Journal of Multivariate Analysis 153 (2017) 189–210

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Multivariate nonparametric test of independence

Yanan Fan [a], Pierre Lafaye de Micheaux [a,b], Spiridon Penev [a,*], Donna Salopek [a]

[a] The University of New South Wales, Department of Statistics, Sydney, 2052 NSW, Australia

[b] CREST, ENSAI, Campus de Ker Lann, Rue Blaise Pascal-BP37203, 35172 BRUZ CEDEX, France

ARTICLE INFO

Article history:
Received 16 December 2015
Available online 12 October 2016

AMS 2000 subject classifications:
62G20
62P05

Keywords:
Central limit theorem
Empirical characteristic function
Multivariate K sample independence

ABSTRACT

The problem of testing mutual independence of p random vectors in a general setting where the dimensions of the vectors can be different and the distributions can be discrete, continuous or both is of great importance. We propose such a test which utilizes multivariate characteristic functions and is a generalization of known results. We characterize the limiting distribution of the test statistic under the null hypothesis. The limiting null distribution is approximated and the method is validated. Numerical results based on simulations are investigated and our methodology is implemented in the R package `IndependenceTests`. Power comparisons are also presented for some partial cases of our general test, where some competitive procedures exist.

1. Introduction

Very often, at certain stage of a statistical inference procedure, the question arises if a certain number p of random vectors (with any combination of component sub-vectors) can be assumed to be mutually independent. Such tests are of great importance to functional magnetic resonance imaging (fMRI) for example. In this situation, it is important to find which areas of the brain are involved in certain activities and if there are any statistical associations between these brain activities. Recovering the underlying components of these activities from a given set of linearly mixed observations can be an ill-defined problem. However, the assumption of independence among the sources provides a surprisingly powerful and effective technique for a wide range of problems in various practical domains. One of the techniques to deal with this is Independent Component Analysis (ICA). The main goal of ICA is to extract multivariate sources by using an explicit assumption about the independence of the sources. Therefore, the need for testing for multivariate independence arises naturally.

In the influential paper [1], the distance correlation is introduced as a new dependence measure between two random vectors. The empirical version of this measure is easily computable in their R package called `Energy`. This distance correlation is based on functionals of the characteristic functions so it can cover a wide variety of situations such as purely continuous, purely discrete, or mixed components of the random vectors. More details on the properties of this distance correlation, the issue of testing independence of two random vectors, and uniqueness are discussed in [29–31].

When the number of vectors p is equal to 2, the distance correlation and its latest extensions are indicative about independence and could be used to construct an independence test, too. In fact, [31, p. 2783] discusses such a test. However, when the number of vectors p is larger than 2, it is not sufficient to just look at pairs of vectors only when testing for mutual independence. Hence there is a need to generalize the test of multivariate independence for the case of $p > 2$.

* Corresponding author.
E-mail addresses: y.fan@unsw.edu.au (Y. Fan), pierre.lafaye-de-micheaux@ensai.fr (P.L. de Micheaux), s.penev@unsw.edu.au (S. Penev), dm.salopek@unsw.edu.au (D. Salopek).

http://dx.doi.org/10.1016/j.jmva.2016.09.014
0047-259X/

Such a test for $p > 2$, based on a Cramér–von Mises type functional of a process defined from the empirical characteristic functions, has been proposed earlier in [3]. However, in order to determine the asymptotic distribution of the resulting test-statistic, the authors resort to a simplifying assumption. Although they do not assume joint multivariate normality, they still assume that each of the sub-vectors is marginally normally distributed. [23] also deals with the case of arbitrary $p > 2$, but only tests for joint independence of all the *p univariate* components using only Monte Carlo approximations of the distribution of the test-statistic. The paper [19] also discusses testing independence based on empirical characteristic functions, but are only concerned with the joint independence of all the *p univariate* components. They also modify (by staying within the framework of testing joint independence of all the *p* components) their original test that uses values of the empirical characteristic functions on compact supports only. The assumption of compact support is restrictive since a test of such type would be inconsistent against virtually all possible deviations from the null hypothesis (there are counterexamples of two different distribution functions with identical characteristic functions on a compact interval (see [32])). However, considering characteristic functions or their empirical counterparts on the whole space (to avoid test inconsistency) introduces many technical difficulties (especially because of their periodicity).

The modified test of [32] has an asymptotically distribution free version for testing independence of all *p univariate* components only, but this modification can only be applied when all components of the joint distribution are *continuous*. In another development, [2] proposes a non-parametric test of independence between random *vectors* based on characterization of mutual independence defined from probabilities of half-spaces in a combinatorial formula of Möbius. Their paper, which generalizes [14], is related to our current project. They use the combinatorial formula expressed by using the cumulative distribution functions of (sub)vectors. Now our approach, by using the characteristic functions, allows us to easily accommodate continuous, discrete, and mixed components. In addition, our approach is supported by earlier works showing that goodness-of-fit and independence tests based on the empirical characteristic functions are very competitive for testing the real multinormal distribution (see, e.g., [16,22]).

As pointed out in [2], without the assumption that each sub-vector is one-dimensional with a continuous cumulative distribution function, any test of independence can no longer be distribution free. Hence both tests in [2] and in the current paper, are naturally not distribution free. The paper [2] deals with this issue by computing bootstrap approximations whereas we evaluate the asymptotic distribution of our test statistic.

With the recent advances of copulas in statistical applications, another focus of a stream of papers has been in deriving tests of independence among random vectors based on Cramér–von Mises functionals of the empirical copula process. The paper [13] studies the limiting distribution of such statistics under contiguous sequences of alternatives and analyzes asymptotic relative efficiencies in some classes of copula alternatives. A summary of the efforts in this direction, as well as further references can be seen in [20,25]. The copula approach is tempting to apply in this setting on ideological grounds. Indeed, the copula function is meant to precisely "extract" the dependence structure by "leaving aside" the marginals thus making a copula-based approach very suitable to testing multivariate independence. Both papers utilize the same empirical process based on the characterization of stochastic vectorial independence in terms of copulas to construct the test-statistic. However, this approach is limited only to *continuous* random vectors. Besides, their asymptotic distribution arising from the testing procedure, is even further restricted by the requirement that the copula has continuous partial derivatives. They also apply bootstrap approximations which may be more computationally intensive than our proposed procedures. For the above reasons we do not consider further comparison of our methods with copula-based approaches in this paper.

Our paper proposes a test in the most general form. That is, our test statistic is based on a Cramér–von Mises type functional of a process defined from the empirical characteristic functions, does not need any of the restrictive assumptions such as in [3] and treats the case when $p \geq 2$, where each of the p components can itself be a vector of arbitrary length $q_i, i = 1, \ldots, p$. We offer a theory about the asymptotic distribution of our characteristic functions-based test statistic as opposed to the use of the bootstrap approach as in [2]. This allows us to determine the p-values for our test and are able to reduce the computational time in some cases without compromising the ability of the test to keep the correct level of significance asymptotically (see Section 6). Our paper, by stating the asymptotic distribution of the test statistic, represents a generalization of [1] and of [28]. We note that in [1], (for the case of $p = 2$ only), a bound of the control of the first type error of the test is given. This bound is fast to compute but being a bound only, may not be accurate enough. Further, we solve the issues with numerical calculation of our test statistic via its decomposition into feasible components and calculate critical values numerically based on asymptotic approximations. We also have performed extensive investigation of the power of our test. We have included an example of the case $p \geq 3$ where, to the level of generality considered, the only competitor to our test is the bootstrap-based procedure from the program `dependogram` in [2], and have demonstrated the favourable performance of the new test in terms of computational time and ability to keep the size close to the nominal level. For the case of $p = 2$ we have compared our test to the tests of [2,15,28] and have demonstrated the favourable performance of our test in comparison to all competitors. We also offer a variety of weight functions thus further extending the applicability of our testing procedure.

The paper is organized as follows. In Section 2, we define our test statistic and in Section 3, we study its asymptotic distribution. We discuss the normalization of our test statistic in Section 4 and in Section 5, we propose and investigate several choices of weight functions in our construction of the test statistics. In Section 6, we discuss the numerical implementation. We also demonstrate the favourable performance of our test both under the null hypothesis and under the alternative, on specific simulated examples, by comparing it to the tests from [15,28]. Section 7 contains the proofs.

2. Notation and setting

Let $\underset{\sim}{\mathbf{X}}_1, \ldots, \underset{\sim}{\mathbf{X}}_n$ be a random sample representing n independent copies of a vector

$$\underset{\sim}{\mathbf{X}} = \begin{pmatrix} \mathbf{X}^1 \\ \vdots \\ \mathbf{X}^p \end{pmatrix} \in \mathbb{R}^{q_1} \times \cdots \times \mathbb{R}^{q_p}.$$

Accordingly, $\underset{\sim}{\boldsymbol{t}} = (\boldsymbol{t}_1{}^\top, \ldots, \boldsymbol{t}_p{}^\top)^\top \in \mathbb{R}^{q_1} \times \cdots \times \mathbb{R}^{q_p}$ is a column vector of length $q := \sum_{\ell=1}^p q_\ell$, $q_\ell \geq 1$, where $^\top$ denotes the transposition operator. The purpose is to test the hypothesis of mutual independence of the p sub-components of $\underset{\sim}{\mathbf{X}}$. Suppose $\varphi(\underset{\sim}{\boldsymbol{t}})$ is the true (unknown) characteristic function of $\underset{\sim}{\mathbf{X}}$ at $\underset{\sim}{\boldsymbol{t}}$ and $\varphi_\ell(\boldsymbol{t}_\ell)$ the true (unknown) characteristic function of $\mathbf{X}^\ell$ at $\boldsymbol{t}_\ell$, $\ell = 1, \ldots, p$. Then the null hypothesis is

$$\mathcal{H}_0 : \varphi(\underset{\sim}{\boldsymbol{t}}) = \prod_{\ell=1}^p \varphi_\ell(\boldsymbol{t}_\ell), \quad \forall_{\underset{\sim}{\boldsymbol{t}} \in \mathbb{R}^{q_1} \times \cdots \times \mathbb{R}^{q_p}}. \tag{1}$$

The empirical estimator of $\varphi(\underset{\sim}{\boldsymbol{t}})$ is

$$\hat{\varphi}_n(\underset{\sim}{\boldsymbol{t}}) = n^{-1} \sum_{j=1}^n e^{i\underset{\sim}{\boldsymbol{t}}^\top \underset{\sim}{\mathbf{X}}_j} = n^{-1} \sum_{j=1}^n \prod_{\ell=1}^p e^{i\boldsymbol{t}_\ell{}^\top \mathbf{X}_j^\ell},$$

where i is the complex number such that $i^2 = -1$. Similarly, the empirical estimator of $\varphi_\ell(\boldsymbol{t}_\ell)$ is $\hat{\varphi}_{n,\ell}(\boldsymbol{t}_\ell) = n^{-1} \sum_{j=1}^n e^{i\boldsymbol{t}_\ell{}^\top \mathbf{X}_j^\ell}$.

We denote

$$\tilde{\varphi}_n(\underset{\sim}{\boldsymbol{t}}) = \prod_{\ell=1}^p \hat{\varphi}_{n,\ell}(\boldsymbol{t}_\ell).$$

Further, for a suitably chosen non-negative weight function $w(\underset{\sim}{\boldsymbol{t}})$, we denote the weighted L_2 norm of the complex-valued function $f(\cdot)$ by $\|f\|_w^2 = \int_{\mathbb{R}^q} |f(\underset{\sim}{\boldsymbol{t}})|^2 w(\underset{\sim}{\boldsymbol{t}})\, d\underset{\sim}{\boldsymbol{t}}$. The proposed family of test statistics for testing $\mathcal{H}_0$ is in the form

$$n\mathrm{T}_n(w) = \|\sqrt{n}\mathrm{D}_n\|_w^2 = n \int_{\mathbb{R}^{q_1} \times \cdots \times \mathbb{R}^{q_p}} |\mathrm{D}_n(\underset{\sim}{\boldsymbol{t}})|^2 w(\underset{\sim}{\boldsymbol{t}})\, d\underset{\sim}{\boldsymbol{t}}, \tag{2}$$

where

$$\mathrm{D}_n(\underset{\sim}{\boldsymbol{t}}) = \hat{\varphi}_n(\underset{\sim}{\boldsymbol{t}}) - \tilde{\varphi}_n(\underset{\sim}{\boldsymbol{t}})$$

and the weight functions $w(\cdot)$ satisfying

$$w(\underset{\sim}{\boldsymbol{t}}) = \prod_{\ell=1}^p v(\boldsymbol{t}_\ell).$$

Here $v(\boldsymbol{t}_\ell) = v(-\boldsymbol{t}_\ell)$, $\ell = 1, 2, \ldots, p$ are symmetric, non-negative, and continuous. We consider the cases when $\int_{\mathbb{R}^{q_\ell}} v(\boldsymbol{t}_\ell) d\boldsymbol{t}_\ell = 1$ and cases in which the latter integral is in fact infinity. Certain choices of $v(\cdot)$ lead to simpler expressions and we discuss them in Section 5.

By introducing the notation $f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell) = 1 - e^{i\boldsymbol{t}_\ell{}^\top \mathbf{X}_j^\ell}$ and $g_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell) = e^{i\boldsymbol{t}_\ell{}^\top \mathbf{X}_j^\ell}$, we can rewrite $\hat{\varphi}_{n,\ell}(\boldsymbol{t}_\ell)$ as follows:

$$\hat{\varphi}_{n,\ell}(\boldsymbol{t}_\ell) = n^{-1} \sum_{j=1}^n g_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell) = 1 - n^{-1} \sum_{j=1}^n f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell).$$

We recall the well-known multinomial formula. Let $A = \{k_1, \ldots, k_{|A|}\} \neq \emptyset$ be any finite set with cardinality $|A|$. If $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{|A|}$, with the notation $\boldsymbol{u} = (u_1, \ldots, u_{|A|})^\top := (u^{(k_1)}, \ldots, u^{(k_{|A|})})^\top$, then

$$\sum_{B \subset A} \left(\prod_{i \in B} u^{(i)} \right) \left(\prod_{j \in A \setminus B} v^{(j)} \right) = \prod_{i \in A} \left(u^{(i)} + v^{(i)} \right). \tag{3}$$

Using (3), we can write the following

$$\hat{\varphi}_n(\underset{\sim}{\boldsymbol{t}}) = n^{-1} \sum_{j=1}^n \prod_{\ell=1}^p g_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell) = n^{-1} \sum_{j=1}^n \sum_{B \subset I_p} (-1)^{|B|} \prod_{\ell \in B} f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell), \tag{4}$$

and

$$\tilde{\varphi}_n(\underset{\sim}{t}) = \prod_{\ell=1}^{p}\left(n^{-1}\sum_{j=1}^{n} g_{\mathbf{X}_j^\ell}(t_\ell)\right) = \sum_{B\subset I_p}(-1)^{|B|}\prod_{\ell\in B}\left(n^{-1}\sum_{j=1}^{n} f_{\mathbf{X}_j^\ell}(t_\ell)\right). \tag{5}$$

Here $I_p = \{1,\dots,p\}$ and the summation is calculated over all $B \subset I_p$.

Now, after some algebraic transformations, we obtain an explicit expression to calculate (2). We formulate the outcome as our first Lemma.

Lemma 1. (a) *Without any further restrictions on the weight functions we have*

$$\begin{aligned} nT_n(w) = n\sum_{B\subset I_p;B\neq\emptyset}\sum_{B'\subset I_p;B'\neq\emptyset}(-1)^{|B|+|B'|}\Bigg[&n^{-2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\prod_{\ell\in B\cap B'}\beta_{j,j',\ell}\prod_{\ell\in B\setminus B'}\gamma_{j,\ell}\prod_{\ell\in B'\setminus B}\gamma_{j',\ell}\\ &-2\left\{n^{-1}\sum_{j=1}^{n}\prod_{\ell\in B\cap B'}\left(n^{-1}\sum_{j'=1}^{n}\beta_{j,j',\ell}\right)\prod_{\ell\in B\setminus B'}\gamma_{j,\ell}\prod_{\ell\in B'\setminus B}\left(n^{-1}\sum_{j'=1}^{n}\gamma_{j',\ell}\right)\right\}\\ &+\prod_{\ell\in B\cap B'}\left(n^{-2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\beta_{j,j',\ell}\right)\prod_{\ell\in B\setminus B'}\left(n^{-1}\sum_{j'=1}^{n}\gamma_{j',\ell}\right)\prod_{\ell\in B'\setminus B}\left(n^{-1}\sum_{j'=1}^{n}\gamma_{j',\ell}\right)\Bigg]. \end{aligned}$$

Here

$$\gamma_{j,\ell} := \int_{\mathbb{R}^{q_\ell}} f_{\mathbf{X}_j^\ell}(t_\ell)v(t_\ell)dt_\ell = \int_{\mathbb{R}^{q_\ell}} f_{\mathbf{X}_j^\ell}(-t_\ell)v(t_\ell)dt_\ell \in \mathbb{R}$$

and

$$\beta_{j,j',\ell} = \beta_{j',j,\ell} := \gamma_{j,\ell}+\gamma_{j',\ell}-\gamma_{j,j',\ell} = \int_{\mathbb{R}^{q_\ell}} f_{\mathbf{X}_j^\ell}(t_\ell)f_{\mathbf{X}_{j'}^\ell}(-t_\ell)v(t_\ell)dt_\ell \in \mathbb{R}.$$

Note that the calculation of the term $\gamma_{j,j',\ell} := \int_{\mathbb{R}_{q_\ell}} f_{\mathbf{X}_j^\ell-\mathbf{X}_{j'}^\ell}(t_\ell)v(t_\ell)dt_\ell \in \mathbb{R}$ *involves the same pattern like the calculation of* $\gamma_{j,\ell}$. *In addition, when* $j=j'$, $\beta_{j,j',\ell}=2\gamma_{j,\ell}$.

(b) *When all integrals* $\int_{\mathbb{R}^{q_\ell}} v(t_\ell)dt_\ell = 1$, $\ell = 1,\dots,p$, *we obtain a simplified version of* (a)*:*

$$nT_n(w) = n^{-1}\sum_{j=1}^{n}\sum_{j'=1}^{n}\prod_{\ell=1}^{p}\xi_{j,j',\ell} - 2\sum_{j=1}^{n}\prod_{\ell=1}^{p}\left(n^{-1}\sum_{j'=1}^{n}\xi_{j,j',\ell}\right) + n\prod_{\ell=1}^{p}\left(n^{-2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\xi_{j,j',\ell}\right),$$

where $\xi_{j,j',\ell} = \int_{\mathbb{R}^{q_\ell}}\cos\left(t_\ell{}^{\top}\left(\mathbf{X}_j^\ell-\mathbf{X}_{j'}^\ell\right)\right)v(t_\ell)dt_\ell \in \mathbb{R}$.

Note that the formula in (a) is more general. In particular, it covers the cases of heavy-tailed weight functions whose integrals are infinite, like the weights considered in [1,31].

3. Asymptotic distribution of the test statistic

In this section, we study the asymptotic distribution of our test-statistic. Before we proceed, we need the following notation and condition.

Let $\psi(\underset{\sim}{t}) = (1-Re[\varphi(\underset{\sim}{t})])^{1/2}$, $\|\underset{\sim}{t}\| = \max(|t_1|,\dots,|t_q|)$ and let μ_q stands for the q-dimensional Lebesgue measure. Suppose $\bar{\varphi}(h) = \sup\{z : 0\le z\le 1, \mu_q\{\underset{\sim}{t} : \|\underset{\sim}{t}\| < 1/2, \psi(\underset{\sim}{t}) < z\} < h\}$. Then assume that the condition

$$\int_0^1 \frac{\bar{\varphi}(h)}{h(\log\frac{1}{h})^{1/2}}dh < \infty \tag{6}$$

is satisfied. Note that this condition is necessary and sufficient for $\sqrt{n}\left(\hat{\varphi}_n-\varphi\right)$ to converge weakly to some Gaussian process on every compact set (see [5, Theorem 3.1]).

Theorem 1. *Suppose condition* (6) *holds. If* $\mathbf{X}^1,\dots,\mathbf{X}^p$ *are mutually independent, then the process* $\sqrt{n}D_n$ *converges in the space* $C(\mathbb{R}^q,\mathbb{C})$ *of continuous functions from* $\mathbb{R}^q$ *to* $\mathbb{C}$, *to a zero mean complex valued Gaussian process* D *having covariance function given by*

$$\begin{aligned} C(\underset{\sim}{s},\underset{\sim}{t}) &:= \mathbb{E}\left\{D(\underset{\sim}{s})\overline{D(\underset{\sim}{t})}\right\}\\ &= \prod_{\ell=1}^{p}\varphi_\ell(-t_\ell+s_\ell) - \prod_{\ell=1}^{p}\{\varphi_\ell(-t_\ell)\varphi_\ell(s_\ell)\} - \sum_{\ell=1}^{p}\{\varphi_\ell(-t_\ell+s_\ell)-\varphi_\ell(-t_\ell)\varphi_\ell(s_\ell)\}\prod_{j=1;j\neq\ell}^{p}\varphi_j(-t_j)\varphi_j(s_j). \end{aligned}$$

After some calculations, the expression for the covariance function can be simplified to

$$\prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)-\prod_{\ell=1}^{p}\{\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\}\left\{1-p+\sum_{\ell=1}^{p}\frac{\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)}{\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)}\right\}.$$

The pseudo-covariance function is given by

$$\begin{aligned}P(\underset{\sim}{\boldsymbol{s}},\underset{\sim}{\boldsymbol{t}}) &:= \mathbb{E}\{\mathrm{D}(\underset{\sim}{\boldsymbol{s}})\mathrm{D}(\underset{\sim}{\boldsymbol{t}})\}\\ &= \prod_{\ell=1}^{p}\varphi_\ell(\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)-\prod_{\ell=1}^{p}\{\varphi_\ell(\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\}-\sum_{j=1}^{p}\left\{\varphi_j(\boldsymbol{t}_j+\boldsymbol{s}_j)-\varphi_j(\boldsymbol{t}_j)\varphi_j(\boldsymbol{s}_j)\right\}\prod_{\ell=1;\ell\neq j}^{p}\varphi_\ell(\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell).\end{aligned}$$

Now we are in a position to formulate the asymptotic distribution of the test-statistic under the null hypothesis of independence.

Theorem 2. *Assume that*

$$\int_{\mathbb{R}^{q_\ell}}|\varphi_\ell(\boldsymbol{x}_\ell+\boldsymbol{y}_\ell)|\,v(\boldsymbol{x}_\ell)d\boldsymbol{x}_\ell<\infty \tag{7}$$

for all $\ell=1,\ldots,p$ *and all* $\boldsymbol{y}_\ell\in\mathbb{R}^{q_\ell}$ *holds. Further on, assume that*

$$\int_{\mathbb{R}^q}\int_{\mathbb{R}^q}|C(\underset{\sim}{\boldsymbol{s}},\underset{\sim}{\boldsymbol{t}})|^2\,w(\underset{\sim}{\boldsymbol{s}})w(\underset{\sim}{\boldsymbol{t}})d\underset{\sim}{\boldsymbol{s}}d\underset{\sim}{\boldsymbol{t}}<\infty \tag{8}$$

holds. Then the statistic $nT_n(w)=\|\sqrt{n}D_n\|_w^2$ *converges in distribution, under* $\mathcal{H}_0$*, to*

$$nT_n(w)=\int_{\mathbb{R}^{q_1}\times\cdots\times\mathbb{R}^{q_p}}\left|\sqrt{n}D_n(\underset{\sim}{\boldsymbol{t}})\right|^2 w(\underset{\sim}{\boldsymbol{t}})\,d\underset{\sim}{\boldsymbol{t}}\xrightarrow{\mathcal{L}}\sum_{k=1}^{\infty}\frac{\lambda_k}{2}\{(1+|p_k|)\xi_k+(1-|p_k|)\eta_k\}. \tag{9}$$

Here $\xi_k,\eta_k,k=1,2,\ldots$ *are independent pairs of independent chi-square random variables, each with one degree of freedom and* λ_k *(respectively* f_k*) are the eigenvalues, possibly of multiplicity larger than* 1 *(respectively eigenfunctions) of the integral operator* $\boldsymbol{K}(\cdot)(\cdot)$ *defined by*

$$\boldsymbol{K}(f)(\underset{\sim}{\boldsymbol{x}})=\int_{\mathbb{R}^q}f(\underset{\sim}{\boldsymbol{y}})C(\underset{\sim}{\boldsymbol{x}},\underset{\sim}{\boldsymbol{y}})w(\underset{\sim}{\boldsymbol{y}})d\underset{\sim}{\boldsymbol{y}}, \tag{10}$$

i.e., λ_k *and* f_k *are the solutions of the equation*

$$\lambda f(\underset{\sim}{\boldsymbol{x}})=\int_{\mathbb{R}^q}f(\underset{\sim}{\boldsymbol{y}})C(\underset{\sim}{\boldsymbol{x}},\underset{\sim}{\boldsymbol{y}})w(\underset{\sim}{\boldsymbol{y}})d\underset{\sim}{\boldsymbol{y}}. \tag{11}$$

The values p_k *are defined as*

$$p_k=\int_{\mathbb{R}^q}f_k(\underset{\sim}{\boldsymbol{s}})f_k(-\underset{\sim}{\boldsymbol{s}})w(\underset{\sim}{\boldsymbol{s}})d\underset{\sim}{\boldsymbol{s}}.$$

If the multiplicity of λ_k *is equal to one, then* $|p_k|=1$*. If the multiplicity of* λ_k *is equal to certain* $m>1$*, then one of the corresponding* $|p_k|$ *values is equal to one, and the remaining* $(m-1)$ *values are equal to* 0*. That is, each summand in the right hand side of* (9) *is either in the form* $\lambda_k\xi_k$ *or in the form* $\frac{\lambda_k}{2}\zeta_k$ *with* ξ_k *being chi-square distributed with one degree of freedom and* ζ_k *being chi-square distributed with two degrees of freedom.*

Of course, we only require (7) for the case of non-integrable weights since otherwise it is automatically satisfied. Condition (8), stating that $C\in L^2(\mathbb{R}^q\times\mathbb{R}^q,\mathcal{B}(\mathbb{R}^q)\times\mathcal{B}(\mathbb{R}^q),w\times w)$, ensures that $\mathbf{K}$ is a compact operator [4, Proposition 4.7, p.43]. For compact normal operators (which is the case for $\mathbf{K}$), algebraic and geometric multiplicities are equal and finite (see, e.g., [21, p. 129]).

Remark 1. Analysis of the proof of Theorems 1 and 2 in Section 7 gives a clear indication that our proposed test is expected to be universally consistent for all fixed alternatives. The reason is that the property $\mathbb{E}[\mathrm{D}_n(\underset{\sim}{\boldsymbol{t}})]=0$ will always be violated when the null hypothesis of independence is violated. Hence under the alternative, the statistic $nT_n(w)$ will contain a deterministic component that will tend to infinity as $n\to\infty$.

In practice, we do not know the true function $C(\underset{\sim}{\boldsymbol{s}},\underset{\sim}{\boldsymbol{t}})$. However, we show in Theorem 4 that the following

$$\hat{C}_n(\underset{\sim}{\boldsymbol{s}},\underset{\sim}{\boldsymbol{t}})=\prod_{\ell=1}^{p}\hat{\varphi}_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)-\left\{\prod_{\ell=1}^{p}\hat{\varphi}_\ell(-\boldsymbol{t}_\ell)\hat{\varphi}_\ell(\boldsymbol{s}_\ell)\right\}\left\{1-p+\sum_{\ell=1}^{p}\frac{\hat{\varphi}_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)}{\hat{\varphi}_\ell(-\boldsymbol{t}_\ell)\hat{\varphi}_\ell(\boldsymbol{s}_\ell)}\right\} \tag{12}$$

is a suitable estimate to use instead. By replacing this estimated covariance function in (11), we can find the approximate solution of the eigenvalues and eigenfunctions. At this stage, there are at least two options available. One option is Monte Carlo. Assume that we have N i.i.d. vectors $\underset{\sim}{\mathbf{Y}}_1, \ldots, \underset{\sim}{\mathbf{Y}}_N$ with density w (these can be simulated given that the density w is known). The law of large numbers and the extended continuous mapping theorem (see [33, Theorem 1.11.1]) imply that

$$\frac{1}{N}\sum_{j=1}^{N} f(\underset{\sim}{\mathbf{Y}}_j)C(\underset{\sim}{\boldsymbol{x}}, \underset{\sim}{\mathbf{Y}}_j) \xrightarrow{P} \int_{\mathbb{R}^q} f(\underset{\sim}{\boldsymbol{y}})C(\underset{\sim}{\boldsymbol{x}}, \underset{\sim}{\boldsymbol{y}})w(\underset{\sim}{\boldsymbol{y}})d\underset{\sim}{\boldsymbol{y}}. \tag{13}$$

On the other hand, for a given suitably designed (adapted to w) cubature rule, we can write

$$\sum_{j=1}^{N} \omega_j f(\underset{\sim}{\boldsymbol{y}}_j)C(\underset{\sim}{\boldsymbol{x}}, \underset{\sim}{\boldsymbol{y}}_j) \underset{N\to\infty}{\longrightarrow} \int_{\mathbb{R}^q} f(\underset{\sim}{\boldsymbol{y}})C(\underset{\sim}{\boldsymbol{x}}, \underset{\sim}{\boldsymbol{y}})w(\underset{\sim}{\boldsymbol{y}})d\underset{\sim}{\boldsymbol{y}} = \lambda f(\underset{\sim}{\boldsymbol{x}}), \tag{14}$$

where the ω_j's are the weights of the cubature and the $\underset{\sim}{\boldsymbol{y}}_j$'s are (fixed) nodes of the cubature. We note that if $(\omega_j, \underset{\sim}{\boldsymbol{y}}_j)$ coincide with $(1/N, \underset{\sim}{\mathbf{Y}}_j)$, then the Monte-Carlo approximation (13) is a particular cubature rule (14). For the Monte Carlo case, we have the approximation

$$\sum_{j=1}^{N} \sqrt{\omega_j} f(\underset{\sim}{\mathbf{Y}}_j)\sqrt{\omega_i}C(\underset{\sim}{\mathbf{Y}}_i, \underset{\sim}{\mathbf{Y}}_j)\sqrt{\omega_j} = \lambda\sqrt{\omega_i} f(\underset{\sim}{\mathbf{Y}}_i) + o_p(1) \quad i = 1, \ldots, N. \tag{15}$$

The $o_P(1)$ term above indicates quantities converging in probability towards 0 when N goes to infinity. Let $g_j = \sqrt{\omega_j} f(\underset{\sim}{\mathbf{Y}}_j) = \sqrt{\omega_j} f_j, j = 1, \ldots, N$ and $\boldsymbol{g} = (g_1, \ldots, g_N)^\top$. The system (15) can be presented in matrix form as follows:

$$\begin{pmatrix} \sqrt{\omega_1} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \sqrt{\omega_N} \end{pmatrix} \Gamma \begin{pmatrix} \sqrt{\omega_1} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \sqrt{\omega_N} \end{pmatrix} \boldsymbol{g} = \lambda \boldsymbol{g} + \boldsymbol{o}_P(1). \tag{16}$$

Here Γ is the $N \times N$ random matrix $\left\{C(\underset{\sim}{\mathbf{Y}}_i, \underset{\sim}{\mathbf{Y}}_j)\right\}_{1\le i,j\le N}$. Again, in a practical situation, we do not know $C(\underset{\sim}{\boldsymbol{s}}, \underset{\sim}{\boldsymbol{t}})$ which we need to calculate the above values $C(\underset{\sim}{\mathbf{Y}}_i, \underset{\sim}{\mathbf{Y}}_j), i = 1, \ldots, N, j = 1, \ldots, N$. The consistent estimator $\hat{C}_n(\underset{\sim}{\boldsymbol{s}}, \underset{\sim}{\boldsymbol{t}})$ from (12) will be used instead. Then we will get an empirical counterpart of (16) in the form

$$\begin{pmatrix} \sqrt{\omega_1} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \sqrt{\omega_N} \end{pmatrix} \hat{\Gamma}_n \begin{pmatrix} \sqrt{\omega_1} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \sqrt{\omega_N} \end{pmatrix} \boldsymbol{g} = \lambda \boldsymbol{g} + \boldsymbol{o}_P(1). \tag{17}$$

The system (17) allows us to get estimated eigenvalues $\hat{\lambda}_{k,n,N}$ and corresponding discretized eigenvectors $\hat{\boldsymbol{g}}_{k,n,N}, k = 1, \ldots, N$. From these, the estimated eigenvectors $\hat{\boldsymbol{f}}_{k,n,N}, k = 1, \ldots, N$, can be obtained. The final eigenfunctions will then be defined as

$$\hat{f}_{k,n,N}(\underset{\sim}{\boldsymbol{s}}) = \frac{1}{\hat{\lambda}_{k,n,N}} \frac{1}{N} \sum_{j=1}^{N} \hat{C}_n(\underset{\sim}{\boldsymbol{s}}, \underset{\sim}{\mathbf{Y}}_j)\hat{f}^j_{k,n,N}$$

where $\hat{f}^j_{k,n,N}$ denotes the jth component of $\hat{\boldsymbol{f}}_{k,n,N}$.

The cubature option would offer significant saving in computations. In this case, a smaller N can be chosen in the approximation (14) and a suitable possibly not-too-dense mesh of points $\underset{\sim}{\boldsymbol{x}}$ could be chosen when trying to approximately solve (11) by discretizing it. However, it is difficult to give safe general recommendations for constructing the mesh, as well as for the choice of the cubature rule with its weights and nodes. The Monte Carlo method, on the other hand, works quite generally and, for a higher computational cost, delivers reasonable results when N is large enough.

We also notice that when the weights are not integrable, we do not have a density w to simulate from when using the Monte Carlo approach. However importance sampling-based techniques can be used to implement Monte Carlo in this situation. With the above constructions for the Monte Carlo approach, we have the following results.

Theorem 3. *Let $N \in \mathbb{N}\setminus\{0\}$ be a fixed integer. Let $A_n = (a_{ij,n})$ be a sequence of $N \times N$ random complex matrices defined on some sample space Ω, with (possibly complex) eigenvalues $\lambda_{(1),n}, \ldots, \lambda_{(N),n}$, accounting for multiplicity, ordered by the increasing value of their modulus. Let $A = (a_{ij})$ be a $N \times N$ positive definite random complex matrix defined on Ω with (real positive) eigenvalues $0 < \lambda_{(1)} \le \cdots \le \lambda_{(N)}$, accounting for multiplicity. Suppose $\boldsymbol{f}_{k,n}$ and $\boldsymbol{f}_k$, $k = 1, \ldots, N$ are the associated eigenvectors. If A_n converges almost surely towards A, i.e.*

$$\forall_{i=1,\ldots,N},\ \forall_{j=1,\ldots,N,}\ a_{ij,n} \xrightarrow{a.s.} a_{ij}, \quad \text{when } n \to \infty,$$

then

$$\lambda_{(k),n} \xrightarrow{a.s.} \lambda_{(k)}, \quad \forall_{k=1,\ldots,N}$$

and, if A_n is positive definite for large values of n, then

$$\|(I_N - P_{A,k})\boldsymbol{f}_{k,n}\|_2 \xrightarrow{a.s.} 0, \quad \forall_{k=1,\ldots,N},$$

where $P_{A,k}$ is the orthogonal projector on the space spanned by all the orthonormal eigenvectors associated with λ_k.

This theorem is a general result which is interesting in its own right. We require it to improve the accuracy when we approximate the limiting distribution of the statistic $nT_n(w)$ by its empirical version. This then allows us to approximate the critical levels of our proposed multivariate nonparametric test of independence.

Theorem 4. (a) *Assume that all eigenvalues of the integral operator*

$$\mathbf{K}(f)(\underset{\sim}{\boldsymbol{x}}) = \int_{\mathbb{R}^q} f(\underset{\sim}{\boldsymbol{y}})C(\underset{\sim}{\boldsymbol{x}}, \underset{\sim}{\boldsymbol{y}})w(\underset{\sim}{\boldsymbol{y}})d\underset{\sim}{\boldsymbol{y}}$$

are of multiplicity one and that conditions (7) and (8) hold.

For all integer $K \geq 1$, we have, when $N \to \infty$:

$$\sum_{k=1}^{K} \left(\lambda_k - \hat{\lambda}_k\right) \xi_k \xrightarrow{P} 0. \tag{18}$$

Here $\hat{\lambda}_k$, $k = 1, \ldots, K$ are the solutions of a discretized version of

$$\lambda f(\underset{\sim}{\boldsymbol{x}}) = \int_{\mathbb{R}^q} f(\underset{\sim}{\boldsymbol{y}})C(\underset{\sim}{\boldsymbol{x}}, \underset{\sim}{\boldsymbol{y}})w(\underset{\sim}{\boldsymbol{y}})d\underset{\sim}{\boldsymbol{y}} \tag{19}$$

whereby, in addition to the discretization, C is replaced with $\hat{C}_n$ as given in (12).

The ξ_k, $k = 1, \ldots$ in (18) are independent chi-square random variables with one degree of freedom.

(b) *In the case where the eigenvalues λ_k of the operator $\mathbf{K}(f)(\underset{\sim}{\boldsymbol{x}})$ or the eigenvalues $\hat{\lambda}_k$ obtained from (17) may be of multiplicity bigger than one, the convergence in (18) needs to be written in a more general form. Denote by ξ_k and η_k, $k = 1, \ldots$ a set of independent pairs of independent chi-square random variables with one degree of freedom. Then, for all $K \in \mathbb{N} \setminus \{0\}$ and when $N \to \infty$:*

$$\sum_{k=1}^{K} \frac{1}{2}\lambda_k \left((1 + |p_k|)\xi_k + (1 - |p_k|)\eta_k\right) - \sum_{k=1}^{K} \frac{1}{2}\hat{\lambda}_k \left((1 + |\hat{p}_k|)\xi_k + (1 - |\hat{p}_k|)\eta_k\right) \xrightarrow{P} 0. \tag{20}$$

In (20), the p_k are defined as $p_k = \int f_k(\underset{\sim}{\boldsymbol{s}})f_k(-\underset{\sim}{\boldsymbol{s}})w(\underset{\sim}{\boldsymbol{s}})d\underset{\sim}{\boldsymbol{s}}$. The $\hat{p}_k$ are defined by cubature approximation of $\hat{p}_k = \int \hat{f}_k(\underset{\sim}{\boldsymbol{s}})\hat{f}_k(-\underset{\sim}{\boldsymbol{s}})w(\underset{\sim}{\boldsymbol{s}})d\underset{\sim}{\boldsymbol{s}}$. The $\hat{f}_k(\underset{\sim}{\boldsymbol{s}})$ values, on a dense grid, are obtained by solving (17) for large N whereas the $\hat{f}_k(-\underset{\sim}{\boldsymbol{s}})$ values, on a dense grid, are obtained by solving an equation of the same type as (17), with $\hat{\Gamma}_n$ replaced by its complex conjugate $\overline{\hat{\Gamma}_n}$.

We note immediately that in (20), we have $|p_k| = 1$ when a corresponding eigenvalue λ_k is of multiplicity one. When the multiplicity of λ_k is certain $m > 1$ then one of the corresponding $|p_k|$ values is equal to one, and the remaining $(m - 1)$ values are equal to 0. The same type of statement also holds true for the relation between the values of $|\hat{p}_k|$ and $\hat{\lambda}_k$.

In a practical implementation, all $\hat{\lambda}_k$ will be of multiplicity 1 with a probability one and then the corresponding η_k would be redundant. In such a case the kth summand in (20) will be reducing to $\left(\lambda_k - \hat{\lambda}_k\right)\xi_k$ in accordance with (18).

4. Normalization

When $p = 2$, let $\underset{\sim}{\mathbf{X}} = (\mathbf{X}^\top, \mathbf{Y}^\top)^\top$ be the two vector components of $\underset{\sim}{\mathbf{X}}$, with a corresponding subdivision of $\underset{\sim}{\boldsymbol{t}} = (\boldsymbol{t}_1{}^\top, \boldsymbol{t}_2{}^\top)^\top$. In this situation, the normalization of the statistic $n\mathrm{T}_n(w)$ is very useful as pointed out in [1]. Let

$$S_2 = \left\| \sqrt{\left\{1 - \left|\hat{\varphi}_{n,1}(\boldsymbol{t}_1)\right|^2\right\}\left\{1 - \left|\hat{\varphi}_{n,2}(\boldsymbol{t}_2)\right|^2\right\}} \right\|_w^2$$

and set

$$\frac{n\mathrm{T}_n(w)}{S_2} = n\mathcal{I}_n^2 = \frac{n\mathcal{V}_n^2}{S_2} \tag{21}$$

so that our notation agrees with those used in [1,31], i.e., $\mathcal{I}_n^2 = \mathrm{T}_n(w)/S_2$ and $\mathrm{T}_n(w) = \mathcal{V}_n^2$, respectively. Here $\mathcal{V}_n$ is the empirical distance covariance and is discussed in [31].

In [1] it is shown, by an application of Cauchy–Schwartz inequality, that their $\mathcal{I}_n \in [0, 1]$ and of course $\mathcal{I}_n^2 \in [0, 1]$. So in their situation $\mathcal{I}_n$ is an empirical measure for the association between two random vectors of arbitrary dimensions.

Now for $p > 2$ and arbitrary dimensions $q_1, \ldots, q_p$, $\sum_{k=1}^p q_k = q$ we consider the following generalization H_n of the quantity S_2:

$$H_n = \left\| \sqrt{1 - \left(1 + \sum_{\ell=1}^{p} (|\hat{\varphi}_{n,\ell}|^{-2} - 1)\right) \prod_{\ell=1}^{p} |\hat{\varphi}_{n,\ell}|^2} \right\|_w^2 \tag{22}$$

which we borrow from [1]. Observe that when $p = 2, H_n = S_2$. Thus the statistic $n\mathrm{T}_n(w)/H_n$ is a generalization of $n\mathcal{I}_n^2 = n\mathcal{V}_n^2/S_2$ for the case $p > 2$. However, the property $T_n/H_n \in [0, 1]$ is no longer valid in general for arbitrary n and arbitrary weight function when $p > 2$. Although from the proof of Theorem 4, and when n is sufficiently large, the property $T_n/H_n \in [0, 1]$ holds with a very high probability. There is one more benefit of the normalization: under the null hypothesis of independence the weak limit of $n\mathrm{T}_n(w)/H_n$ will have expected value equal to 1. For this reason, we give next more details about the calculation of H_n. More specifically, we can show that the following relation between the empirical norming H_n and the empirical function $C(\underset{\sim}{\boldsymbol{t}}, \underset{\sim}{\boldsymbol{t}})$:

$$H_n = \left\| \sqrt{\hat{C}_n(\underset{\sim}{\boldsymbol{t}}, \underset{\sim}{\boldsymbol{t}})} \right\|_w^2. \tag{23}$$

The proof of this relation is delegated to the end of the Proofs section. Also there, we demonstrate in detail (compare the formula (28)) the most general formula to assemble the calculation of H_n from simple-to-calculate ingredients involving the data. We also show that, like in Lemma 1, we get a simplification of (28) which helps to significantly alleviate the calculations for the particular case where the weight function is integrable, i.e., when $\int_{\mathbb{R}^q} w(\underset{\sim}{\boldsymbol{t}})\, d\underset{\sim}{\boldsymbol{t}} = 1$.

5. Suggestions for choices of the v functions

We propose and investigate several choices of weight functions in the construction of the test-statistic. In the paper [31] (in particular, on p. 2771) a strong preference is expressed towards a choice of non-integrable weights when defining the distance correlation. The weights proposed in [31] are listed below as our *Choice 3* weights. In the paper [29] even statements about uniqueness of those non-integrable weights have been derived under requirements for "rigid motion invariance and scale invariance" (see, e.g., [29, p. 2279]). With respect to use of weights in hypothesis testing context, we are of the opinion that using a *variety* of different (both integrable and non-integrable) weights can only be beneficial. The behaviour under the alternative may strongly be influenced by the choice of the weight. Specifically under the alternatives, these effects can influence the power of the test. Let $\|\cdot\|_p$ be the p-norm, namely $\|\underset{\sim}{\boldsymbol{t}}\|_p = \left(\sum_{j=1}^q |t_j|^p\right)^{1/p}$ for any $\underset{\sim}{\boldsymbol{t}} \in \mathbb{R}^q$ and let $|\underset{\sim}{\boldsymbol{t}}|_q$ be the Euclidean norm (e.g., 2-norm) for a vector $\underset{\sim}{\boldsymbol{t}} \in \mathbb{R}^q$. Our suggested choices of weights are:

5.1. Choice 1

Let $a > 0$.

$$v(\boldsymbol{t}_\ell) = (2\pi)^{-q_\ell/2} a^{-q_\ell} e^{-\frac{1}{2a^2}\|\boldsymbol{t}_\ell\|_2^2} = (2\pi)^{-q_\ell/2} a^{-q_\ell} \exp\left(-\frac{1}{2a^2}\sum_{k=1}^{q_\ell} t_{\ell,k}^2\right).$$

We have $\int_{\mathbb{R}^{q_l}} v(\boldsymbol{t}_\ell) d\boldsymbol{t}_\ell = 1$ and $\gamma_{j,\ell} = 1 - \exp\left(-\frac{a^2}{2}\left\|\mathbf{X}_j^\ell\right\|_2^2\right) = 1 - \xi_{j,\ell}$. Moreover,

$$\xi_{j,j',\ell} = \exp\left(-\frac{a^2}{2}\left\|\mathbf{X}_j^\ell - \mathbf{X}_{j'}^\ell\right\|_2^2\right) = \prod_{k=1}^{q_\ell} \exp\left\{-\frac{a^2}{2}\left(X_{j,k}^\ell - X_{j',k}^\ell\right)^2\right\} \text{ and } \xi_{j,j,\ell} = 1.$$

5.2. Choice 2

Let $a > 0$.

$$v(\boldsymbol{t}_\ell) = a^{-q_\ell} 2^{-q_\ell} e^{-\frac{1}{a}\|\boldsymbol{t}_\ell\|_1} = a^{-q_\ell} 2^{-q_\ell} \exp\left(-\frac{1}{a}\sum_{k=1}^{q_l} |t_{l,k}|\right).$$

We have $\int_{\mathbb{R}^{q_l}} v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell = 1$ and $\gamma_{j,\ell} = 1 - \prod_{k=1}^{q_\ell}\left\{1 + \left(aX_{j,k}^\ell\right)^2\right\}^{-1} = 1 - \xi_{j,\ell}$. Moreover,

$$\xi_{j,j',\ell} = \prod_{k=1}^{q_\ell}\left\{1 + a^2\left(X_{j,k}^\ell - X_{j',k}^\ell\right)^2\right\}^{-1} \text{ and } \xi_{j,j,\ell} = 1.$$

5.3. Choice 3

For all $0 < a < 2$, let $v(\boldsymbol{t}_\ell) = (C(q_\ell, a))^{-1}|\boldsymbol{t}_\ell|_{q_\ell}^{-q_\ell - a}$ with

$$C(d, a) = \frac{2\pi^{d/2}\Gamma(1 - a/2)}{a2^a\Gamma((d + a)/2)}.$$

We have $\int_{\mathbb{R}^{q_\ell}} v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell = \infty$ and $\gamma_{j,l} = \|\mathbf{X}_j^\ell\|_2^a$. We also have $\xi_{j,\ell} = \infty$.

5.4. Choice 4

Let $a \in \mathbb{R}$ and $v(\boldsymbol{t}_\ell) = \prod_{j=1}^{q_\ell}\frac{\sin^2(at_j)}{\pi|a|t_j^2}$. We have $\int_{\mathbb{R}^{q_\ell}} v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell = 1$ and

$\gamma_{j,\ell} = 1 - \prod_{k=1}^{q_\ell}\frac{-2|X_{j,k}^\ell| + |X_{j,k}^\ell - 2a| + |X_{j,k}^\ell + 2a|}{4|a|} = 1 - \xi_{j,\ell}$. We also have

$$\xi_{j,j',\ell} = \prod_{k=1}^{q_\ell}\frac{-2|X_{j,k}^\ell - X_{j',k}^\ell| + |X_{j,k}^\ell - X_{j',k}^\ell - 2a| + |X_{j,k}^\ell - X_{j',k}^\ell + 2a|}{4|a|}$$

and $\xi_{j,j,\ell} = 1$.

5.5. Choice 5

Let $a > 0$ and $v(\boldsymbol{t}_\ell) = \pi^{-q_\ell/2}2^{q_\ell}a^{3q_\ell/2}\prod_{j=1}^{q_\ell}t_j^2e^{-at_j^2}$. We have $\int_{\mathbb{R}^{q_\ell}} v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell = 1$ and

$$\gamma_{j,\ell} = 1 - \prod_{k=1}^{q_\ell}\frac{\left\{2a - (X_{j,k}^\ell)^2\right\}e^{-(X_{j,k}^\ell)^2/(4a)}}{2a} = 1 - \xi_{j,\ell}.$$

We also have $\xi_{j,j,\ell} = 1$ and

$$\xi_{j,j',\ell} = \prod_{k=1}^{q_\ell}\frac{\left\{2a - (X_{j,k}^\ell - X_{j',k}^\ell)^2\right\}e^{-(X_{j,k}^\ell - X_{j',k}^\ell)^2/(4a)}}{2a}.$$

6. Numerical implementation and simulation experiments

6.1. Choice of design parameters

In practice N should be taken large to have a good convergence of the Monte-Carlo approximation (13) and to make sure that (17) delivers an accurate enough discretized version of (19). The eigenvectors represent approximations to the eigenfunctions evaluated at N points. In theory, the larger N the better, however in a practical sense increasing N requires a solution of a large-dimensional eigenvalue–eigenvector problem and numerical inaccuracies due to the curse of dimension may destroy the theoretical advantages. Hence N should stay within reasonable limits. Our extensive numerical experimentation shows that $N = 1000$ is a good default. For the same reason, many of the eigenvalues that are small in magnitude will be calculated with a non-negligible numerical error so only the first $K \ll N$ eigenvalues (and associated eigenvectors) should be taken into account. Some efficient numerical procedures allow us to calculate the first K eigenvalues only thus saving significant computational time.

Another point is that for the limiting distribution of the statistic $nT_n(w)$ as given in (9) in Theorem 2 we need to consider the case of eigenvalues of multiplicity larger than one. Part (b) of Theorem 4 suggests a general result covering the case of roots of multiplicity larger than one for the eigenvalue problem in (17). It may indeed be the case that roots of multiplicity larger than one do exist for (17). However, due to numerical inaccuracies in solving this equation system, the resulting eigenvalues would never be precisely equal to each other and discovering these cases will be virtually impossible. Hence, in the numerical work, when calculating the approximation to the limiting distribution of the test-statistic, we always use the expression $\sum_{k=1}^K\hat{\lambda}_k\xi_k$. For calculating the approximation for the p-values of our test-statistic, we need a reliable

approximation for the cumulative distribution function of the latter. This is a particular case of a quadratic form in normal random variables. Several approaches have been proposed in the literature to evaluate numerically the quantiles of such quadratic forms including methods that rely on numerical inversion of the characteristic function (see, e.g., [7,17]). Some of these methods are compared in [9] and are implemented in an R package called `CompQuadForm`. We use the `imhof()` function of this package to compute the critical values for our test. All the above taken into account in our numerical experimentation, leads us to a recommendation of $K = 200$ as a default.

The above default values of N and K would of course not be universally good. Since we rely on the asymptotic statements of Theorems 2 and 4 for calculating the critical values, the decision on a certain number K of eigenvalues is also related to the sample size n that allows for a good approximation in (18). The interplay between N, K and n is complicated in general. To get a rough idea we consider the case of Choice 1 from Section 5 when all components of the data vector are independent standard Gaussian. Then a simple calculation leads us to:

$$\int C\left(\underset{\sim}{t}, \underset{\sim}{t}\right) \prod_{\ell=1}^{p} v(t_\ell) dt_\ell = 1 - \left\{1 - p + \sum_{\ell=1}^{p} (2a^2+1)^{q_\ell/2}\right\} (2a^2+1)^{-p/2}.$$

On the other hand,

$$\mathbb{E}\left(n\mathrm{T}_n(w)\right) = \int \mathbb{E}\left(\left|\sqrt{n}\mathrm{D}_n\left(\underset{\sim}{t}\right)\right|^2\right) w\left(\underset{\sim}{t}\right) d\underset{\sim}{t} \longrightarrow \int \mathbb{E}\left\{\mathrm{D}\left(\underset{\sim}{t}\right)\overline{\mathrm{D}\left(\underset{\sim}{t}\right)}\right\} w\left(\underset{\sim}{t}\right) d\underset{\sim}{t} := H_\infty.$$

Hence we know

$$H_\infty = 1 - \left\{1 - p + \sum_{\ell=1}^{p} (2a^2+1)^{q_\ell/2}\right\} (2a^2+1)^{-p/2}$$

in this case. We have also that $H_n \longrightarrow H_\infty$ (where H_n is defined in (22)) and that

$$\mathbb{E}\left\{n\mathrm{T}_n(w)\right\} \longrightarrow \mathbb{E}\left(\sum_{k=1}^{\infty} \lambda_k \xi_k\right) = \sum_{k=1}^{\infty} \lambda_k = H_\infty.$$

Also in the general situation, when the theoretical H_∞ is not available, one can still use a comparison of $\left|\sum_{k=1}^{K} \hat{\lambda}_{k,N} - H_n\right|$ to a small threshold value $\epsilon > 0$ as a criterion of a choice of K. Since both $\sum_{k=1}^{K} \hat{\lambda}_{k,N}$ and H_n would be close to H_∞, they are also expected to be close to each other. Hence for large n and N, a recommendation for the value of K is to choose it in such a way that $\left|\sum_{k=1}^{K} \hat{\lambda}_{k,N} - H_n\right| < \epsilon$ is satisfied. Difficulties in choosing K usually only appear for large dimensions q and p. For dimensions such as 2, 3, 4 or 5, the eigenvalues usually are nicely behaving, showing a strongly decreasing pattern and in such cases we choose a small threshold such as 10^{-8} and only keep the eigenvalues above this threshold.

Finally, the five weight choices we propose in Section 5, involve an additional one-dimensional parameter a. The flexibility to choose a allows us to increase the power of our test in comparison to the power of the test that is implemented in the `dCov` program of Székely and Rizzo (for the same level of significance). The asymptotic properties of the test imply that the choice of a does not matter, but for fixed sample sizes such as $n = 50$, 100 or 200 this choice makes some difference. Depending on the scenarios chosen for model comparison in Section 6.2, the values of a for our weight choice 1 were chosen in the range from 0.2 to 4. If the data is standardized individually (i.e., each observation vector is transformed by subtracting its arithmetic mean and dividing by the standard deviation estimator) then a value of $a = 1$ turned out to be universally good value to use for scenarios with weight choice 1.

6.2. Comparison with other competing programs

We have implemented our multivariate independence tests in an R package called `IndependenceTests`. The program that performs the testing in the package is called `mdcov`. We present below power comparisons on some unusual bivariate distributions, to demonstrate the good power of our test by comparing it with the tests from [28] (R program called `dCov`) from the `energy` package and [15] (R program `HHG`). To the best of our knowledge, the only other widely available computer program that performs a multivariate independence test at the level of generality considered in this paper (that is, for $p \geq 3$), is the function `dependogram()` implementing the test from the paper [2]. Their program calculates the critical values using bootstrap. We compare the performance of our test with the above program. The outcomes from this program are labelled BBL in this section. The reader is referred to [2] for more details on their program. Here we only note that the implemented bootstrap method there is very specific and the computational effort is controlled by two parameters: N (which, to avoid conflict of notation with N used in our paper, will be denoted by N^* here), and B (the number of bootstrap replications). The role of N^* (related to number of points chosen on q_i-dimensional spheres ($i = 1, \ldots, p$)) is related to the accuracy of the bootstrap results and so is B, too. It can be seen from Table 1 that for the bootstrap approximation to work reasonably well with respect to size, we need the replications to be at least $B = 200$ (which is also the default value in the package). The number of computations of the test-statistic becomes proportional to $B2^p(N^*)^{q-p}$ (compare again [2], specifically p. 1818).

Table 1
Empirical size computed for $n = 100$ over $M = 1,000$ repetitions for the different methods. Data generated from independent q dimensional Gaussian. Gaussian weights are used for `mdcov`, with $a = 1$, and with $N^* = 10$, $B = 200$ for BBL.

Method	q	N	K	Emp. size
`mdcov`	2	1000	200	0.047(0.007)
`HHG`	2	–	–	0.047(0.007)
`dCov`	2	–	–	0.05 (0.007)
`BBL`	2	–	–	0.053 (0.007)
`mdcov`	$4(2+2)$	1000	200	0.056(0.007)
`HHG`	$4(2+2)$	–	–	0.054(0.007)
`dCov`	$4(2+2)$	–	–	0.051(0.007)
`BBL`	$4(2+2)$	–	–	0.046 (0.006)

Table 2
Empirical size for $q = 6, p = 3$ computed for $n = 100$ over $M = 1,000$ repetitions ($N^* = 10, B = 200$ for BBL). The acronym mixture stands for a 3-dimensional correlated standard Gaussian (with a joint correlation coefficient of 0.5); an independent two-dimensional correlated Poisson Y_1, Y_2 with $Y_1 = Z_1 + Z_2, Y_2 = Z_2 + Z_3$ and $Z_1 \sim \mathcal{P}(1), Z_2 \sim \mathcal{P}(3), Z_3 \sim \mathcal{P}(1)$ being independent; and an independent one-dimensional Bernoulli with probability of success equal to 1/2.

Data	Method	a	N	K	Emp. size
3 indep. pairs of dependent Gauss.	`mdcov`	0.2	1000	200	0.048(0.007)
3 indep. pairs of dependent uniforms	`mdcov`	2	1000	200	0.049(0.007)
3 indep. pairs of dependent Gauss.	`mdcov` Cuba	0.2	272	272	0.048(0.007)
3 indep. pairs of dependent Gauss.	`BBL`	–	–	–	0.054(0.007)
3 indep. pairs of dependent uniforms	`BBL`	–	–	–	0.049(0.007)
Mixture $(3+2+1)$	`mdcov`	0.3	1000	200	0.048(0.007)
Mixture $(3+2+1)$	`BBL`	–	–	–	0.061(0.007)

This leads to an exponential increase of the numerical calculations of the test-statistic when $q - p$ is large. The need to increase B beyond 200 when $q-p$ is large can increase significantly the computing time for the bootstrap and it can become uncompetitive for $q > 4$. When $q = 4$ and $p = 2$, (i.e., $q-p = 2$), computational times are comparable. However, even when B does not increase much and p is small *but* $q-p$ is large, the `BBL` method becomes much less competitive computationally. For example, in the case of the model coded $Y_j = X_j\epsilon_j$ in the Table 3 (where $p = 2$ only but $q - p = 8$) it was not possible to apply the `BBL` approach for any reasonable amount of hours of time. At the same time, it is not hard to calculate accurately the p-values based on our asymptotic approximation, as illustrated in Tables 1–3. The size is evaluated by the empirical (out of $M = 1,000$ repetitions) ratio of rejections of the null hypothesis of independence and is presented in the last columns of Tables 1 and 2. The empirical values in these columns need to be compared to the ideal value of $\alpha = 0.05$. Table 2 represents simulations for models containing $p \geq 3$ vectors for which the `dCov` and `HHG` are not tailored hence we only discuss `mdcov` results. These are all related to designs with a total dimension $q = 6$. We experimented with a set of 3 two-dimensional standard Gaussian vectors where within pair there was a correlation of 0.5, and also with a set of 3 independent pairs of dependent uniform vectors (the dependency being created via Gaussian copula with correlation coefficient of 0.5). Finally we also chose a 6-dimensional vector consisting of an independent 3-dimensional correlated standard Gaussian (with a joint correlation coefficient of 0.5), two-dimensional correlated Poisson and an independent one-dimensional Bernoulli component (i.e., $q = 6, q_1 = 3, q_2 = 2, q_3 = 1.$). The choice of the variance parameter a of the weight function also has significant impact for sample sizes such as 100, 200 or 300.

Besides the Monte Carlo-based evaluation of the solution to the eigenvalue–eigenfunction problem (11) we present in Table 2 a result obtained via application of a Gaussian cubature formula based on the celebrated Stroud's method. These cubature formulae require a fixed number of nodes. The number of nodes is $2^{q+1} + 4q^2$ when $q > 2$ which, for dimension 6 discussed in Table 2, gives $N = 272$ for the numerical cubature in this case. For $q = 2$ the number N turns out to be only equal to 44. In principle, the cubature method is much faster than the Monte Carlo method. However we abstain from recommending the cubature method for all situations. Indeed, for $q = 2$, $N = 44$ in the examples considered in the Table 3, the quadrature method performed poorly and the empirical size of the resulting tests was closer to 1% rather than to the nominal α of 5%. Also, for large q, the rate of falling of the eigenvalues slows and for another distribution types and for higher dimension q the fixed number of nodes of the cubature formulae may not be enough for accurate solution of the integral equation. However for the multivariate Gaussian example considered in Table 2, the outcome for the empirical level of the cubature method was only slightly worse than the Monte Carlo result that used $N = 1,000$, $K = 200$.

Table 3 represents size and power comparisons between `HHG`, `dCov`, `mdcov` and `BBL`. The reason to compare with these two programs is the following. The recent paper [15] does point out the superiority of `dCov` over a large variety of classical tests. However, in the comparison in [15], also some lack of sufficient power of `dCov` for small sample sizes such as $n = 50$ has been demonstrated, whereby `HHG` has shown much better power at these sample sizes. The demonstration has been performed on a set of unusual bivariate distributions. The details of these distributions are in the Supplementary

Table 3
Size and power (with standard error in parentheses) of HHG test, dCov test, our mdcov test and BBL for small sample sizes ($n = 50$) for some unusual relations. Results based on $M = 1,000$ repetitions. The design parameters for mdcov were $N = 1,000$, $K = 200$. For BBL: $N^* = 10$, $B = 200$.

Distribution	Independence	a	HHG	dCov	mdcov	BBL
4 indep. clouds	Yes	1	0.058(0.007)	0.058(0.007)	0.059(0.007)	0.048(0.007)
W	Yes	4	0.038(0.006)	0.046(0.007)	0.043(0.006)	0.032(0.006)
W	No	4	0.999(0.001)	0.935(0.007)	1(0)	0.894(0.009)
Diamond	Yes	4	0.047(0.007)	0.042(0.006)	0.039(0.006)	0.046(0.007)
Diamond	No	4	0.747(0.014)	0.040(0.006)	0.520(0.016)	0.039(0.006)
Parabola	Yes	4	0.048(0.007)	0.044(0.006)	0.048(0.007)	0.047(0.007)
Parabola	No	4	0.979(0.004)	0.917(0.008)	0.998(0.001)	0.634(0.015)
2 Parabolas	Yes	4	0.042(0.006)	0.041(0.006)	0.043(0.006)	0.046(0.006)
2 Parabolas	No	4	1(0)	0.331(0.015)	1(0)	0.367(0.015)
Circle	Yes	4	0.058(0.007)	0.051(0.007)	0.045(0.006)	0.052(0.007)
Circle	No	4	0.978(0.005)	0.096(0.009)	0.990(0.003)	0.107(0.010)
$Y_j = X_j\epsilon_j$	Yes	0.4	0.051(0.007)	0.046(0.007)	0.057(0.007)	–
$Y_j = X_j\epsilon_j$	No	0.4	0.956(0.006)	0.415(0.016)	0.608(0.015)	–
GEVmodel1	Yes	0.9	0.046(0.007)	0.044(0.006)	0.037(0.006)	0.049(0.007)
GEVmodel1	No	0.9	0.881(0.010)	0.362(0.015)	0.941(0.007)	0.828(0.012)
GEVmodel2	Yes	0.2	0.057(0.007)	0.053(0.007)	0.054(0.007)	0.048(0.007)
GEVmodel2	No	0.2	0.377(0.015)	0.675(0.015)	0.662(0.015)	0.387(0.015)

material to [15].[1] We demonstrate in Table 3 that mdcov is a worthy contender to both HHG and dCov in terms of power and size. The bivariate relations are called Four independent clouds, W, Diamond, Parabola, Two parabolas, and Circle. The Four independent clouds represent data generated under the independence hypothesis so there we were looking at the size. To make the comparison fair for the other curves, we simulated data first under the same dependence structures like in [15] but in addition we also simulated independent data while keeping the same marginals like in the dependent case. We believe that the power comparison is fairer in this way. To complete the comparison we also include the results for BBL in this table. Looking at the lines that represent the power, we see now that especially for the Circle, for the Parabola and for the Two parabolas case where dCov and BBL were significantly deficient to HHG, our program is now quite competitive to HHG. It also shows significant improvement for the case of Diamond.

The model coded $Y_j = X_j\epsilon_j$ is also discussed in [15]. There, two dependent five-dimensional data vectors are generated ($p = 2$, $q_1 = q_2 = 5$.) Again, the power for mdcov represents a significant improvement over dCov (albeit still being slightly lower than HHG). We should also state that when the sample size is increased to $n = 100$, all three programs showed much better power.

We also include simulations coded as the GEVmodel1 and GEVmodel2. They involve data generation with non-symmetric extreme value distributions of second type (Fréchet) and third type (Weibull), to which a random normal noise has been added. The dependence model for GEVmodel1 is as follows: $X = W + Z_1, Y = X + F + Z_2$ where W is Weibull distributed, F is Fréchet distributed and Z_1, Z_2 are independent $\mathcal{N}(0, 0.2^2)$. The dependence model for GEVmodel2 is as follows: $X = W + Z_1, Y = 3W + Z_2$ where W is Weibull distributed, and Z_1, Z_2 are independent $\mathcal{N}(0, 2^2)$. (The related independence models just keep the corresponding marginal distributions the same as in the respective dependent models). Our choice of these models was motivated by the wish to also include non-symmetric distributions with heavier tails. For these data, mdcov was strongly competitive to HHG, dCov and BBL.

The outcomes of our simulations can be summarized as follows:

1. When using Gaussian weights, with $a = 1$, we get very precise coverage for $q = 2, 4, 6$ at sample sizes as low as $n = 200$ or 500.
2. If we stay with dimension $q = 6$ (at which the computations start getting more challenging and computationally more intensive), we need the number N to be around $N = 1,000$. Smaller values of N may decrease accuracy of the approximation of the eigenvalues whereas larger values would lead to much more intensive calculations due to the large dimension of the linear system in (17). Illustrative outcomes for $q = 6$ are presented in Table 2. We notice that in all cases where our Monte Carlo method was applied for the sake of approximating the limiting distribution, the resulting empirical level of the test was well within the expected limits of $0.05 \pm 1.96 * \sqrt{0.05 * 0.95/1000} = (0.0365, 0.0635)$.
3. For $N = 1,000$, the sum of all eigenvalues appears stable, i.e., virtually does not vary when N is varied around the 1,000 mark. The pattern in which the sorted out eigenvalues go down to zero is significantly influenced by the distribution

[1] These authors have generously made their package publicly available and we could read the codes they have used for their data generation. Some of the data generation mechanisms could be criticized conceptually since instead of generating n independent uniformly distributed random variables on the circle or in $[-1, 1]$, a set of equidistantly spaced deterministic values have been generated. This misconception has been propagated from the paper [24]. In the case of the W example in this paper, this implies as a consequence that the data generation has lead to non-identically distributed independent random variables U and V instead of the claimed identically distributed dependent variables U and V. Therefore some of the comparisons in the HHG paper are not completely valid and here we have corrected their data generation procedures to deliver fair rigorous comparisons. The deviations in the power performance for dCov and HHG are relatively minor but the comparison is more theoretically sound.

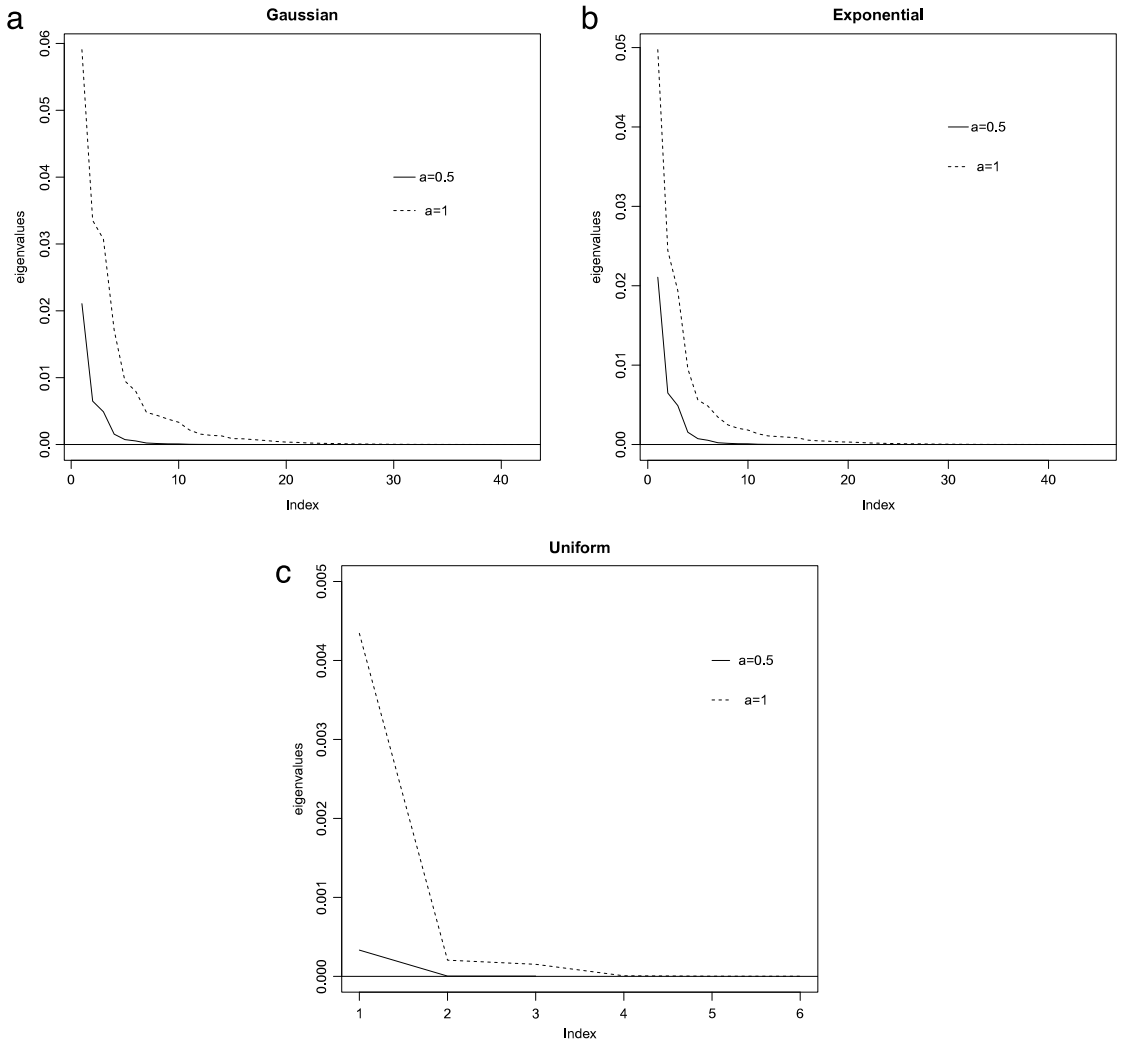

Fig. 1. Patterns of decay of the largest eigenvalues of the matrix $\hat{\Gamma}_n$ for Gaussian, exponential and uniform data ($n = 5,000$, $N = 2,000$) and for values $a = 1$ and $a = 0.5$.

of the single multivariate random vector. The distribution of the data vector in the sample influences the choice of K. Theoretical investigation of the choice of K is difficult and is not discussed in this paper. Fig. 1 shows screeplots which offer two take home messages. Firstly, the pattern of decrease towards zero of the sequence of eigenvalues varies significantly depending on the distribution of a single data vector: if the components arise from a multivariate normal or exponential distribution, the decrease of the sorted out eigenvalues happens much less quickly. Secondly, the variance parameter a of the weight also influences the pattern of convergence towards zero.

4. For dimensions q of the order considered here, the asymptotic approximations to the critical values turned out to be quite good and in agreement with the bootstrap-based approximations based on the methodology presented in [2]. Regarding power, we did apply the BBL method on the examples considered in Table 3. However, the behaviour of BBL was quite comparable to `dCov` in the cases when it was possible to run it (and hence was deficient in comparison to `HHG` and to `mdcov` in these examples. In a nutshell, one can say that the performance of BBL was either comparable or deficient to `mdcov` for the cases when BBL could be run for a reasonable time. In addition, by using quick numerical calculations such as the ones illustrated in Fig. 1 we are able to determine suitable values of a and K to control the accuracy of the approximation of the critical values. This option can help us to reduce the computational time. There is no such option in the BBL method. If, to be on the safe side, we decide to increase N^* and B, then the computational time for BBL can blow up significantly. This difficulty is avoided in our asymptotic approximations since, despite also using simulations, they avoid the repeated calculation of the *statistic* on resampled data vectors. The calculation of the statistic is the numerically costly operation that dramatically increases the time needed by BBL with increasing n. The computational cost also increases linearly with the number of resamplings B and, more importantly, exponentially in the difference $q - p$. If $q - p$ is of the order of 6 or more, the BBL method of [2] could virtually not be used.

7. Proofs

Proof of Lemma 1. We start with the decomposition

$$|\mathrm{D}_n(\underset{\sim}{t})|^2 = \mathrm{D}_n(\underset{\sim}{t})\,\overline{\mathbf{D}_n(\underset{\sim}{t})} = \hat{\varphi}_n(\underset{\sim}{t})\,\hat{\varphi}_n(-\underset{\sim}{t}) - 2Re\left\{\hat{\varphi}_n(\underset{\sim}{t})\,\tilde{\varphi}_n(-\underset{\sim}{t})\right\} + \tilde{\varphi}_n(\underset{\sim}{t})\,\tilde{\varphi}_n(-\underset{\sim}{t}).$$

Using (4) and (5), we obtain

$$\begin{aligned}
|\mathrm{D}_n(\underset{\sim}{t})|^2 &= n^{-2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\sum_{B\subset I_p}\sum_{B'\subset I_p}(-1)^{|B|+|B'|}\prod_{\ell\in B}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)\prod_{\ell'\in B'}f_{\mathbf{X}_{j'}^{\ell'}}(-\boldsymbol{t}_{\ell'})\\
&\quad - 2Re\left\{n^{-1}\sum_{j=1}^{n}\sum_{B\subset I_p}\sum_{B'\subset I_p}(-1)^{|B|+|B'|}\prod_{\ell\in B}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)\prod_{\ell'\in B'}\left(n^{-1}\sum_{j'=1}^{n}f_{\mathbf{X}_{j'}^{\ell'}}(-\boldsymbol{t}_{\ell'})\right)\right\}\\
&\quad + \sum_{B\subset I_p}\sum_{B'\subset I_p}(-1)^{|B|+|B'|}\prod_{\ell\in B}\left\{n^{-1}\sum_{j=1}^{n}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)\right\}\prod_{\ell\in B'}\left\{n^{-1}\sum_{j'=1}^{n}f_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell)\right\}\\
&= \sum_{B\subset I_p}\sum_{B'\subset I_p}(-1)^{|B|+|B'|}\left\{\mathrm{C}_1(\underset{\sim}{t}) - 2\mathrm{C}_2(\underset{\sim}{t}) + \mathrm{C}_3(\underset{\sim}{t})\right\},
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{C}_1(\underset{\sim}{t}) &= n^{-2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\left\{\prod_{\ell\in B\cap B'}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)f_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell)\right\}\prod_{\ell\in B\setminus B'}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)\prod_{\ell\in B'\setminus B}f_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell),\\
\mathrm{C}_2(\underset{\sim}{t}) &= n^{-1}Re\left[\sum_{j=1}^{n}\left[\prod_{\ell\in B\cap B'}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)\left\{n^{-1}\sum_{j'=1}^{n}f_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell)\right\}\right]\prod_{\ell\in B\setminus B'}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)\prod_{\ell\in B'\setminus B}\left\{n^{-1}\sum_{j'=1}^{n}f_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell)\right\}\right]
\end{aligned}$$

and

$$\mathrm{C}_3(\underset{\sim}{t}) = \left[\prod_{\ell\in B\cap B'}\left\{n^{-1}\sum_{j=1}^{n}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)\right\}\left\{n^{-1}\sum_{j'=1}^{n}f_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell)\right\}\right]\prod_{\ell\in B\setminus B'}\left\{n^{-1}\sum_{j=1}^{n}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)\right\}\prod_{\ell\in B'\setminus B}\left\{n^{-1}\sum_{j'=1}^{n}f_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell)\right\}.$$

Hence we obtain

$$\int|\mathrm{D}_n(\underset{\sim}{t})|^2 w(\underset{\sim}{t})\,d\underset{\sim}{t} = \sum_{B\subset I_p}\sum_{B'\subset I_p}(-1)^{|B|+|B'|}\left\{\int\mathrm{C}_1(\underset{\sim}{t})\,w(\underset{\sim}{t})\,d\underset{\sim}{t} - 2\int\mathrm{C}_2(\underset{\sim}{t})\,w(\underset{\sim}{t})\,d\underset{\sim}{t} + \int\mathrm{C}_3(\underset{\sim}{t})\,w(\underset{\sim}{t})\,d\underset{\sim}{t}\right\}.$$

We present below the calculation of all terms involved in this formula. Note that *all* of them are real valued. First, because of the antisymmetry of the sine function and of the assumed symmetry of the $v(\boldsymbol{t}_\ell)$ function, we realize that for $\boldsymbol{x}\in\mathbb{R}^{q_\ell}$:

$$\int_{\mathbb{R}^{q_\ell}}\left\{1 - e^{i\boldsymbol{t}_\ell{}^\top\boldsymbol{x}}\right\}v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell = \int_{\mathbb{R}^{q_\ell}}\left\{1-\cos\left(\boldsymbol{t}_\ell{}^\top\boldsymbol{x}\right)\right\}v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell$$

is real-valued. Then, of course

$$\gamma_{j,\ell} := \int_{\mathbb{R}^{q_\ell}}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell = \int_{\mathbb{R}^{q_\ell}}f_{\mathbf{X}_j^\ell}(-\boldsymbol{t}_\ell)v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell$$

and

$$\gamma_{j,j',\ell} := \int_{\mathbb{R}_{q_\ell}}f_{\mathbf{X}_j^\ell-\mathbf{X}_{j'}^\ell}(\boldsymbol{t}_\ell)v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell$$

are also real-valued. Finally,

$$\beta_{j,j',\ell} = \beta_{j',j,\ell} := \gamma_{j,\ell} + \gamma_{j',\ell} - \gamma_{j,j',\ell} = \int_{\mathbb{R}^{q_\ell}}f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)f_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell)v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell\in\mathbb{R}$$

is also real-valued. Note that when $j=j'$, $\beta_{j,j',\ell} = 2\gamma_{j,\ell}$ holds.

Putting everything together we obtain the formula from part a) of the Theorem.

Similarly, from (4) and (5) using the functions $g_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)$ instead of $f_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)$, we obtain

$$\begin{aligned}|\mathrm{D}_n(\underset{\sim}{\boldsymbol{t}})|^2 &= n^{-2}\sum_{j=1}^n\sum_{j'=1}^n\prod_{\ell=1}^p g_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)g_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell) - 2Re\left[n^{-1}\sum_{j=1}^n\prod_{\ell=1}^p\left\{n^{-1}\sum_{j'=1}^n g_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)g_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell)\right\}\right]\\ &\quad+\prod_{\ell=1}^p\left\{n^{-2}\sum_{j=1}^n\sum_{j'=1}^n g_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)g_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell)\right\}\\ &:= \mathrm{G}_1(\underset{\sim}{\boldsymbol{t}}) - 2\mathrm{G}_2(\underset{\sim}{\boldsymbol{t}}) + \mathrm{G}_3(\underset{\sim}{\boldsymbol{t}}).\end{aligned}$$

Proceeding like before, we get

$$\int|\mathrm{D}_n(\underset{\sim}{\boldsymbol{t}})|^2 w(\underset{\sim}{\boldsymbol{t}})\,d\underset{\sim}{\boldsymbol{t}} = \int \mathrm{G}_1(\underset{\sim}{\boldsymbol{t}})\,w(\underset{\sim}{\boldsymbol{t}})\,d\underset{\sim}{\boldsymbol{t}} - 2\int \mathrm{G}_2(\underset{\sim}{\boldsymbol{t}})\,w(\underset{\sim}{\boldsymbol{t}})\,d\underset{\sim}{\boldsymbol{t}} + \int \mathrm{G}_3(\underset{\sim}{\boldsymbol{t}})\,w(\underset{\sim}{\boldsymbol{t}})\,d\underset{\sim}{\boldsymbol{t}}$$

whose computation involves terms such as:

$$\int_{\mathbb{R}^{q_\ell}} g_{\mathbf{X}_j^\ell}(\boldsymbol{t}_\ell)g_{\mathbf{X}_{j'}^\ell}(-\boldsymbol{t}_\ell)v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell = \int_{\mathbb{R}^{q_\ell}}\cos\left\{\boldsymbol{t}_\ell{}^\top\left(\mathbf{X}_j^\ell - \mathbf{X}_{j'}^\ell\right)\right\}v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell := \xi_{j,j',\ell}\in\mathbb{R}.$$

We also get

$$\xi_{j,\ell} := \int_{\mathbb{R}^{q_\ell}}\cos\left(\boldsymbol{t}_\ell{}^\top\mathbf{X}_j^\ell\right)v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell.$$

Furthermore when $j = j'$, we have $\xi_{j,j',\ell} = \int_{\mathbb{R}^{q_\ell}} v(\boldsymbol{t}_\ell)d\boldsymbol{t}_\ell$, $\ell = 1,\ldots,p$.

These expressions are all equal to one under the assumption (b) of the Theorem. Hence putting everything together, we end up with the desired formula

$$nT_n(w) = n^{-1}\sum_{j=1}^n\sum_{j'=1}^n\prod_{\ell=1}^p\xi_{j,j',\ell} - 2\sum_{j=1}^n\prod_{\ell=1}^p\left(n^{-1}\sum_{j'=1}^n\xi_{j,j',\ell}\right) + n\prod_{\ell=1}^p\left(n^{-2}\sum_{j=1}^n\sum_{j'=1}^n\xi_{j,j',\ell}\right).$$

Proof of Theorem 1. We need to prove that $\sqrt{n}\mathrm{D}_n$ converges in $C(\mathbb{R}^q,\mathbb{C})$ to a complex Gaussian process D, where the space $C(\mathbb{R}^q,\mathbb{C})$ is endowed with the topology of uniform convergence on compacts. For this purpose, we define the metric ρ on $C(\mathbb{R}^q,\mathbb{C})$ defined by

$$\rho(x,y) := \sum_{j=1}^\infty 2^{-j}\frac{\rho_j(x,y)}{1+\rho_j(x,y)},$$

where

$$\rho_j(x,y) = \sup_{\|\underset{\sim}{\boldsymbol{t}}\|\le j}|x(\underset{\sim}{\boldsymbol{t}}) - y(\underset{\sim}{\boldsymbol{t}})|$$

under the usual sup norm. For all mappings $x\in C(\mathbb{R}^q,\mathbb{C})$, let $r_j(x)$ be the restriction of x to the ball $\mathcal{B}_j^q$ of radius j. By [18, Proposition 14.6], it suffices to show that $r_j(\sqrt{n}\mathrm{D}_n)\rightsquigarrow r_j(\mathrm{D})$ in order to show the convergence $\sqrt{n}\mathrm{D}_n\rightsquigarrow\mathrm{D}$.

We have

$$\sqrt{n}\mathrm{D}_n(\underset{\sim}{\boldsymbol{t}}) = \sqrt{n}\left\{\hat{\varphi}_n(\underset{\sim}{\boldsymbol{t}}) - \prod_{\ell=1}^p\hat{\varphi}_{n,\ell}(\boldsymbol{t}_\ell)\right\} = \frac{1}{\sqrt{n}}\sum_{j=1}^n e^{i\underset{\sim}{\boldsymbol{t}}^\top\mathbf{X}_j} - \sqrt{n}\tilde{\varphi}_n(\underset{\sim}{\boldsymbol{t}}).$$

This is an empirical characteristic function process for which tightness on compacts was shown in Csörgő's Theorem [6, p. 294]. Our condition (6) quotes Csörgő's condition. This is a mild condition on the tails of the joint distribution of the data.

Now, we compute the moments of this complex Gaussian process. Recall that

$$\hat{\varphi}_n(\underset{\sim}{\boldsymbol{t}}) = n^{-1}\sum_{j=1}^n\prod_{\ell=1}^p e^{i\boldsymbol{t}_\ell{}^\top\mathbf{X}_j^\ell}\quad\text{and}\quad\tilde{\varphi}_n(\underset{\sim}{\boldsymbol{t}}) = \prod_{\ell=1}^p n^{-1}\sum_{j=1}^n e^{i\boldsymbol{t}_\ell{}^\top\mathbf{X}_j^\ell}.$$

Under the null hypothesis, we have

$$\mathbb{E}\{\mathrm{D}_n(\underset{\sim}{\boldsymbol{t}})\} = \mathbb{E}\left\{\hat{\varphi}_n(\underset{\sim}{\boldsymbol{t}})\right\} - \mathbb{E}\{\tilde{\varphi}_n(\underset{\sim}{\boldsymbol{t}})\} = n^{-1}\sum_{j=1}^n\prod_{\ell=1}^p\varphi_\ell(\boldsymbol{t}_\ell) - \prod_{\ell=1}^p n^{-1}\sum_{j=1}^n\varphi_\ell(\boldsymbol{t}_\ell) = 0,$$

$$\begin{aligned}
\mathrm{D}_n(\underset{\sim}{s})\,\overline{\mathrm{D}_n(\underset{\sim}{t})} &= \hat{\varphi}_n(-\underset{\sim}{t})\,\hat{\varphi}_n(\underset{\sim}{s}) - \hat{\varphi}_n(-\underset{\sim}{t})\,\tilde{\varphi}_n(\underset{\sim}{s}) - \tilde{\varphi}_n(-\underset{\sim}{t})\,\hat{\varphi}_n(\underset{\sim}{s}) + \tilde{\varphi}_n(-\underset{\sim}{t})\,\tilde{\varphi}_n(\underset{\sim}{s}) \\
&= n^{-2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\prod_{\ell=1}^{p} e^{-i\boldsymbol{t}_\ell{}^\top \mathbf{X}_j^\ell} e^{i\boldsymbol{s}_\ell{}^\top \mathbf{X}_{j'}^\ell} - n^{-1}\sum_{j=1}^{n}\prod_{\ell=1}^{p}\left(e^{-i\boldsymbol{t}_\ell{}^\top \mathbf{X}_j^\ell} n^{-1}\sum_{j'=1}^{n} e^{i\boldsymbol{s}_\ell{}^\top \mathbf{X}_{j'}^\ell}\right) \\
&\quad - n^{-1}\sum_{j=1}^{n}\prod_{\ell=1}^{p}\left(e^{i\boldsymbol{s}_\ell{}^\top \mathbf{X}_j^\ell} n^{-1}\sum_{j'=1}^{n} e^{-i\boldsymbol{t}_\ell{}^\top \mathbf{X}_{j'}^\ell}\right) + \prod_{\ell=1}^{p}\left(n^{-2}\sum_{j=1}^{n}\sum_{j'=1}^{n} e^{-i\boldsymbol{t}_\ell{}^\top \mathbf{X}_j^\ell} e^{i\boldsymbol{s}_\ell{}^\top \mathbf{X}_{j'}^\ell}\right).
\end{aligned}$$

Hence

$$\begin{aligned}
\mathbb{E}[n\mathrm{D}_n(\underset{\sim}{s})\,\overline{\mathrm{D}_n(\underset{\sim}{t})}] &= n^{-1}\sum_{j=1}^{n}\sum_{j'=1;j'\neq j}^{n}\prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + n^{-1}\sum_{j=1}^{n}\prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell) \\
&\quad - \sum_{j=1}^{n}\prod_{\ell=1}^{p}\left\{n^{-1}\sum_{j'=1;j'\neq j}^{n}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + n^{-1}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)\right\} \\
&\quad - \sum_{j=1}^{n}\prod_{\ell=1}^{p}\left\{n^{-1}\sum_{j'=1;j'\neq j}^{n}\varphi_\ell(\boldsymbol{s}_\ell)\varphi_\ell(-\boldsymbol{t}_\ell) + n^{-1}\varphi_\ell(\boldsymbol{s}_\ell-\boldsymbol{t}_\ell)\right\} \\
&\quad + n\prod_{\ell=1}^{p}\left[n^{-2}\sum_{j=1}^{n}\left\{\sum_{j'=1;j'\neq j}^{n}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + \varphi_\ell(\boldsymbol{s}_\ell-\boldsymbol{t}_\ell)\right\}\right] \\
&= (n-1)\prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + \prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell) \\
&\quad - 2\sum_{j=1}^{n}\prod_{\ell=1}^{p}\left\{\frac{n-1}{n}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + n^{-1}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)\right\} \\
&\quad + n\prod_{\ell=1}^{p}\left[n^{-2}\sum_{j=1}^{n}\{(n-1)\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + \varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)\}\right] \\
&= (n-1)\prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + \prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell) \\
&\quad - 2n^{-p+1}\prod_{\ell=1}^{p}\{(n-1)\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + \varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)\} \\
&\quad + n^{-p+1}\prod_{\ell=1}^{p}\{(n-1)\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + \varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)\} \\
&= n\prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + \left\{\prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell) - \prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\right\} \\
&\quad - n\prod_{\ell=1}^{p}\left[\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + \frac{1}{n}\{\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell) - \varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\}\right] \\
&= \prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell) - \prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) \\
&\quad - \sum_{j=1}^{p}\left\{\varphi_j(-\boldsymbol{t}_j+\boldsymbol{s}_j) - \varphi_j(-\boldsymbol{t}_j)\varphi_j(\boldsymbol{s}_j)\right\}\prod_{\ell=1;\ell\neq j}^{p}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell) + O\left(\frac{1}{n}\right)
\end{aligned}$$

where the remainder of order $O(\frac{1}{n})$ has the explicit form

$$O\left(\frac{1}{n}\right) = -\frac{1}{n}\sum_{B\subset I_p,|B|\geq 2}\frac{1}{n^{|B|-2}}\left[\prod_{k\in B}\{\varphi_k(-\boldsymbol{t}_k+\boldsymbol{s}_k) - \varphi_k(-\boldsymbol{t}_k)\varphi_k(\boldsymbol{s}_k)\}\right]\left\{\prod_{\ell\in I_p\setminus B}\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_l(\boldsymbol{s}_\ell)\right\}.$$

Now we get after some elementary transformations:

$$\begin{aligned}
C(\underset{\sim}{s},\underset{\sim}{t}) &= \prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)-\prod_{\ell=1}^{p}\{\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\}\\
&\quad-\prod_{j=1}^{p}\{\varphi_j(-\boldsymbol{t}_j)\varphi_j(\boldsymbol{s}_j)\}\sum_{\ell=1}^{p}\frac{\{\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)-\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\}}{\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)}\\
&= \prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)-\prod_{\ell=1}^{p}\{\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\}\left[1+\sum_{\ell=1}^{p}\frac{\{\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)-\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\}}{\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)}\right]\\
&= \prod_{\ell=1}^{p}\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)-\prod_{\ell=1}^{p}\{\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\}\left\{1-p+\sum_{\ell=1}^{p}\frac{\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)}{\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)}\right\}.
\end{aligned}$$

The derivation of the pseudo-covariance function starts with the decomposition

$$\mathrm{D}_n(\underset{\sim}{s})\mathrm{D}_n(\underset{\sim}{t})=\hat{\varphi}_n(\underset{\sim}{s})\hat{\varphi}_n(\underset{\sim}{t})-\hat{\varphi}_n(\underset{\sim}{s})\tilde{\varphi}_n(\underset{\sim}{t})-\tilde{\varphi}_n(\underset{\sim}{s})\hat{\varphi}_n(\underset{\sim}{t})+\tilde{\varphi}_n(\underset{\sim}{s})\tilde{\varphi}_n(\underset{\sim}{t})$$

and follows similar steps like the derivation of the covariance function.

Remark 2. An alternative derivation of the limiting covariance structure has been suggested by an anonymous referee. It is based on defining the functional

$$\mathfrak{F}(g)(\underset{\sim}{t})=g(\underset{\sim}{t})-\prod_{j=1}^{p}g(\boldsymbol{0},\boldsymbol{t}_j,\boldsymbol{0})$$

and applying the functional delta method to $\sqrt{n}\{\mathfrak{F}(\hat{\varphi}_n)-\mathfrak{F}(\varphi)\}$. Hadamard differentiability of this functional under the null hypothesis can be shown and its formal Hadamard derivative can be derived easily. By substituting in this formal expression the $\varphi_j, j=1,\ldots,p$ functions, we can get after some algebraic transformations the covariance function and the pseudo-covariance functions as stated in Theorem 1.

Proof of Theorem 2. A complex Gaussian process indexed by $\underset{\sim}{t}\in\mathbb{R}^q$ is characterized by a triplet. This includes the mean vector $\mu(\underset{\sim}{t})$ (which is 0 in our case), the covariance function $C(\underset{\sim}{s},\underset{\sim}{t})$, and the pseudo-covariance function $P(\underset{\sim}{s},\underset{\sim}{t})=C(\underset{\sim}{s},-\underset{\sim}{t})$. These satisfy the following properties:

1. $\overline{C(\underset{\sim}{s},\underset{\sim}{t})}=C(-\underset{\sim}{s},-\underset{\sim}{t})$
2. $C(\underset{\sim}{s},\underset{\sim}{t})$ is hermitian: $\overline{C(\underset{\sim}{s},\underset{\sim}{t})}=C(\underset{\sim}{t},\underset{\sim}{s})$.
3. $P(\underset{\sim}{s},\underset{\sim}{t})$ is symmetric: $P(\underset{\sim}{s},\underset{\sim}{t})=P(\underset{\sim}{t},\underset{\sim}{s})$.
4. $P(\underset{\sim}{s},\underset{\sim}{t})$ is centro-hermitian: $\overline{P(\underset{\sim}{s},\underset{\sim}{t})}=P(-\underset{\sim}{t},-\underset{\sim}{s})$.
5. It holds: $\overline{C(-\underset{\sim}{s},\underset{\sim}{t})}=P(\underset{\sim}{s},\underset{\sim}{t})=C(\underset{\sim}{s},-\underset{\sim}{t})$.
6. $C(\underset{\sim}{s},\underset{\sim}{t})$ is positive definite: $\forall_{u\in L_1(\mathbb{R}^q)}$ it holds
$$\iint C(\underset{\sim}{s},\underset{\sim}{t})u(\underset{\sim}{s})\overline{u(\underset{\sim}{t})}w(\underset{\sim}{s})w(\underset{\sim}{t})d\underset{\sim}{s}d\underset{\sim}{t}=\mathbb{E}\left\{\int \mathrm{D}(\underset{\sim}{s})u(\underset{\sim}{s})w(\underset{\sim}{s})d\underset{\sim}{s}\right\}^2\geq 0.$$
7. $C(\cdot,\cdot)$ is continuous because $\varphi(\underset{\sim}{t})$ is continuous.

In particular, the finite m-dimensional joint distributions at a set of points $\underset{\sim}{t}_1,\underset{\sim}{t}_2,\ldots,\underset{\sim}{t}_m$ represent a complex m-dimensional Gaussian vector $\mathbf{Z}$ with distribution denoted $\mathbf{Z}\sim CN_m(\mu,C_m,P_m)$ and characterized by a mean vector μ, a covariance matrix C_m and a pseudo-covariance matrix P_m. This means that $\mathbf{Z}=\mathbf{W}_1+i\mathbf{W}_2$ where $\mathbf{W}=(\mathbf{W}_1{}^\top,\mathbf{W}_2{}^\top)^\top$ has a $2m$-dimensional (real) multivariate normal distribution with $E\mathbf{W}_1=\mu_1, E\mathbf{W}_2=\mu_2, \mathbb{C}\mathrm{ov}(\mathbf{W}_1,\mathbf{W}_1)=\Sigma_{11}, \mathbb{C}\mathrm{ov}(\mathbf{W}_2,\mathbf{W}_2)=\Sigma_{22}, \mathbb{C}\mathrm{ov}(\mathbf{W}_1,\mathbf{W}_2)=\Sigma_{12}$. The relationships

$$\Sigma_{11}=\frac{1}{2}\Re(C_m+P_m),\ \Sigma_{12}=\frac{1}{2}\Im(-C_m+P_m),\ \Sigma_{21}=\frac{1}{2}\Im(C_m+P_m),\ \Sigma_{22}=\frac{1}{2}\Re(C_m-P_m)$$

hold. We also have

$$\begin{aligned}
C_m&=\mathbb{E}\left[(\mathbf{Z}-\mu)(\overline{\mathbf{Z}-\mu})'\right]=\Sigma_{11}+\Sigma_{22}+i(\Sigma_{21}-\Sigma_{12}),\\
P_m&=\mathbb{E}\left[(\mathbf{Z}-\mu)(\mathbf{Z}-\mu)'\right]=\Sigma_{11}-\Sigma_{22}+i(\Sigma_{21}+\Sigma_{12})
\end{aligned}$$

(see, e.g., [8], which generalizes the particular case of $\Sigma_{12}=\Sigma_{21}=0$, $\Sigma_{11}=\Sigma_{22}$ considered originally by Wooding [35]).

Observe that $\lambda f(-\boldsymbol{x})=\int_{\mathbb{R}^q}f(-\boldsymbol{y})\overline{C(\boldsymbol{x},\boldsymbol{y})}w(\boldsymbol{y})d\boldsymbol{y}$ holds, i.e., $f(-\boldsymbol{x})$ is an eigenfunction of $\overline{C(\cdot,\cdot)}$ associated with λ. Moreover, $\lambda\overline{f(-\boldsymbol{x})}=\int_{\mathbb{R}^q}\overline{f(-\boldsymbol{y})}C(\boldsymbol{x},\boldsymbol{y})w(\boldsymbol{y})d\boldsymbol{y}$, i.e., $\overline{f(-\boldsymbol{x})}$ is an eigenfunction of $C(\cdot,\cdot)$ associated with λ. If λ is of multiplicity 1, then $\overline{f(-\boldsymbol{x})}=e^{i\theta}f(\boldsymbol{x})$ for some θ in $[0,2\pi)$. Proving the Theorem boils down to checking that the

conditions of Mercer's theorem in [11] are satisfied. Then we would have $C(\underset{\sim}{s}, \underset{\sim}{t}) = \sum_{k=1}^{\infty} \lambda_k f_k(\underset{\sim}{s})\overline{f_k(\underset{\sim}{t})}$, where $\langle f, g\rangle_w = \int f(\underset{\sim}{t})\overline{g(\underset{\sim}{t})}w(\underset{\sim}{t})\,d\underset{\sim}{t}$. It is not hard to show that these quantities satisfy

$$\int C(\underset{\sim}{s}, \underset{\sim}{t}) f_j(\underset{\sim}{t})\, w(\underset{\sim}{t})\, d\underset{\sim}{t} = \sum_{k=1}^{\infty} \lambda_k f_k(\underset{\sim}{s}) \int f_j(\underset{\sim}{t})\overline{f_k(\underset{\sim}{t})} w(\underset{\sim}{t})\, d\underset{\sim}{t} = \lambda_j f_j(\underset{\sim}{s}).$$

Similarly:

$$\int P(\underset{\sim}{s}, \underset{\sim}{t})\overline{f_k(\underset{\sim}{t})} w(\underset{\sim}{t})\, d\underset{\sim}{t} = \lambda_k \overline{f_k(-\underset{\sim}{s})} \quad \text{and} \quad \int P(\underset{\sim}{s}, \underset{\sim}{t}) f_k(-\underset{\sim}{s})\, w(\underset{\sim}{s})\, d\underset{\sim}{s} = \lambda_k f_k(\underset{\sim}{t}).$$

Recall the limiting complex valued Gaussian stochastic process D discussed in Theorem 1. A stochastic expansion of D on the $\{f_k\}$ orthonormal basis gives:

$$\mathrm{D}(\underset{\sim}{t}) = \sum_{k=1}^{\infty} \langle \mathrm{D}, f_k\rangle_w f_k(\underset{\sim}{t}).$$

Hence

$$\int |\mathrm{D}(\underset{\sim}{t})|^2 w(\underset{\sim}{t})\, d\underset{\sim}{t} = \sum_k |\langle \mathrm{D}, f_k\rangle|_w^2.$$

Now

$$\langle \mathrm{D}, f_k\rangle_w = \int \mathrm{D}(\underset{\sim}{t})\overline{f_k(\underset{\sim}{t})} w(\underset{\sim}{t})\, d\underset{\sim}{t} \sim CN_1(0, \lambda_k, \lambda_k\overline{p_k}).$$

Indeed, $\mathrm{D}(\underset{\sim}{t})$ is a zero mean complex Gaussian process and by approximating the integral by weighted Riemann sums, each of which is Gaussian, shows us to the statement that the limiting random variable is Gaussian. We also get easily the variance and the pseudo-variance as follows:

$$\begin{aligned}\mathbb{E}\left[\langle \mathrm{D}, f_k\rangle_w \overline{\langle \mathrm{D}, f_k\rangle_w}\right] &= \iint C(\underset{\sim}{s}, \underset{\sim}{t})\overline{f_k(\underset{\sim}{s})} f_k(\underset{\sim}{t}) w(\underset{\sim}{t}) w(\underset{\sim}{s}) d\underset{\sim}{t} d\underset{\sim}{s} \\ &= \int \lambda_k f_k(\underset{\sim}{s})\overline{f_k(\underset{\sim}{s})} w(\underset{\sim}{s}) d\underset{\sim}{s} = \lambda_k\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left[\langle \mathrm{D}, f_k\rangle_w \langle \mathrm{D}, f_k\rangle_w\right] &= \iint P(\underset{\sim}{s}, \underset{\sim}{t})\overline{f_k(\underset{\sim}{t}) f_k(\underset{\sim}{s})} w(\underset{\sim}{t}) w(\underset{\sim}{s}) d\underset{\sim}{t} d\underset{\sim}{s} \\ &= \lambda_k \int \overline{f_k(\underset{\sim}{s}) f_k(-\underset{\sim}{s})} w(\underset{\sim}{s}) d\underset{\sim}{s} = \lambda_k \overline{p_k}.\end{aligned}$$

Note that if λ_k is of multiplicity 1 then we have

$$p_k = \int f_k(\underset{\sim}{s}) f_k(-\underset{\sim}{s}) w(\underset{\sim}{s}) d\underset{\sim}{s} = \int f_k(\underset{\sim}{s})\overline{e^{i\theta} f_k(\underset{\sim}{s})} w(\underset{\sim}{s}) d\underset{\sim}{s} = e^{-i\theta}.$$

In particular, $|p_k| = 1$ in this case. In the case of multiple roots and $|p_k| \neq 1$ we then must have $p_k = 0$ because of the orthogonality of the eigenfunctions in the scalar product representation $p_k = \int f_k(\boldsymbol{s})\overline{\overline{f_k(-\boldsymbol{s})} w(\boldsymbol{s}) d\boldsymbol{s}}$. We use [8] with R as in their Eq. (13), with $G = 1/2$ and $K = 0$ such that $\underset{\sim}{Z}^{\mathcal{H}} R \underset{\sim}{Z} = |\underset{\sim}{Z}|^2 \sim \alpha_1 \chi_1^2 + \alpha_2 \chi_1^2$, where α_1 and α_2 are the eigenvalues of the matrix

$$\Gamma_P R = \begin{pmatrix} \lambda_k & \lambda_k \overline{p}_k \\ \lambda_k p_k & \lambda_k \end{pmatrix} \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} = \frac{\lambda_k}{2} \begin{pmatrix} 1 & \overline{p}_k \\ p_k & 1 \end{pmatrix}.$$

That is, we need to multiply by $\lambda_k/2$ the eigenvalues of the identity matrix. We thus obtain (for $|p_k| \neq 1$)

$$|\langle D, f_k\rangle|^2 \sim \frac{\lambda_k}{2}(\xi_k + \eta_k) \tag{24}$$

where the two χ_1^2 random variables ξ_k and η_k are independent.

When $|p_k| = 1$, we obtain

$$|\langle D, f_k\rangle|^2 \sim \lambda_k \chi_1^2$$

because the first eigenvalue of the rank-1 matrix $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is 2.

We note about the complications in applying Mercer's theorem in our case. They are due to the fact that the functional $nT_n(w) = \|\sqrt{n}\mathrm{D}_n\|_w^2$ defined on the left hand side in (9) as our test statistic is not continuous; it is even not defined on $C(\mathbb{R}^q, \mathbb{C})$ but only on the subset of squared-integrable functions with respect to the measure $w(\underset{\sim}{t})d\underset{\sim}{t}$. Thus the standard continuous mapping theorem cannot be invoked to claim directly the convergence in (9). We need a generalization on a uniform integrability as presented in [3, Theorem 2.3]. We need to show that

$$A := \lim_{j\to\infty} \limsup_{n\to\infty} \int_{\mathbb{R}^q \setminus \mathcal{B}_j^q} \mathbb{E}|\sqrt{n}\mathrm{D}_n(\underset{\sim}{t})|^2 w(\underset{\sim}{t})\, d\underset{\sim}{t} = 0.$$

We already showed that

$$\mathbb{E}\left(n\mathrm{D}_n(\underset{\sim}{t})\,\overline{\mathrm{D}_n(\underset{\sim}{t})}\right) = C(\underset{\sim}{t}, \underset{\sim}{t}) + \frac{1}{n}O(1)_{(\underset{\sim}{t})}.$$

Moreover,

$$\int C(\underset{\sim}{t}, \underset{\sim}{t})w(\underset{\sim}{t})d\underset{\sim}{t} \le \int |C(\underset{\sim}{t}, \underset{\sim}{t})|w(\underset{\sim}{t})d\underset{\sim}{t} < \infty \tag{25}$$

holds. From the form of the remainder we see that

$$\lim_{j\to\infty} \limsup_{n\to\infty} \int_{\mathbb{R}^q \setminus \mathcal{B}_j^q} O(1)_{(\underset{\sim}{t})} w(\underset{\sim}{t})d\underset{\sim}{t} \to 0.$$

Also, since (7) holds then not only (25) but also

$$\int |C(\underset{\sim}{s}, \underset{\sim}{t})|w(\underset{\sim}{t})d\underset{\sim}{t} < \infty$$

holds for all $\underset{\sim}{s}$ holds. Hence by [4, Theorem 1.6, p. 28] we have that the operator $\mathbf{K}(f)(\underset{\sim}{x}) = \int_{\mathbb{R}^q} f(\underset{\sim}{y})C(\underset{\sim}{x}, \underset{\sim}{y})w(\underset{\sim}{y})d\underset{\sim}{y}$ is a bounded linear operator. Further by the assumption

$$\iint |C(\underset{\sim}{s}, \underset{\sim}{t})|^2 w(\underset{\sim}{s})w(\underset{\sim}{t})d\underset{\sim}{s}d\underset{\sim}{s} < \infty$$

and by [4, Proposition 4.7, p. 43], $\mathbf{K}(f)(\underset{\sim}{x})$ is also a compact operator. Then by [11, p. 69], Mercer's theorem applies and hence we establish the convergence of (9) as required.

Proof of Theorem 3. Let $\mathcal{N} = \left\{\omega \in \Omega; A_n(\omega) \underset{n\to\infty}{\longrightarrow} A(\omega)\right\}$. Obviously, $\Pr(\Omega \setminus \mathcal{N}) = 0$. Let $\omega \in \mathcal{N}$ be fixed. Let $\sigma(A_n)$ be the spectrum of A_n which is the set of eigenvalues for A_n and similarly, $\sigma(A)$ be the spectrum of A. We write $P_A(z) = \alpha_0 + \alpha_1 z + \cdots + \alpha_N z^N$ as the characteristic polynomial of $A(\omega)$. Recall that P_A is a polynomial of exact degree N and the eigenvalues are the roots of the polynomial (i.e., $P_A(z) = det(A - zI)$.) Similarly, let $P_n(z) = \alpha_{0,n} + \alpha_{1,n}z + \cdots + \alpha_{N,n}z^N$ be the characteristic polynomial of $A_n(\omega)$. We have, for all $k = 1, \ldots, N$, that α_k is a (finite) linear combination of the N^2 values a_{ij} (because $P_A(z) = \det(A(\omega) - zI)$) and that $\alpha_{k,n}$ is a linear combination of the N^2 values $a_{ij,n}$. Now, since for all $i, j = 1, \ldots, N$, $a_{ij,n} \underset{n\to\infty}{\longrightarrow} a_{ij}$, we have that for all $k = 1, \ldots, N$, $\alpha_{k,n} \underset{n\to\infty}{\longrightarrow} \alpha_k$, where $\alpha_{k,n} = \alpha_k + \epsilon_{k,n}$, and $\max_k |\epsilon_{k,n}| \underset{n\to\infty}{\longrightarrow} 0$ when $\omega \in \mathcal{N}$. Suppose $\mu \in \sigma(A)$ be one of the eigenvalues of $A(\omega)$ with multiplicity m. Then for all $\eta > 0$ sufficiently small, no other distinct root of $P_A(z)$ different from μ will be in the disc $\{|z - \mu| \le \eta\}$ with a centre μ. Let us fix such a $\eta > 0$ and let

$$M = \min_{|z-\mu|=\eta} |P_A(z)|.$$

Note that $M > 0$ since there are no roots of $P_A(z)$ on the boundary $|z - \mu| = \eta$. Moreover, since all the N coefficients of the polynomial $P_n(z) - P_A(z)$ converge to 0 as $n \to \infty$, then there exists a $L \in \mathbb{N}$ such that for all $n > L$,

$$|P_n(z) - P_A(z)| < M \le |P_A(z)|, \quad \text{for } |z - \mu| = \eta.$$

By *Rouché's Theorem* (see, e.g., [27, p. 229]), $P_n(z)$ has exactly m roots in $|z - \mu| < \eta$. Since η can be taken arbitrarily small, these m roots of P_n (which are m eigenvalues of $A_n(\omega)$) must converge to μ.

Now suppose $\lambda_n \in \sigma(A_n)$ is one of these m roots and its associated eigenvector $\mathbf{f}_n \in K_n := \{\mathbf{f}_n \in \mathbb{C}^N : (A_n - \lambda_n \mathbf{I})\mathbf{f}_n = 0\}$, i.e.

$$A_n(\omega)\mathbf{f}_n = \lambda_n \mathbf{f}_n.$$

Then $K_n \underset{n\to\infty}{\longrightarrow} \bar{K} \subseteq K := \{\mathbf{f} \in \mathbb{C}^N : (A - \mu\mathbf{I})\mathbf{f} = 0\}$ since $A_n(\omega) \underset{n\to\infty}{\longrightarrow} A(\omega)$ and $\lambda_n \underset{n\to\infty}{\longrightarrow} \mu$. That is, the accumulation point of K_n must converge to an element of $ker\{A - \mu\mathbf{I}\}$ (see [34, Section 25 p.83]). Hence the result follows.

Proof of Theorem 4. We start with part (a).

It is well known that $\hat{\varphi}_n(\underset{\sim}{t}) \xrightarrow{a.s.}_{n\to\infty} \varphi(\underset{\sim}{t})$ for all $\underset{\sim}{t}$ (see, e.g., [12]). Then we have

$$\hat{C}_n(\underset{\sim}{t}, \underset{\sim}{s}) \xrightarrow{a.s.}_{n\to\infty} C(\underset{\sim}{t}, \underset{\sim}{s}) \quad \text{for all } \underset{\sim}{t}, \underset{\sim}{s} \in \mathbb{R}^q. \tag{26}$$

Let us fix N first. Applying (26) we then have, conditionally on the generated $\underset{\sim}{\mathbf{Y}}_1, \underset{\sim}{\mathbf{Y}}_2, \ldots, \underset{\sim}{\mathbf{Y}}_N$ values:

$$\left(\hat{C}_n(\underset{\sim}{\mathbf{Y}}_i, \underset{\sim}{\mathbf{Y}}_j)\right)_{i,j} \xrightarrow{a.s.}_{n\to\infty} \left(C(\underset{\sim}{\mathbf{Y}}_i, \underset{\sim}{\mathbf{Y}}_j)\right)_{i,j}.$$

The almost sure convergence holds for any of the (i, j)th entry on the left to its matching entry on the right. Thus by Theorem 3 we have

$$\lambda_{(k),n,N} \xrightarrow{a.s.}_{n\to\infty} \lambda_{(k),N}, \quad k = 1, 2, \ldots, N,$$

and thus $\lambda_{(k),n,N} \xrightarrow{P}_{n\to\infty} \lambda_{(k),N}$, $k = 1, 2, \ldots, N$. Note that $C(\cdot, \cdot)$ is continuous. Moreover, clearly the uniform bound from above

$$|C(\underset{\sim}{t}, \underset{\sim}{t})| \le 2(p+1) := \kappa$$

holds. Proposition 10 in [26] is precisely tailored for such continuous and uniformly bounded kernels and their empirical counterparts. According to this Proposition, for any $\tau > 0$ the bound

$$\Pr\left[\sum_{k\ge1} \left(\lambda_{(k),N} - \lambda_{(k)}\right)^2 \le \frac{8\kappa^2\tau}{N}\right] \ge 1 - 2e^{-\tau} \tag{27}$$

holds. Hence, for arbitrary $\epsilon > 0$, by setting $\tau = N\epsilon/8\kappa^2$ and substituting in (27) we get

$$\Pr\left[\sum_{k\ge1} (\lambda_{(k),N} - \lambda_{(k)})^2 \le \epsilon\right] \ge 1 - 2e^{-\tau}.$$

Hence

$$\Pr\left[\sum_{k\ge1} (\lambda_{(k),N} - \lambda_{(k)})^2 \ge \epsilon\right] \longrightarrow_{N\to\infty} 0.$$

Therefore for all $k \ge 1$, we also have $0 \le \Pr\left[(\lambda_{(k),N} - \lambda_{(k)})^2 \ge \epsilon\right] \longrightarrow_{N\to\infty} 0$ and it immediately follows that $\lambda_{(k),N} \xrightarrow{P}_{N\to\infty} \lambda_{(k)}$. Now,

$$\begin{aligned} 0 &\le \Pr\left(|\lambda_{(k),n,N} - \lambda_{(k)}| > \epsilon\right) \\ &\le \Pr\left(|\lambda_{(k),n,N} - \lambda_{(k),N}| > \frac{\epsilon}{2}\right) + \Pr\left(|\lambda_{(k),N} - \lambda_{(k)}| > \frac{\epsilon}{2}\right) \to_{n,N\to\infty} 0. \end{aligned}$$

Denoting $\hat{\lambda}_k = \lambda_{(k),n,N}$ and $\lambda_k = \lambda_{(k)}$ we have $\lambda_k - \hat{\lambda}_k \xrightarrow{P}_{n\to\infty} 0$. Hence by of [10, Theorem 6' (a), p. 42], the statement follows.

The part (b) of the Theorem follows the same steps as part (a). The convergence of the estimated eigenfunctions $\hat{f}_{k,n,N}(\cdot)$ to the (spaces of the) true ones $f_k(\cdot)$ is also obtained in two steps, first letting $n \to \infty$ while (our) N is fixed via Theorem 3, and then using [26, Theorem 12] when (our) N goes to infinity.

We note that [26, Proposition 10] concerns the *extended* enumeration of the eigenvalues of the true and of the empirical kernel functions. The extended enumeration is a sequence of real numbers where every non-zero eigenvalue appears as many times as its multiplicity. Hence the multiple eigenvalues case of part (b) is also included and the only modification from part (a) follows the same argument that was brought forward in the justification of (24) in the proof of Theorem 2.

Proof of formula (23). We introduce $\vartheta = 1 - \varphi$. Then we can write:

$$\begin{aligned} C(\underset{\sim}{t}, \underset{\sim}{t}) &= 1 - \prod_{\ell=1}^{p} \{\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{t}_\ell)\} - \sum_{\ell=1}^{p} \{1 - \varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{t}_\ell)\} \prod_{j=1;j\ne\ell}^{p} \varphi_j(-\boldsymbol{t}_j)\varphi_j(\boldsymbol{t}_j) \\ &= 1 - (1-p) \sum_{B\subset I_p} \sum_{B'\subset I_p} (-1)^{|B|+|B'|} \prod_{\ell\in B\cap B'} \vartheta_\ell(-\boldsymbol{t}_\ell)\vartheta_\ell(\boldsymbol{t}_\ell) \prod_{\ell\in B\setminus B'} \vartheta_\ell(-\boldsymbol{t}_\ell) \prod_{\ell\in B'\setminus B} \vartheta_\ell(\boldsymbol{t}_\ell) \\ &\quad - \sum_{\ell=1}^{p} \sum_{B\subset I_p\setminus\ell} \sum_{B'\subset I_p\setminus\ell} (-1)^{|B|+|B'|} \prod_{j\in B\cap B'} \vartheta_j(-\boldsymbol{t}_j)\vartheta_j(\boldsymbol{t}_j) \prod_{j\in B\setminus B'} \vartheta_j(-\boldsymbol{t}_j) \prod_{j\in B'\setminus B} \vartheta_j(\boldsymbol{t}_j). \end{aligned}$$

Using the notation $b_\ell = n^{-2}\sum_{j=1}^n\sum_{j'=1}^n \beta_{j,j',\ell}$ and $g_\ell = n^{-1}\sum_{j=1}^n \gamma_{j,\ell}$ we get:

$$\begin{aligned}
\left\|\sqrt{\hat{C}_n(\underset{\sim}{t},\underset{\sim}{t})}\right\|_w^2 &= 1-(1-p)\sum_{B\subset I_p}\sum_{B'\subset I_p}(-1)^{|B|+|B'|}\prod_{\ell\in B\cap B'}b_\ell\prod_{\ell\in B\setminus B'}g_\ell\prod_{\ell\in B'\setminus B}g_\ell\\
&\quad-\sum_{\ell=1}^p\sum_{B\subset I_p\setminus\ell}\sum_{B'\subset I_p\setminus\ell}(-1)^{|B|+|B'|}\prod_{j\in B\cap B'}b_j\prod_{j\in B\setminus B'}g_j\prod_{j\in B'\setminus B}g_j\\
&= 1-(1-p)\sum_{B\subset I_p}\sum_{B'\subset I_p}\prod_{\ell\in B\cap B'}b_\ell\prod_{\ell\in(B\cup B')\setminus(B\cap B')}(-g_\ell)\\
&\quad-\sum_{\ell=1}^p\sum_{B\subset I_p\setminus\ell}\sum_{B'\subset I_p\setminus\ell}\prod_{j\in B\cap B'}b_j\prod_{j\in(B\cup B')\setminus(B\cap B')}(-g_j)\\
&= 1-(2-p)\sum_{B\subset I_p}\sum_{B'\subset I_p}\prod_{\ell\in B\cap B'}b_\ell\prod_{\ell\in(B\cup B')\setminus(B\cap B')}(-g_\ell)\\
&\quad+\sum_{B\subset I_p}\sum_{B'\subset I_p}(|B\cup B'|-p+1)\prod_{\ell\in B\cap B'}b_\ell\prod_{\ell\in(B\cup B')\setminus(B\cap B')}(-g_\ell)\\
&= 1+\sum_{B\subset I_p}\sum_{B'\subset I_p}(|B\cup B'|-1)\prod_{\ell\in B\cap B'}b_\ell\prod_{\ell\in(B\cup B')\setminus(B\cap B')}(-g_\ell)\\
&= \sum_{B,B'\subset I_p;|B\cup B'|>1}|B\cup B'|\prod_{\ell\in B\cap B'}b_\ell\prod_{\ell\in(B\cup B')\setminus(B\cap B')}(-g_\ell)+2\sum_{B\subset I_p;|B|>1}|B|\prod_{\ell\in B}(-g_\ell). \qquad (28)
\end{aligned}$$

It is easy to see that

$$1-\left\{1+\sum_{l=1}^p(|\hat{\varphi}_{n,\ell}(\boldsymbol{t}_\ell)|^{-2}-1)\right\}\prod_{l=1}^p|\hat{\varphi}_{n,\ell}(\boldsymbol{t}_\ell)|^2=\widehat{C}_n(\underset{\sim}{t},\underset{\sim}{t}) \qquad (29)$$

hence $H_n=\left\|\sqrt{\hat{C}_n(\underset{\sim}{t},\underset{\sim}{t})}\right\|_w^2$ holds. Indeed, since $C(\underset{\sim}{t},\underset{\sim}{t})=\mathbb{E}\left\{\mathrm{D}(\underset{\sim}{t})\overline{\mathrm{D}(\underset{\sim}{t})}\right\}$ and

$$\begin{aligned}
C(\underset{\sim}{s},\underset{\sim}{t}) &= \prod_{\ell=1}^p\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)-\prod_{\ell=1}^p\{\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\}\\
&\quad-\sum_{\ell=1}^p\{\varphi_\ell(-\boldsymbol{t}_\ell+\boldsymbol{s}_\ell)-\varphi_\ell(-\boldsymbol{t}_\ell)\varphi_\ell(\boldsymbol{s}_\ell)\}\prod_{j=1;j\neq\ell}^p\varphi_j(-\boldsymbol{t}_j)\varphi_j(\boldsymbol{s}_j)
\end{aligned}$$

we see that

$$C(\underset{\sim}{t},\underset{\sim}{t})=1-\prod_{\ell=1}^p|\varphi_\ell(\boldsymbol{t}_\ell)|^2-\sum_{\ell=1}^p\left\{1-|\varphi_\ell(\boldsymbol{t}_\ell)|^2\right\}\prod_{j=1;j\neq\ell}^p\left|\varphi_j(\boldsymbol{t}_j)\right|^2. \qquad (30)$$

Substituting the empirical characteristic functions in (30) we get (29).

Note that, again like in Lemma 1, a simplification occurs in (28) for the particular case where the weight function is integrable, i.e., when $\int_{\mathbb{R}^q}w(\underset{\sim}{t})\,d\underset{\sim}{t}=1$. In this case, the denominator $H_n=\left\|\sqrt{1-\left\{1+\sum_{\ell=1}^p(|\hat{\varphi}_{n,\ell}|^{-2}-1)\right\}\prod_{\ell=1}^p|\hat{\varphi}_{n,\ell}|^2}\right\|_w^2$ is equal to

$$\begin{aligned}
&\int_{\mathbb{R}^q}\left[1-\left\{1+\sum_{\ell=1}^p(|\hat{\varphi}_{n,\ell}(\boldsymbol{t}_\ell)|^{-2}-1)\right\}\prod_{\ell=1}^p|\hat{\varphi}_{n,\ell}(\boldsymbol{t}_\ell)|^2\right]w(\underset{\sim}{t})d\underset{\sim}{t}\\
&=1-(1-p)\prod_{\ell=1}^p\int_{\mathbb{R}^{q_\ell}}|\hat{\varphi}_{n,\ell}(\boldsymbol{t}_\ell)|^2v(\underset{\sim}{t}_\ell)d\underset{\sim}{t}_\ell-\sum_{\ell=1}^p\prod_{\ell'=1;\ell'\neq\ell}^p\int_{\mathbb{R}^{q_{\ell'}}}|\hat{\varphi}_{n,\ell'}(\boldsymbol{t}_{\ell'})|^2v(\underset{\sim}{t}_{\ell'})d\underset{\sim}{t}_{\ell'}\\
&=1-(1-p)\prod_{\ell=1}^p x_\ell-\sum_{\ell'=1}^p\frac{\prod_{\ell=1}^p x_\ell}{x_{\ell'}}=1-\left\{1+\sum_{\ell=1}^p(x_\ell^{-1}-1)\right\}\prod_{\ell=1}^p x_\ell
\end{aligned}$$

where

$$x_\ell := \int_{\mathbb{R}^{q_\ell}} |\hat{\varphi}_{n,\ell}(\underset{\sim}{t}_\ell)|^2 v(\underset{\sim}{t}_\ell) d\underset{\sim}{t}_\ell = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\int_{\mathbb{R}^{q_\ell}} e^{i\underset{\sim}{t}_\ell{}^\top(\mathbf{x}_j^\ell - \mathbf{x}_{j'}^\ell)} v(\underset{\sim}{t}_\ell) d\underset{\sim}{t}_\ell = \frac{1}{n^2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\xi_{j,j',\ell}.$$

Acknowledgements

The research of Pierre Lafaye de Micheaux has been funded by the Natural Sciences and Engineering Research Council of Canada (grant RGPIN-2014-06348). The research of Spiridon Penev was supported under Australian Research Council's Discovery Project funding Scheme (project DP160103489). We are grateful to the editor, the associate editor and to the two anonymous referees. Their criticism, insightful comments and suggestions helped us to improve significantly the content, style and presentation of the paper.

References

[1] N.K. Bakirov, M.L. Rizzo, G.J. Székely, A multivariate nonparametric test of independence, J. Multivariate Anal. 97 (8) (2006) 1742–1756.
[2] R. Beran, M. Bilodeau, P. Lafaye de Micheaux, Nonparametric tests of independence between random vectors, J. Multivariate Anal. 98 (9) (2007) 1805–1824.
[3] M. Bilodeau, P. Lafaye de Micheaux, A multivariate empirical characteristic function test of independence with normal marginals, J. Multivariate Anal. 95 (2) (2005) 345–369.
[4] J.B. Conway, A Course in Functional Analysis, second ed., Springer Verlag, New York, 1997.
[5] S. Csörgő, Multivariate empirical characteristic functions, Z. Wahrscheinlichkeitstheor. Verwandte Geb. 55 (2) (1981) 203–229.
[6] S. Csörgő, Testing for independence by the empirical characteristic function, J. Multivariate Anal. 16 (3) (1985) 290–299.
[7] R. Davies, The distribution of a linear combination of chi-square random variables, J. R. Stat. Soc. Ser. C. Appl. Stat. 29 (3) (1980) 323–333.
[8] G. Ducharme, P. Lafaye de Micheaux, B. Marchina, The complex multinomial distribution, quadratic forms in complex random vectors and an omnibus goodness-of-fit test for complex normal distribution, Ann. Inst. Statist. Math. 68 (1) (2016) 77–104.
[9] P. Duchesne, P. Lafaye De Micheaux, Computing the distribution of quadratic forms: further comparisons between the Liu-Tang-Zhang approximation and exact methods, Comput. Statist. Data Anal. 54 (4) (2010) 858–862.
[10] T.S. Ferguson, A Course in Large Sample Theory, Chapman & Hall/CRC, 1996.
[11] J.C. Ferreira, V.A. Menegatto, Positive definiteness, reproducing kernel hilbert spaces and beyond, Ann. Funct. Anal. 4 (1) (2013) 64–88.
[12] A. Feuerverger, R.A. Mureika, The empirical characteristic function and its applications, Ann. Statist. 5 (1) (1977) 88–97.
[13] C. Genest, J.-F. Quessy, B. Remillard, Asymptotic local efficiency of Cramér–von Mises tests for multivariate independence, Ann. Statist. 35 (1) (2007) 166–191.
[14] K. Ghoudi, R.J. Kulperger, B. Rémillard, A nonparametric test of serial independence for time series and residuals, J. Multivariate Anal. 79 (2) (2001) 191–218.
[15] R. Heller, Y. Heller, M. Gorfine, A consistent multivariate test of association based on ranks of distances, Biometrika 100 (2) (2013) 503–510.
[16] N. Henze, Invariant tests for multivariate normality: a critical review, Statist. Papers 43 (4) (2002) 467–506.
[17] J.P. Imhof, Computing the distribution of quadratic forms in normal variables, Biometrika 48 (1961) 419–426.
[18] O. Kallenberg, Foundations of modern probability, in: Probability and its Applications (New York), Springer-Verlag, New York, 1997.
[19] A. Kankainen, N.G. Ushakov, A consistent modification of a test for independence based on the empirical characteristic function, J. Math. Sci. (N. Y.) 89 (5) (1998) 1486–1494. Stability problems for stochastic models, Part 1 (Moscow, 1996).
[20] I. Kojadinovic, M. Holmes, Tests of independence among continuous random vectors based on Cramér–von Mises functionals of the empirical copula process, J. Multivariate Anal. 100 (6) (2009) 1137–1154.
[21] C.S. Kubrusly, Spectral Theory of Operators on Hilbert Spaces, Birkhäuser/Springer, New York, 2012.
[22] C.J. Mecklin, D.J. Mundfrom, A Monte Carlo comparison of the type I and type II error rates of tests of multivariate normality, J. Stat. Comput. Simul. 75 (2) (2005) 93–107.
[23] S.G. Meintanis, G. Iliopoulos, Fourier methods for testing multivariate independence, Comput. Statist. Data Anal. 52 (4) (2008) 1884–1895.
[24] M.A. Newton, Introducing the discussion paper by Székely and Rizzo, Ann. Appl. Stat. 3 (4) (2009) 1233–1235.
[25] J.-F. Quessy, Applications and asymptotic power of marginal-free tests of stochastic vectorial independence, J. Statist. Plann. Inference 140 (11) (2010) 3058–3075.
[26] L. Rosasco, M. Belkin, E. De Vito, On learning with integral operators, J. Mach. Learn. Res. 11 (2010) 905–934.
[27] W. Rudin, Real and Complex Analysis, third ed., McGraw-Hill Book Co., New York, 1987.
[28] G.J. Székely, M.L. Rizzo, Brownian distance covariance, Ann. Appl. Stat. 3 (4) (2009) 1236–1265.
[29] G.J. Székely, M.L. Rizzo, On the uniqueness of distance covariance, Statist. Probab. Lett. 82 (12) (2012) 2278–2282.
[30] G.J. Székely, M.L. Rizzo, Energy statistics: a class of statistics based on distances, J. Statist. Plann. Inference 143 (8) (2013) 1249–1272.
[31] G.J. Székely, M.L. Rizzo, N.K. Bakirov, Measuring and testing dependence by correlation of distances, Ann. Statist. 35 (6) (2007) 2769–2794.
[32] N.G. Ushakov, Selected Topics in Characteristic Functions, in: Modern Probability and Statistics, VSP, Utrecht, 1999.
[33] A.W. van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, in: Springer Series in Statistics, Springer-Verlag, New York, 1996.
[34] J.H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, 1988.
[35] R.A. Wooding, The multivariate distribution of complex normal variables, Biometrika 43 (1956) 212–215.