Y_s, Y_k)\right\}\Bigg]$$

$$= n^{-7}\sum_{i,j,k,l,r,s,t=1}^{n} p_{Q_1}(X_i, X_j, X_k, X_l, X_r)p_{Q_2}(Y_i, Y_j, Y_k, Y_s, Y_t) = n^{-7}\sum_{i,j,k,l,r,s,t=1}^{n} h\left((X_i, Y_i), \dots, (X_t, Y_t)\right)$$

$$= n^{-7}\sum_{i,j,k,l,r,s,t=1}^{n} \bar{h}(W_i, \dots, W_t),$$

where $\bar{h}(W_1, \dots, W_7) = \frac{1}{7!}\sum h(W_{\pi(1)}, \dots, W_{\pi(7)})$ with the summation over all permutations $(\pi(1), \dots, \pi(7))$ of $\{1, \dots, 7\}$. Hence $\mathrm{Acov}_n^2(X, Y)$ is $V$-statistics for the kernel $\bar{h}$ of degree 7. Since $h$ is bounded by $16\pi^2$, $\mathrm{E}\bar{h}|(W_i, \dots, W_t)| < \infty$ for all $1 \le i, \dots, t \le 7$. By the strong law of large number for $V$-statistics (Theorem 3.3.1 of [9]), $\mathrm{Acov}_n^2(X, Y)$ converges to $\mathrm{Acov}^2(X, Y)$ almost surely. □

**Proof of Theorem 4.** (i) We use the same notation as in the proof of Theorem 3. Let

$$\bar{h}_1\left((x, y)\right) = \mathrm{E}\left\{\bar{h}\left((x, y), (X^2, Y^2), \dots, (X^7, Y^7)\right)\right\}, \quad \zeta_1 = \mathrm{var}(\bar{h}_1((X, Y)))$$
$$\bar{h}_2\left((x, y), (x', y')\right) = \mathrm{E}\left\{\bar{h}\left((x, y), (x', y'), (X^3, Y^3), \dots, (X^7, Y^7)\right)\right\}.$$

Under $H_0$, we have

$$\begin{aligned}\mathrm{E}\left\{h\left((X^1, Y^1), \dots, (X^7, Y^7)\right)|(X^1, Y^1)\right\} =& \mathrm{E}\Big\{\big(\theta_{Q_1}(X^1, X^2, X^3) - \theta_{Q_1}(X^1, X^4, X^3) - \theta_{Q_1}(X^2, X^5, X^3)\\ &+ \theta_{Q_1}(X^4, X^5, X^3)\big)|X^1\Big\}\\ &\times \mathrm{E}\Big\{\big(\theta_{Q_2}(Y^1, Y^2, Y^3) - \theta_{Q_2}(Y^1, Y^6, Y^3) - \theta_{Q_2}(Y^2, Y^7, Y^3)\\ &+ \theta_{Q_2}(Y^6, Y^7, Y^3)\big)|Y^1\Big\}\\ =& 0.\end{aligned}$$

Similarly, $\mathrm{E}\left\{h\left((X^1, Y^1), \dots, (X^7, Y^7)\right)|(X^i, Y^i)\right\} = 0$ for $i \in \{2, \dots, 7\}$. This means that

$$\mathrm{E}\left\{\bar{h}\left((X^1, Y^1), \dots, (X^7, Y^7)\right)|(X^1, Y^1)\right\} = 0.$$

Hence $\zeta_1 = 0$. On the other hand, we can verify that

$$\begin{aligned}\bar{h}_2\left((x, y), (x', y')\right) =& \frac{1}{21}\mathrm{E}\left\{\left(\theta_{Q_1}(x, x', X^3) - \theta_{Q_1}(x, X^4, X^3) - \theta_{Q_1}(x', X^5, X^3) + \theta_{Q_1}(X^4, X^5, X^3)\right)\right\}\\ &\times \mathrm{E}\left\{\left(\theta_{Q_2}(y, y', Y^3) - \theta_{Q_2}(y, Y^6, Y^3) - \theta_{Q_2}(y', Y^7, Y^3) + \theta_{Q_2}(Y^6, Y^7, Y^3)\right)\right\}\\ =:& \frac{1}{21}g_1(x, x')g_2(y, y').\end{aligned}$$

It can be verified that $\mathrm{var}\left\{\bar{h}_2\left((X^1, Y^1), (X^2, Y^2)\right)\right\} > 0$. To see this, it is enough to show that both $g_1(X^1, X^2)$ and $g_2(Y^1, Y^2)$ are not degenerate. It is easy to know $\mathrm{E}\left\{g_1(X^1, X^2)\right\} = 0$, $\mathrm{E}\left\{g_2(Y^1, Y^2)\right\} = 0$. We just prove $\mathrm{E}\left\{g_1^2(X^1, X^2)\right\} > 0$, and $\mathrm{E}\left\{g_2^2(Y^1, Y^2)\right\} > 0$ can be obtained similarly. For simple, we drop the $Q_1$. It can be shown with some computations that

$$\begin{aligned}\mathrm{E}\left\{g_1^2(X^1, X^2)\right\} =& \mathrm{E}_{1,2}\mathrm{E}_3^2\theta(X^1, X^2, X^3) - 2\mathrm{E}_1\mathrm{E}_{2,3}^2\theta(X^1, X^2, X^3) + \mathrm{E}_{1,2,3}^2\theta(X^1, X^2, X^3)\\ =& \mathrm{E}_{1,2}\left\{\mathrm{E}_3\theta(X^1, X^2, X^3) - \mathrm{E}_{2,3}\theta(X^1, X^2, X^3)\right\}^2 - \mathrm{E}_1\left\{\mathrm{E}_{2,3}\theta(X^1, X^2, X^3) - \mathrm{E}_{1,2,3}\theta(X^1, X^2, X^3)\right\}^2\\ =& \mathrm{E}_{1,2}\left\{\mathrm{E}_3\theta(X^1, X^2, X^3) - \mathrm{E}_{2,3}\theta(X^1, X^2, X^3)\right\}^2 - \mathrm{E}_1\left[\mathrm{E}_2\left\{\mathrm{E}_3\theta(X^1, X^2, X^3) - \mathrm{E}_{2,3}\theta(X^1, X^2, X^3)\right\}\right]^2 \ge 0\end{aligned}$$

where $\mathrm{E}_I$ means the expectation is taken over the joint distribution of the variables in set $I$. The last inequality is obtained by Jensen's inequality, that is,

$$\mathrm{E}_2\left\{\mathrm{E}_3\theta(X^1, X^2, X^3) - \mathrm{E}_{2,3}\theta(X^1, X^2, X^3)\right\}^2 \ge \left[\mathrm{E}_2\left\{E_3\theta(X^1, X^2, X^3) - \mathrm{E}_{2,3}\theta(X^1, X^2, X^3)\right\}\right]^2,$$

and equality holds if and only if $\mathrm{E}_3\theta(X^1, X^2, X^3) - \mathrm{E}_{2,3}\theta(X^1, X^2, X^3) = 0$ almost surely with respect to the distribution of $X^2$. Thus, $\mathrm{E}\left\{g_1^2(X^1, X^2)\right\} > 0$ since $\mathrm{E}_3\theta(X^1, X^2, X^3)$ is non-degenerate.

Therefore, $\mathrm{Acov}_n^2(X, Y)$ is a degenerate $V$-statistic of rank 2. Since $\bar{h}_2$ is bounded, we have $\mathrm{E}\bar{h}_2^2 < \infty$. Let $\tau_i$s be the eigenvalues of the map that sends $l \in L^2(\mathcal{H}_1 \times \mathcal{H}_2, \Lambda)$ to the function

$$l(x, y) \mapsto \int \bar{h}_2\left((x, y), (x', y')\right) l(x', y') d\Lambda(x', y'),$$

where $\Lambda$ is the joint probability measure of $X$ and $Y$, and $L^2(\mathcal{H}_1 \times \mathcal{H}_2, \Lambda)$ denotes the space of square integrable functions on $\mathcal{H}_1 \times \mathcal{H}_2$ with respect to $\Lambda$. Then, by the theory of degenerate $V$-statistics (see Theorem C.9 in the supplementary material of [13]),

$$n\mathrm{Acov}_n^2(X, Y) \xrightarrow{d} \frac{7 \times (7-1)}{2} \sum_{i=1}^{\infty} \tau_i Z_i^2,$$

as $n \to \infty$, where $Z_i$s are independent standard normal variables. Let $\gamma_i = 21\tau_i$, we have $n\mathrm{Acov}_n^2(X, Y) \xrightarrow{d} \sum_{i=1}^{\infty} \gamma_i Z_i^2$, as $n \to \infty$.

(ii) If $X$ and $Y$ are dependent, then $\mathrm{Acov}^2(X, Y) > 0$. By the standard theory of $V$-statistics (see page 212 in [17]), $n^{1/2}\{\mathrm{Acov}_n^2(X, Y) - \mathrm{Acov}^2(X, Y)\} = 7n^{-1/2} \sum_{i=1}^{n} [\bar{h}_1((X_i, Y_i)) - \mathrm{Acov}^2(X, Y)] + o_p(1)$. Since $[\bar{h}_1((X_i, Y_i)) - \mathrm{Acov}^2(X, Y)]$ are i.i.d., by central limit theorem and Slutsky's theorem, we have

$$n^{1/2}\{\mathrm{Acov}_n^2(X, Y) - \mathrm{Acov}^2(X, Y)\} \xrightarrow{d} N(0, 7^2\zeta_1),$$

where $\zeta_1 = \mathrm{E}[\bar{h}_1((X_i, Y_i)) - \mathrm{Acov}^2(X, Y)]^2$. □

## References

[1] L. Ambrosio, G. Da Prato, A. Mennucci, Introduction to Measure Theory and Integration, Springer, New York, 2012.
[2] G. Da Prato, Introduction to Stochastic Analysis and Malliavin Calculus, Springer, New York, 2014.
[3] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer, New York, 2006.
[4] A. Gretton, O. Bousquet, A. Smola, B. Schölkopf, Measuring statistical dependence with Hilbert–Schmidt norms, in: International Conference on Algorithmic Learning Theory, 2005, pp. 63–77.
[5] A. Gretton, K. Fukumizu, C.H. Teo, L. Song, B. Schölkopf, A.J. Smola, A kernel statistical test of independence, in: Advances in Neural Information Processing Systems, 2008, pp. 585–592.
[6] P. Hall, M. Hosseini-Nasab, On properties of functional principal components analysis, J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (2006) 109–126.
[7] T. Hsing, R. Eubank, Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators, John Wiley and Sons, 2015.
[8] P. Kokoszka, M. Reimherr, Introduction to Functional Data Analysis, CRC Press, 2017.
[9] V. Koroljuk, Y.V. Borovskich, Theory of U-Statistics, Springer, New York, 1994.
[10] R. Lyons, Distance covariance in metric spaces, Ann. Probab. 41 (2013) 3284–3305.
[11] J.S. Morris, Functional regression, Annu. Rev. Stat. Appl. 2 (2015) 321–359.
[12] W. Pan, X. Wang, H. Zhang, H. Zhu, J. Zhu, Ball covariance: A generic measure of dependence in Banach space, J. Amer. Statist. Assoc. 115 (2020) 307–317.
[13] N. Pfister, P. Bühlmann, B. Schölkopf, J. Peters, Kernel-based tests for joint independence, J. R. Stat. Soc. Ser. B Stat. Methodol. 80 (2018) 5–31.
[14] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer, New York, 2005.
[15] D.N. Reshef, Y.A. Reshef, H.K. Finucane, S.R. Grossman, G. McVean, P.J. Turnbaugh, E.S. Lander, M. Mitzenmacher, P.C. Sabeti, Detecting novel associations in large data sets, Science 334 (2011) 1518–1524.
[16] B. Schweizer, E.F. Wolff, et al., On nonparametric measures of dependence for random variables, Ann. Statist. 9 (1981) 879–885.
[17] J. Shao, Mathematical Statistics, second ed., Springer, New York, 2003.
[18] C. Shen, C.E. Priebe, J.T. Vogelstein, From distance correlation to multiscale graph correlation, J. Amer. Statist. Assoc. 115 (2020) 280–291.
[19] K.F. Siburg, P.A. Stoimenov, A measure of mutual complete dependence, Metrika 71 (2010) 239–251.
[20] A. Smola, A. Gretton, L. Song, B. Schölkopf, A hilbert space embedding for distributions, in: International Conference on Algorithmic Learning Theory, 2007, pp. 13–31.
[21] G.J. Székely, M.L. Rizzo, Brownian distance covariance, Ann. Appl. Stat. 3 (2009) 1236–1265.
[22] G.J. Székely, M.L. Rizzo, N.K.B. Bakirov, Measuring and testing dependence by correlation of distances, Ann. Statist. 35 (2007) 2769–2794.
[23] D. Tjøstheim, H. Otneim, B. Støve, Statistical dependence: Beyond pearson's $\rho$, 2018, arXiv preprint arXiv:1809.10455.
[24] N. Vakhania, The topological support of gaussian measure in Banach space, Nagoya Math. J. 57 (1975) 59–63.
[25] N. Vakhania, V. Tarieladze, S. Chobanyan, Probability Distributions on Banach Spaces, Springer, New York, 1987.
[26] R. Vershynin, High-Dimensional Probability: An Introduction with Applications in Data Science, Cambridge University Press, 2018.
[27] J.-L. Wang, J.-M. Chiou, H.-G. Müller, Functional data analysis, Annu. Rev. Stat. Appl. 3 (2016) 257–295.
[28] L. Zhu, K. Xu, R. Li, W. Zhong, Projection correlation between two random vectors, Biometrika 104 (2017) 829–843.